Analysis of Thermal Motion Effects on the Electron Density via Computational Simulations

by

J. Robert Michael

A dissertation submitted in partial
fulfilment of the requirements of the degree of
Doctor of Philosophy in Computational Science

Middle Tennessee State University
December, 2014

Dissertation Committee:

Dr. Tibor Koritsanszky, Chair
Dr. Anatoliy Volkov
Dr. Jing Kong
Dr. Abdul Khaliq
Dr. Yuri Melnikov

To my beautiful wife,

Lauren,

who stood beside me every day

and encouraged me when it seemed impossible.

ABSTRACT

The Electron Density (ED) of a molecular structure can only be observed for large ensembles of molecules packed tightly in crystal structures in the solid state. Even then it cannot truly be observed, instead experimental measurements are taken via X-Ray Diffraction (XRD) and the resulting data is fitted to a theoretical ED model describing the probability of finding an electron inside an infinitesimal volume element.

The atoms (made up of negatively charged electrons orbiting a positively charged nucleus), which are bound together by electrostatic forces to form the molecule, are constantly in vibrational motion. Even at very low temperatures, quantum effects cause the nuclei to maintain vibration in approximately harmonic oscillations. This nuclear motion takes place on a much faster time scale than the XRD experiment which yields a thermally averaged view of the molecule in the XRD data.

A topological analysis of the static ED (the ED with non-vibrating nuclei centered at mean positions from the observations) are invaluable to chemistry as it yields many chemical properties of the molecule under observation. It is thus important to partition the observed (thermally averaged, dynamic) ED into contributions from the static ED and contributions from nuclear thermal smearing. This partitioning involves an approximation in which the atomic ED is believed to rigidly follow the motion of the nuclei and the resulting dynamic ED is expressed as the convolution between the static ED and the probability of nuclear displacements. The process of fitting parameters to the observed XRD data involves continually refining static and dynamic parameters (the parameters defining the static ED and the nuclear motion, respectively).

In this computational study, the refinement process is simulated and various aspects of the process are evaluated. Among others, important aspects under evaluation include the accuracy of the convolution approximation, the representation and expression of dynamic parameters, the uncertainty of the refinement parameters, and the expression of the static ED.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. Introduction[1]

The static Electron Density (ED) of a chemical system is the probability density function (PDF) defining the probability of finding an electron in an infinitesimal volume element of space for a stationary nuclear configuration. The topology of this PDF can be used to strictly define many static and reactive properties of a molecule. Fundamental concepts in chemistry such as bonding, stability and even the very definition of an atom are now able to be described quantitatively through a topological analysis of this scalar field.

Though the concepts of atoms and molecules have been around for centuries, many properties of individual molecules are not directly observable through experiment. A large ensemble of molecules in crystals, however, can be examined by an experimental technique known as X-Ray diffraction to observe the average of all molecular properties in the crystal sample. An atom is composed of negatively charged electrons and a positively charged nucleus (which is a combination of protons and neutrons). A molecule, which is composed of more than one atom bound together by electrostatic forces, will vibrate due to atomic motion "governed" by both internal and external degrees of freedom according to the laws of quantum mechanics (QM).

Important to QM is the notion that a molecule can be uniquely defined by the position of its nuclei and the number of its electrons. As a result, there is a delicate interplay between the nuclei and the electrons; the distribution of the electrons and the geometric

---

[1] Portions of this introduction also appear in The Journal of Mathematical Chemistry [79].

positions of nuclei are highly correlated. The interdependence of this relationship between electrons and nuclei is used to solve the Schrödinger equation by assuming a stationary nuclear geometry as initial conditions in the calculations, resulting in an electronic wave function, $\Psi(\{r_i\}; R)$, implicitly dependent upon a given nuclear geometry, R.

These initial conditions are due to the Born-Oppenheimer Approximation [1] in which it is assumed that the electronic and nuclear motion can be separated, $\Psi(\{r_i\}; R) = \psi(\{r_i\}; R)\Phi(R)$. Physically, this is validated due to the mass difference between nuclei and electrons - the contribution of nuclear motion comes from classical mechanics which the electronic motion is due to QM. The quantum nature of the electrons allows for an instantaneous rearrangement around the nuclei for any small nuclear displacement. Because the electrons are able to adapt to any arrangement of nuclei, the electronic wave function and related energy can be calculated for any nuclear structure. This allows for calculation of the 'equilibrium' geometry, the nuclear structure for which there are no more forces acting on the nuclei. This equilibrium structure is found by iteratively calculating the forces on the nuclei and rearranging them to minimize the energy. The potential energy for nuclear motion is a function of nuclear positions and has a global minimum at the equilibrium geometry. This potential energy surface can be used to calculate nuclear vibration amplitudes and frequencies by taking a quadratic approximation to the minimum of the energy surface. The resulting quadratic approximation expresses the internal motion of the molecule as harmonic oscillations of the nuclei.

Modeling X-Ray diffraction, an experimental technique which is used to observe the average ED of atoms and or molecules in crystals, is a non-trivial procedure. The static ED associated with the molecule and the thermal vibrations of nuclei both contribute to the

distribution of the observed diffraction pattern. The difficulty of deriving the static ED from the data which is least biased by the thermal motion is addressed in this study.

The treatment of thermal motion is a far from trivial exercise and continues to be of interest today [2, 3]. First-principle modeling of the static ED itself [4, 5, 6] along with analysis of experimental errors associated with the measurement [7, 8] continues to be improved upon in the current era of charge density research. When the uncertainty of topological properties of the static ED is analyzed, only the uncertainty of those parameters which contribute to the construction of the static density are considered. These 'static parameters' do not include the parameters which describe the motion of the nuclei. Since there is a correlation between these 'dynamic parameters' and the static parameters, any uncertainty quantification which considers only the static parameters is incomplete.

In this study, we develop a computational framework for estimating nuclear thermal motion effects on the molecular ED. This framework allows us to create a thermally averaged ED and evaluate standard uncertainties in the parameters of the static ED. By doing this, we are able to perform uncertainty quantification of topological properties in the static ED using not only static parameters, but dynamic parameters as well.

1.1 Scientific Background

The (static) Electron Density of a molecule with M positively charged nuclei and N negatively charged electrons gives the probability of finding an electron in the infinitesimally small space $d\boldsymbol{r}_x$ at the point $\boldsymbol{r}_x = (x, y, z)$:

$$\rho(\boldsymbol{r}_x) = N \int \Psi(\boldsymbol{r}_x, \boldsymbol{r}_2, \dots, \boldsymbol{r}_N; \boldsymbol{R})\Psi^*(\boldsymbol{r}_x, \boldsymbol{r}_2, \dots, \boldsymbol{r}_N; \boldsymbol{R})d\boldsymbol{r}_2 \dots d\boldsymbol{r}_N \qquad (1.1)$$

where Ψ is the spinless ground-state 'wave function' for the N electrons, distributed in the field of nuclei with fixed positions (given by the 3M vector $\boldsymbol{R}$). The standard definition of a wave function is a complex valued function which contains all possible information about the system, however it cannot be physically observed. The ED, on the other hand, is an everywhere positive, real valued, physically observable function which integrates over all space to the number of electrons in the molecule [9, 10].

The ground-state wave function is an eigenfunction of the stationary Hamiltonian operator ($\hat{H}$) with the eigenvalue being the lowest allowed energy of the molecule. The corresponding Schrödinger equation is given by:

$$\hat{H}\Psi = E\Psi \tag{1.2}$$

$$\hat{H} = \hat{T}_N(\boldsymbol{R}) + \hat{T}_e(\{\boldsymbol{r}_i\}) + \hat{V}_{eN}(\{\boldsymbol{r}_i\}, \boldsymbol{R}) + \hat{V}_{NN}(\boldsymbol{R}) + \hat{V}_{ee}(\{\boldsymbol{r}_i\}) \tag{1.3}$$

where $\boldsymbol{R}$ and $\{\boldsymbol{r}_i\}$ represent the 3M nuclear and 3N electronic coordinates, respectively. $\hat{T}$ and $\hat{V}$ are the Kinetic and Potential Energy operators which operate on either nuclei (N), electrons (e) or a combination of the two. The only term in Eqn. 1.3 which includes both nuclear and electronic variables is the $\hat{V}_{eN}$ (electron-nucleus interaction potential energy operator) which prevents the separation of the wave function into a product of nuclear and electronic functions [11].

In order to solve the Schrödinger equation, it is customary to invoke the Born-Oppenheimer Approximation which states that the true wave function, $\Psi(\{\boldsymbol{r}_i\}, \boldsymbol{R})$, may be approximated by $\phi(\{\boldsymbol{r}_i\}; \boldsymbol{R})\chi(\boldsymbol{R})$ where $\phi(\{\boldsymbol{r}_i\}; \boldsymbol{R})$ is now a parametric function of nuclear coordinates $\boldsymbol{R}$ with independent electronic coordinate variables $\{\boldsymbol{r}_i\}$. Conceptually, this is a valid approximation because the electrons are much less massive and move much faster

than the nuclei, so that the nuclei are almost fixed in relation to the motion of the electrons. This is mathematically validated by truncating the Taylor Series expansion of the Hamiltonian to the zeroth order where the motion of the nuclei does not exist [1, 11].

Using this approximation, the electronic wave function is often approximated for a fixed configuration of R^* to obtain $\phi(\{r_i\}, R^*)$ for which we can find the related electronic energy, $E_a(R^*)$. By adjusting the coordinates of the nuclei, a minimum value for E_a can be obtained. The optimized nuclear positions are referred to as the equilibrium molecular configuration. In this standard process, the forces acting on the nuclei are continually calculated and used to re-adjust the positions of the nuclei until the energy is minimized.

The importance of the wave function to chemistry is without parallel, as it uniquely defines all physical properties of the molecule including bond order, atomic boundaries, and energy of a system. The importance of the ED to quantum physics, then, is equally important as it allows the physical information in the $3N$ variable function, $\Psi(\{r_i\})$, to be expressed as a 3 dimensional function, $\rho(r)$, which makes the computational aspect of topological properties much easier.

We measure the ED by using "X-Ray Diffraction" (XRD) in the course of which closely packed molecules in the crystal lattice are bombarded with X-Ray photons. The distribution of the scattered photon intensity is determined by the periodic crystalline ED because electrons are better at scattering X-Rays than the nuclei [9, 12]. Diffraction methods are necessary tools for studying the ED because unless the molecules are closely packed and interacting (as they are in the solid state), it is virtually impossible to experimentally observe the ED of individual, isolated molecules.

Because the molecules form a tightly packed lattice in a crystal (the ED exhibits translational symmetry), constructive and destructive interference occurs, as described by Bragg's Law, which gives rise to discrete, observable intensities. These observable intensities, I_{hkl}, are each a result of coherent scattering with a crystal plane defined by "Miller Indices" $\boldsymbol{h} = (h, k, l)$. The intensity is measured and, along with the phase, ϕ_{hkl}, is related to the Bragg "Structure Factor" (SF), F_{hkl}, which is the FT of the average ED in the unit cell. Since the SFs are measured only for a discrete set of scattering vectors, ($\boldsymbol{H}$), the average ED is the inverse FT summation of the SFs

$$< \rho(\boldsymbol{r}) >= \sum_{h,k,l}^{\infty} F_{h,k,l} Exp[-2\pi i\,(hx + ky + lz)] \tag{1.4}$$

where $\boldsymbol{r} = x\boldsymbol{a} + y\boldsymbol{b} + z\boldsymbol{c}$. The unit cell is the smallest space element from which the crystal can be built by repeated translations along the lattice vectors ($\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$). The scattering vector, ($\boldsymbol{H}$), is a function of the volume of the unit cell, V, the lattice vectors, and the miller indices $\boldsymbol{H} = \frac{1}{V}\{h(\boldsymbol{b} \times \boldsymbol{c}) + k(\boldsymbol{c} \times \boldsymbol{a}) + l(\boldsymbol{a} \times \boldsymbol{b})\}$.

Unfortunately the average ED cannot fully be retrieved in this way for several reasons. First, the SFs are only *related* to the observed intensities by $I \propto F^2$. Other physical effects such as background radiation, absorption, extinction, and thermal diffuse scattering must be taken into account in order to extract the Bragg SFs from the intensity [13]. Because these phenomena are not trivial to explicitly account for, they manifest themselves as systematic experimental errors giving rise to uncertainties in the estimates of ED parameters [9, 14].

In addition, only a finite number of SFs are available from Bragg diffraction experiments. Furthermore, only the amplitudes of the SFs and not their phases can be measured. All of these problems (that the SFs carry more information than we would like, that we are attempting to perform a finite FT, and that we do not have the phases of the FT) mean that we will always be dealing with an incomplete data set which is why we *must* rely on a model to 'fill in the gaps' given that we are attempting to solve an ill-posed inverse problem.

1.2 Theoretical Electron Density from a Wave Function

In the absence of experimental data, an *ab initio* ED will be constructed to study its topological properties by finding the equilibrium geometry, $\boldsymbol{R}$, and accompanying electronic wave function, $\Psi(\{\boldsymbol{r_i}\}; \boldsymbol{R})$, so that the ED can be calculated at a point (Eqn. 1.1). The molecular structure defined by ρ containing N nuclei is parameterized by the location of the nuclei as defined in the previous section. Assuming that an equilibrium geometry has been found, the nuclei are represented as a $N \times 3$ vector $\boldsymbol{R}$ and the related ED can be written as $\rho(x, y, z)$ where it is understood that the nuclei are stationary.

A common practice, and the one implemented by the Gaussian [15] software package, is to describe [16] the many electron wave function, Ψ, as an antisymmetric linear combination of one-electron Molecular Orbitals (MO), Ψ_i, each of which is described by a linear combination of Atomic Orbitals (AO), ϕ_j.

Finally, each AO is expanded in Gaussian Type Orbitals (GTO), g_k, centered at a nucleus such that

$$g_k(x, y, z) = (x - X_k)^{n_k}(y - Y_k)^{m_k}(z - Z_k)^{l_k} Exp[-\alpha_k \,|\boldsymbol{r} - \boldsymbol{R}_k|^2] \qquad (1.5)$$

giving rise to the expressions for AOs and MOs as

$$\phi_j(\boldsymbol{r}) = \sum_k d_{jk} \, g_k(\boldsymbol{r}) \tag{1.6}$$

$$\Psi_i(\boldsymbol{r}) = \sum_j c_{ji} \, \phi_j(\boldsymbol{r}) \tag{1.7}$$

The final ED can be then be expressed as

$$\rho(\boldsymbol{r}) = \sum_{\mu,\nu} P_{\mu\nu} \, \phi_\mu(\boldsymbol{r})\phi_\nu(\boldsymbol{r}) \tag{1.8}$$

where $P_{\mu\nu}$ is the 'Density Matrix' described by $P_{\mu\nu} = 2\sum_k c_{\mu k}c_{\nu k}$ for closed shell systems. Combination of Eqns. (1.5, 1.6, 1.8) yield an expression of the ED involving products of GTOs which may have different centers. This expression will become important in Chapter 4 when taking a convolution of non-nucleus centered Gaussian functions as a result of products of GTOs.

The advantage of a mathematical description of the ED is that the theoretical aspects of the atom can be obtained using computer systems. Many aspects of chemical bonding including bond strength and bond order can be calculated by analysis of the ED as described in Eqn. 1.7. Even the atom can be defined by boundaries of atomic basins via calculation of points, $\boldsymbol{r}_s$, on S, the surface of zero-flux [10]

$$\nabla\rho(\boldsymbol{r}_s) \cdot \boldsymbol{n}(\boldsymbol{r}_s) = 0 \tag{1.9}$$

where $\boldsymbol{n}(\boldsymbol{r}_s)$ is the normal vector to a point on S.

The current discussion, and the Gaussian software package, make use of theoretical wave function based models, but Jayatilaka and others [17, 18, 19, 20] have been using wave function based models to fit experimental data to a desired ED formalism. This is not, however the only method of fitting an ED model to experimental data.

1.3 The Multipole Model of the Electron Density

In order to express the static ED as an analytical function, it is customary to introduce a model, M, which is a function of a set of parameters, $\boldsymbol{\alpha}$, and attempt to reconstruct the diffraction data, $\boldsymbol{D}$, (with standard uncertainties $\boldsymbol{\sigma}$) [4]. Having an analytic expression (a scattering model), all SF's can be predicted and compared with those observed from experiment. We wish to find the unbiased estimates for the parameters in the least-squares (LS) sense, so that the mean-square error

$$\chi^2 = \sum_i \frac{\left(D_i - M_i(\boldsymbol{\alpha})\right)^2}{\sigma_i^2} \tag{1.10}$$

is minimized.

The thermally averaged ED in the unit cell is taken to be the superposition of (thermally averaged) nucleus centered density units. This is equivalent to taking the convolution ($\circledast$) of the atomic densities with the delta functions at the location of each atom's center. In the following equations, a superscript of at is used to signify that the superscripted density is an atom centered density unit, while its absence signifies that it is not.

$$\rho_{UC}(\boldsymbol{r}) = \sum_j \rho_j^{at}(\boldsymbol{r} - \boldsymbol{R}_j) = \sum_j \rho_j^{at}(\boldsymbol{r}) \circledast \delta(\boldsymbol{r} - \boldsymbol{R}_j) \tag{1.11}$$

When the molecular structure is near its geometry of thermal equilibrium, the total molecular vibration can be described as a combination of atomic vibrations [1]. The motion of each atom can then be described by a harmonic oscillation, $p_j(\boldsymbol{r}')$, which defines the probability of finding atom j at a displacement of $\boldsymbol{r}'$ from its equilibrium position $\boldsymbol{R}_j$. These

Gaussian PDFs aid in the representation of the thermally averaged atomic ED by means of a convolution between them and the static atomic ED:

$$\rho_j^{at}(\boldsymbol{r}) = \rho_j^{at,static}(\boldsymbol{r}) \circledast p_j(\boldsymbol{r}) \tag{1.12}$$

Considering the way the SFs relate to the ED as well as the fact that the FT of a convolution is a product of FTs, it follows that

$$F(\boldsymbol{H}) = FT[\rho_{UC}(\boldsymbol{r})] = \sum_j f_j(\boldsymbol{H})T_j(\boldsymbol{H})Exp[-2\pi i\, \boldsymbol{H}.\boldsymbol{R}_j] \tag{1.13}$$

where $f_j(\boldsymbol{H})$ is the 'atomic structure factor' and $T_j(\boldsymbol{H})$ is called the 'atomic temperature factor' or Debye-Waller factor. These forms represent the PDFs in Fourier space:

$$f_j(\boldsymbol{H}) = \int \rho_j^{at,static}(\boldsymbol{r})Exp[2\pi i\boldsymbol{H}.\boldsymbol{r}]d\boldsymbol{r} \tag{1.14}$$

$$T_j(\boldsymbol{H}) = \int p_j(\boldsymbol{r})\, Exp[2\pi i\boldsymbol{H}.\boldsymbol{r}]d\boldsymbol{r} \tag{1.15}$$

We are thus able to calculate the SFs, $F_{calc}(\boldsymbol{H})$, relating to the model which is determined to fit the observed SFs, $F_{obs}(\boldsymbol{H})$, according to the parameters defined in the model and the least squares problem becomes $\sum_{\boldsymbol{H}}\{w_{\boldsymbol{H}}\,(|F_{calc}(\boldsymbol{H})| - |F_{obs}(\boldsymbol{H})|\,)^2\}$ where the weighting factor is determined by the uncertainty of the intensity data [21].

When we assume that the model to be used has been predetermined, the problem turns into finding the 'unbiased' estimates of the parameters for the model which best fit the data. Some of these parameters include locations of nuclei and basis functions, basis function populations, and those determining the size and shape of the temperature factors.

The formalism we choose is called the "Pseudoatom" (PA) model [6, 22, 23]. The ED is defined to be a superposition of nucleus centered basis functions over the domain of real (density normalized) spherical harmonics.

$$\rho(\boldsymbol{r}) = \sum_j \rho^{(j)}(\boldsymbol{r} - \boldsymbol{R}_j) \tag{1.16}$$

$$\rho^{(j)}(\boldsymbol{r}) = \sum_{l=0}^{L} \sum_{m=-l}^{l} P_{lm}\rho_{lm}(r)y_{lm}(\theta, \phi) \tag{1.17}$$

Each atomic density, $\rho^{(j)}(\boldsymbol{r})$, is expanded over real spherical harmonics, $y_{lm}(\theta, \phi)$ expressed in nucleus-centered local frames. The exponentially decaying radial part, $\rho_{lm}(r)$, is normalized so that the 'populations', of the (l, m) multipole is contained in the P_{lm} term. Chapter 2 explores our contributions to the expression of the PA model and quantifiable results for improving accuracy of model.

1.4 The Convolution Approximation

The key property for interpreting X-Ray Bragg scattering is the coherent elastic SF, the Fourier Transform of the thermal average (T-dependent canonical ensemble average over vibration states) of the crystalline ED (referred to here as the dynamic ED):

$$\langle\rho\rangle_T = \sum_n w_n\rho_n \tag{1.18}$$

where w_n and ρ_n are the Boltzmann factor and ED, respectively, associated with the nth state of the system in thermal equilibrium with its surroundings. To reduce this expression to a closed analytic form applicable of SF data fitting, the harmonic convolution approximation is invoked which includes a hierarchy of approximations [24]: (a) the states accessible by the system (molecule or crystal) are restricted to vibration states (no

electronic transition occurs during the scattering); (b) the adiabatic approximation that implicitly assumes the physical observability of the stationary ED corresponding to the equilibrium nuclear geometry (represented by a $3N$-row vector $\boldsymbol{R}^0$ for N nuclei); (c) the stationary ED is supposed to be expressed as a superposition of partial distributions each assigned to a specific center; (d) even less feasibly, each density unit is defined to rigidly follow the motion of its center as the nuclei vibrate about their equilibrium positions in a harmonic potential.

The normal mode analysis of the system of N vibrating nuclei in thermal equilibrium leads to a $3N$-multivariate normal distribution of nuclear displacements relative to their equilibrium positions ($\boldsymbol{u} = \boldsymbol{R} - \boldsymbol{R}^0$):

$$P(\boldsymbol{u}) = (2\pi)^{-\frac{3N}{2}}|\boldsymbol{U}|^{-\frac{1}{2}}Exp\left[-\frac{1}{2}\boldsymbol{u}\boldsymbol{U}^{-1}\boldsymbol{u}^T\right] \qquad (1.19)$$

where $\boldsymbol{u}^T$ represents the transpose of $\boldsymbol{u}$ and the covariance matrix is the Mean Square Displacement Amplitude (MSDA) matrix ($\boldsymbol{U}$) associated with the expectation values of Cartesian nuclear displacement products (second moments):

$$\boldsymbol{U} = \langle \boldsymbol{u}^T\boldsymbol{u} \rangle_T \qquad (1.20)$$

The temperature dependence of $\boldsymbol{U}$ is embedded in the MSDAs of the normal modes, in its eigenvalues ($\boldsymbol{\delta} = (\delta_{i=1,3N})$:

$$\boldsymbol{U} = \boldsymbol{L}\boldsymbol{\delta}\boldsymbol{L}^T \qquad (1.21)$$

$$\delta_j = \frac{h}{8\pi^2 v_j}\coth\left(\frac{hv_j}{2k_B T}\right) \qquad (1.22)$$

where ν_j is the frequency of the jth normal mode [25, 26].

In the simplest versions of the Bragg SF formalism, that relies on the one-center ED models (such as the conventional isolated-atom or the pseudoatom; nucleus-centered multipole expansion [22, 23], the correlations between nuclear vibrations are ignored. In other words, the density units centered on a given nucleus ($\rho_a(\boldsymbol{r_a}), \boldsymbol{r_a} = \boldsymbol{r} - \boldsymbol{R_a}$) are smeared (convoluted) by a trivariate normal distribution of that nuclear center (marginal PDF: P_a). This means that the $3N$ multivariable normal distribution is taken as a product of N trivariate normal distributions, giving rise to the smeared ED

$$\langle \rho(\boldsymbol{r}) \rangle_T = \sum_a \int_{-\infty}^{\infty} \rho_a(\boldsymbol{r_a} - \boldsymbol{u}) P_a(\boldsymbol{u}) d\boldsymbol{u} \tag{1.23}$$

Note that the Fourier Transform of a marginal PDF (P_a) is the Debye-Waller (DW) factor routinely used in scattering models to dampen the scattering power of an atom due to nuclear vibrations [27, 28]. If the convolution approximation is to be applied to the ED within the LCAO-MO approach, new considerations arise because it is not immediately apparent how to calculate the ADPs of non-nuclear sites at which two-center orbital products are centered. Approximations of practical relevance have been suggested previously to estimate the variance / DW factor of these centers as a linear combination of experimental nuclear ADPs [29] / DW factors [30, 31, 32], but without providing the exact solution given below.

1.5 Expression of Nuclear Motion

Adapting Eqn. 1.14 to each nucleus, the atomic smearing function (as it appears in Eqn. 1.18) is

$$P_j(\boldsymbol{u}_j) = (2\pi)^{-\frac{3}{2}}|\boldsymbol{U}_j|^{-\frac{1}{2}}Exp\left[-\frac{1}{2}\boldsymbol{u}_j\boldsymbol{U}_j^{-1}\boldsymbol{u}_j^T\right] \tag{1.24}$$

where $\boldsymbol{u}_j$ is the displacement of the jth nucleus from its equilibrium position and $\boldsymbol{U}_j$ is the symmetric 3x3 symmetric tensor which represents the variance-covariance of the nuclear PDF.

Since during an XRD experiment, atoms in molecules as well as molecules in the crystal visit different vibration states, the scattering intensity is related to the thermal average (canonical ensemble average) of the ED. The vibration states depend on the internal (intra-molecular) and external (intermolecular) forces. The ultimate goal is to derive the static ED, since in the QTAIM treatment of atoms in molecules, TPs of the static ED are used to reveal important chemical features of the molecular systems.

Upon proper deconvolution of thermal motion from the observed reflexions, the static ED is modeled using the equilibrium structure which is believed to be the space-time average of all structures observed in the crystal sample. A natural, though naive, conclusion might be that the TPs of the static model are also representative of the average of TPs from all structures in the ensemble. In this computational study, we find that this is not the case and show how the TPs of the structure changes as the molecule undergoes thermal vibrations.

One of the most widely used computer program packages for fitting SFs is 'XD2006' [33] which allows the user to refine, among others, the following parameters:

(1) The populations, P_{lm}, for each multipole in the expansions of the atomic densities.

(2) The Anisotropic Displacement Parameters (ADP), which make up the U_j matrix used in the temperature factors. These ADPs represent the likelihood for the nucleus to be displaced in given directions. These symmetric 3x3 matrices are the variance-covariance in $u_j = (x_j, y_j, z_j)$ for each nucleus j.

Once the parameter estimates have been found for a given set of diffraction data, we can use the fitted Model Density to calculate topological properties (TP) of the ED. A topological analysis of the ED includes an analysis of both scalar and non-scalar properties of the ED. Some of the scalar properties may include the value of the ED, $\rho(r)$, or it's Laplacian, $\nabla^2 \rho(r)$ at certain points while some of the non-scalar properties may include the location of Bond Critical Points (BCP) as well as other integrated properties.

The importance of the static ED is paramount because it reveals many important properties of the system in question. The concept of the atom in a molecule, while used throughout history, has only recently been explored with charge density studies and relies heavily upon the topology of the ED. Even physical properties which many chemists take for granted, such as bond types and bond strengths, can be examined by studying the topology of the ED of a molecule. Not only are static properties like bond order revealed through an examination of the ED, but reactive properties can be explored by the static ED as well. The Laplacian of the ED, for example, reveals a charge concentration or depletion depending on the sign. When a molecule with a site of charge depletion meets a molecule with a site that has charge concentration, they are attracted to one another and a reaction

ensues. For this reason it is incredibly important to be able to accurately assess the uncertainty with which we predict topological properties of the ED.

Chapter 3 will explore the effects of nuclear motion on the ED by statistically analyzing the way in which scalar and non-scalar properties are affected by perturbations to the nuclear positions resulting from thermal motion. It is important to note that this analysis does not invoke the convolution approximation. Chapter 4 then uses these results to assess the accuracy and validity of the convolution approximation by comparing dynamic ED resulting from statistical treatment of motion to the dynamic ED resulting from a treatment of motion by convolution.

1.6 Uncertainty Quantification of Topological Properties

We fit the model parameters to the data by taking a first order Taylor Series expansion to the model with respect to each parameter and perform a Linear Least Squares estimate. If the parameters are non-linear, as is the case with the ADPs, then we can do this repeatedly - "refining" the parameters until convergence is attained [21]. By doing this, we can easily construct the variance-covariance matrix (VCM) of the parameter estimates so that we may quantify their uncertainty.

The parameter uncertainties will also cause uncertainties in the evaluated TPs so we need to quantify the uncertainty in the TPs due to the uncertainty in the model parameters. It is a rather trivial process to quantify the uncertainty in scalar properties (ρ, $\nabla^2 \rho$) of the ED due to linear parameters (p_{lm}) by means of Error Propagation [34]. It is not simple, however, simple to find the uncertainty in non-scalar 'properties' of the ED such as location of a BCP between two atoms.

The BCP between two atoms is the point along the bond path (the path of maximum density connecting two nuclei) where the gradient of the ED vanishes and the Hessian has two negative and one positive eigenvalues [10]. A small variation in any of the model parameters will result in a minor deviation of the calculated location of the BCP from that of the established parameters. Unfortunately, because finding a BCP is an iterative, non-linear process, it is difficult if not impossible to find the uncertainty in the location of the BCP due to the uncertainty in the model parameters using methods of error propagation.

There is an even more subtle problem - the way in which we treat (or don't treat) the uncertainty in dynamic parameters in the PA model. When we are finding properties of the ED, it is implicitly understood that we are speaking about properties of the static ED. Recall that the dynamic parameters (the ADPs) are used only to deconvolute the smearing in the data which results from thermal motion. After we have removed this smearing and are working with the *static* ED, the variance / covariance of these dynamic parameters are ignored. For this reason, the VCM which we use in error propagation does not contain any of the ADPs, or at least, those parts of the VCM are not used because the ADPs do not show up in the static model. Indeed, the ADPs are used explicitly to deconvolute the thermal motion in the refinement process and are ignored completely when dealing with the properties of the static density.

The relationship between the ADPs and the static model parameters, however, cannot be ignored. In fact, if the ADPs are held constant during the refinement process, the effect would be seen in other parameters which would 'mimic' thermal smearing. Likewise, the ADPs alone can be used, at least in part, to account for the static ED asphericity which shows the interdependence and correlations which can be "as high as 80-

90%" [35], especially when moving beyond the Harmonic Approximation and into the Gram-Charlier expansion [36].

Chapter 5 explores these issues and assesses the effect that the uncertainty in the ADPs has on the uncertainty in the static ED. This is done via a mixture of error propagation and stochastic sampling in such a way that it is possible to compare the uncertainties of the ED resulting from both including and ignoring the uncertainties in the ADPs.

2. A New Approach to the Pseudoatom Expansion

2.1 The Standard Pseudoatom Expansion

Within the basic PA Multipole Model, the total molecular ED, $\rho_{tot}(\boldsymbol{r})$, is expressed as a sum of nucleus centered density units,

$$\rho_{tot}(\boldsymbol{r}) = \sum_a \rho_a(\boldsymbol{r} - \boldsymbol{R}_a) \tag{2.1}$$

$$\rho_a(\boldsymbol{r}) = \sum_{l=0}^{Lmax} \sum_{m=-l}^{l} \rho_{lm}(r)d_{lm}(\theta,\phi) \tag{2.2}$$

where d_{lm} are density normalized Real Spherical Harmonics (RSH) [37] (the normalization coefficients for which have been precalculated up to $l = 40$ [38]). Because the angular basis functions are predetermined, the only part of the expansion left to find are the Radial Density Functions (RDF), $\rho_{lm}(r)$.

A widely used version of the PA expansion is the "Hansen-Coppens" (HC) formalism [23]. In this expansion, the radial term is m-independent so the PA density for each atom (which we will denote as ρ rather than ρ_a) is

$$\rho(\boldsymbol{r}) = \sum_{l=0}^{Lmax} R_l(\kappa_l r) \sum_{m=-l}^{l} P_{lm} d_{lm}(\theta,\phi) \tag{2.3}$$

where R_l is normalized so that P_{lm} defines the multipole population.

The monopole ($l = 0$) RDF is defined differently than for the higher poles ($l > 0$). For the monopole, the RDF is defined as the sum of a 'core' RDF (which does not scale with κ) and a 'valence' term (which does scale with κ) such that $R_0(\kappa_0 r) = \rho_c(r) + P_{00}\rho(\kappa_0 r)$. Both the core and the valence monopole radial functions are obtained from the Hartree-Fock (HF) wave function of the ground-state isolated atom [4, 39].

According to the HC formalism, the higher order poles ($l > 0$) have RDFs which are simple "Slater functions" [40]

$$R_l(\kappa_l r) = \frac{(\alpha_l \kappa_l)^{n_l+3}}{(n_l + 2)!} r^{n_l} Exp[-\alpha_l \kappa_l r] \tag{2.4}$$

leading to an 'm-independent' (MI) Multipole Model; the RDFs of the higher-order poles are dependent only upon l but not on m. The standard HC expansion is typically terminated at the hexadecapolar level ($Lmax = 4$).

As discussed in [6], it is advantageous to build upon the HC model by including m-dependent (MD) terms. The referenced work performs an analysis with the α-glycine molecule using numerical and analytical RDFs which involve a time consuming genetic evolutionary algorithm for fitting the analytical RDFs (aRDF) to the derived numerical RDFs (nRDF). Currently, we perform the same type of analysis on the formamide molecule, showing the advantages of the convergent m-dependent expansion, but using a new, computationally efficient, fitting protocol.

2.2 The Improved Expansion

The ways in which we seek to improve upon the HC formalism is three fold.

(1) We seek to increase the expansion from the often used $Lmax = 4$ to any arbitrary number. Currently the analysis up to $Lmax = 16$ shows promising results, though that level is not feasible in treatment of XRD data.

(2) We use MD RDFs which involves a unique RDF for each combination of (l, m) of a given atom.

(3) We implement a higher sized basis set for the higher order poles. That is, instead of using a single Slater function for $l > 0$, the size of the basis set is increased to 10.

The final expression of our aRDFs takes the following form:

$$R_{lm}(r) = N_{lm} \sum_{i=0}^{M} d_i S_i(r) \tag{2.5}$$

$$S_i(r) = \sum_{j=0}^{M} C_{i,j} r^{j+l} Exp[-\alpha_{lm} r] \tag{2.6}$$

The combination of (2.5) and (2.6) yields the following comparison of the new vs. old formulations of the analytic RDFs:

$$R_{lm}(r) = N_{lm} \sum_{i=0}^{M} c_i' r^{i+l} e^{-\alpha_{lm} r} \tag{new}$$

$$R_{lm}(r) = N_{lm} r^{n_l} e^{-\alpha_l r} \tag{old}$$

Eqns. 2.5 and 2.6 lead to an efficient optimization since both d_i and C_{ij} are uniquely determined by a choice of α_{lm} - meaning that the optimization is in one dimension. For many cases, the choice of α_{lm} can even be predetermined and the same level of accuracy can be achieved simply by extension of the basis set (M). This does, however, increase the computational complexity of the evaluation of topological properties.

2.3 Numerical Radial Functions

To test this formalism, the program 'Radfun' [41] creates numerical RDFs (nRDF) which numerically projects a target *ab-initio* ED (ρ_Ψ) onto RSHs to create a grid based radial density by making use of the Stockholder partitioning scheme [42]:

$$\rho_a^{stock}(r) = w(r)\rho_\Psi(r) = \frac{\rho_a^0(r)}{\sum_b \rho_b^0(r)}\rho_\Psi(r) \tag{2.7}$$

where $\rho_a^0(r)$ is the isolated atomic density (spherical, since it is not affected by bonding). The importance of investigating the extent to which we can reconstruct a target ED with nRDFs is paramount as it directly shows the error in the resulting aRDFs due to a finite truncation of $Lmax$, as opposed to the error introduced in an imperfect fit.

Projection of the target ED into RSHs takes place in two steps as outlined in [6]. First, the fuzzy partitioning of the Stockholder atoms are determined as defined in Eqn. 2.7. Second, the fuzzy atom is projected onto RSH yielding a radial density function of the distance from the nucleus, r_a

$$R_{lm}^{(a)}(r_a) = \frac{\int \rho_a^{stock}(r)d_{lm}(\mathbf{\Omega}_a)d\mathbf{\Omega}_a}{\int \rho_a^{stock}(r)d_{lm}(\mathbf{\Omega}_a)d\mathbf{r}_a} \tag{2.8}$$

The projection onto RSH is performed numerically in the Radfun application. The results are MD nRDFs defined on a fine grid, (r_i, χ_i), which can perfectly reproduce the target ED, ρ_Ψ, given a high enough expansion in $Lmax$.

2.4 Analytical Radial Functions

The process of fitting the aRDFs is performed in a program called 'eFit' which fits a single aRDF (Eqn. 2.5) to each nRDF for each combination of (at, l, m) as demonstrated in Figure 2.1. The procedure makes use of an orthonormal set of functions $\{S_i\}$ as described

in Eqn. 2.6 where, for a given α_{lm}, the C_{ij} can be chosen such that $\langle S_i|S_j\rangle = \delta_{ij}$. This is accomplished by diagonalizing the overlap matrix $G = \langle g_i|g_j\rangle$ where g_i is a primitive Slater function. Upon successful creation of $\{S_i\}$, the mixing coefficients are found by projecting the nRDFs onto the orthonormal functions so that $d_i = \langle S_i|\chi\rangle$. This, again, uniquely defines d_i as a function of α_{lm} resulting in a one variable search for optimization.
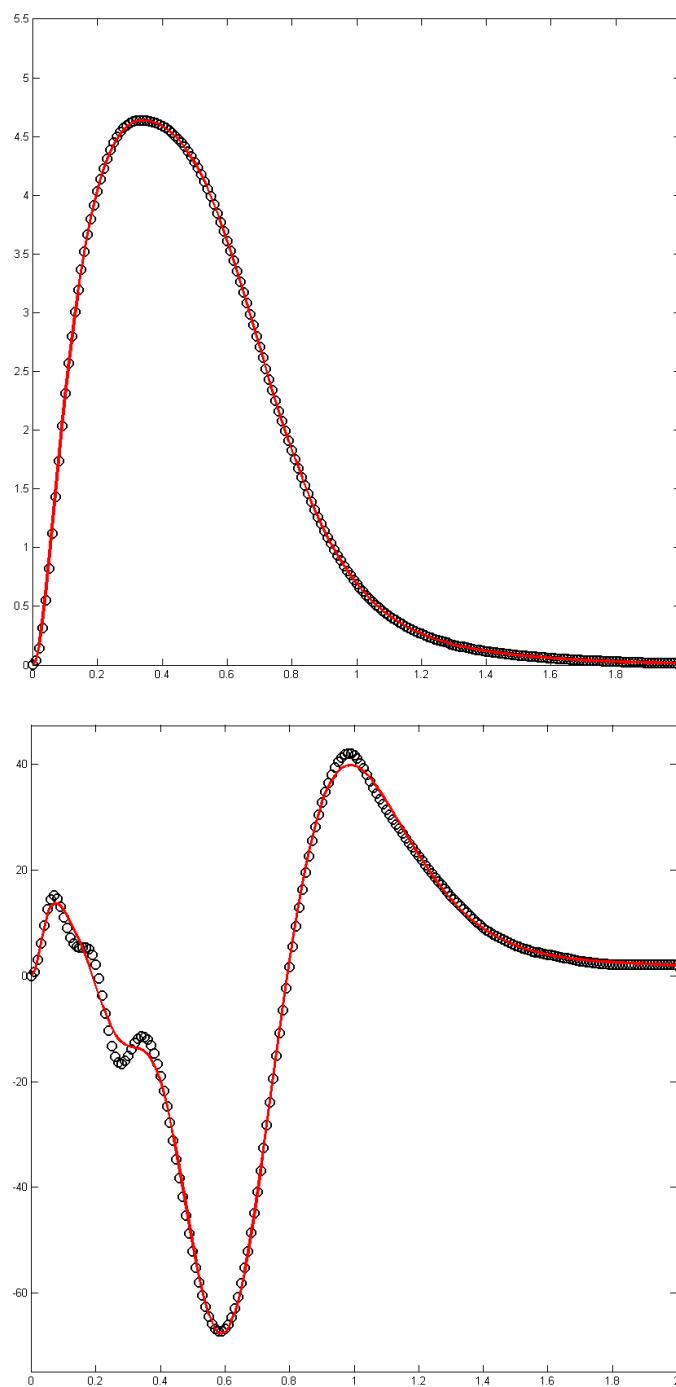
Figure 2.1 nRDFs on a fine grid (black, open circles) with fitted aRDFs (solid red lines) for the carbon atom with $(l, m) = (2,0)$. (a, top) the ED ($e/\text{Å}^3$) and (b,bottom) the Laplacian ($e/\text{Å}^5$) along a radial grid (Å).

This search is performed by sampling a reasonable number of points (50-100) in α and then successive parabolic interpolation in $(\alpha, Cost)$ is performed where

$$Cost = \beta \frac{\langle \rho | \chi \rangle}{\langle \chi \rangle} + (1 - \beta) \frac{\langle \nabla^2 \rho | \nabla^2 \chi \rangle}{\langle \nabla^2 \chi \rangle} \tag{2.9}$$

giving the user the option of fitting the Laplacian as well as the actual ED by input parameter β. This search is performed, independently, for each combination of (at, l, m) which lends the program to virtually linear speedup for parallelism as Figure 2.2 shows.
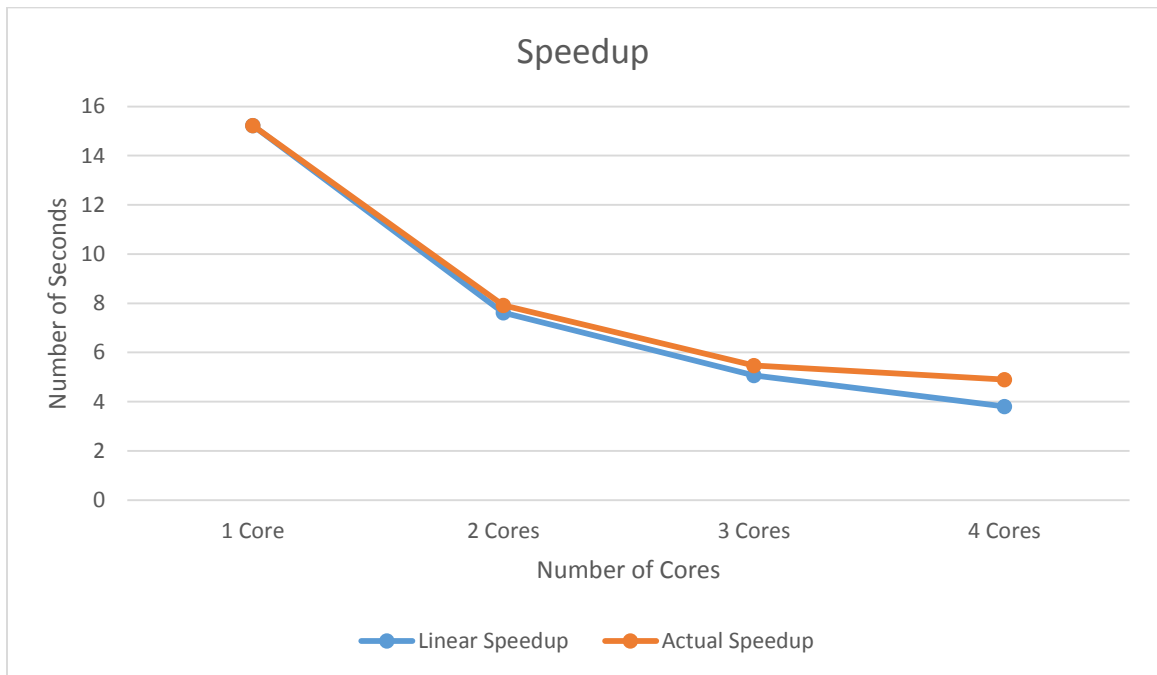
Figure 2.2 Actual speedup for the eFit application on 4 cores plotted against a perfect linear speedup for comparison.

2.5 Accuracy of Formamide Reconstruction

The PA formalisms, thus far, have been described as either nRDFs or aRDFs, the latter of which is also broken down into MD and MI types. Note that the nRDFs are MD and all forms discussed rely on the *Lmax* parameter. As such, the PA type densities will be referred to as ρ_L while the target ED, an *ab-initio* ED of formamide generated at the B3LYP/6-311G** [43] level of theory in Gaussian09 [15], will be referred to as ρ_Ψ. The purpose is to project ρ_Ψ onto ρ_L (with either nRDFs or aRDFs) to quantitatively estimate the error related to the one-center PA representations of the ED.

To simulate the HC model, the basis size for the monopole is 20 and only 1 Slater function for the higher order-poles, indicated by the expression MI-aRDF(20,1) whereas the MD formulation will use 10 for higher order poles, indicated by MD-aRDF(20,10). The higher order RDFs of the MI model are constructed by Principal Component Analysis of the MD functions, yielding the 'best possible' reconstruction of the target density while keeping with the MI RDFs and small (1 Slater) basis functions of higher order terms.

2.5.1 Topological Properties at Bond Critical Points

The ability for the nRDF to accurately reconstruct the target ED, ρ_Ψ, is apparent from Table 2.1 which reports the percent error in reconstructing the ED and Laplacian at the BCPs. These numbers can be compared with those obtained for the $C=O$ BCP by the MI-aRDF(20,1) model which exhibits a 4.923% and 40.643% error for the ED and Laplacian, respectively, at *Lmax* = 16. The situation is worse at the standard cutoff value of *Lmax* = 4 with ED and Laplacian errors of 3.963% and 110.414%. Even in the case of the MD-aRDF(20,10) model, the error due to fitting is noticeable. While all other BCPs exhibit no higher than a 2.66% error (for either ρ or $\nabla^2\rho$) at *Lmax* = 16, the $C=O$ BCP for

the MD-aRDF(20,10) model shows ED and Laplacian errors of 0.177% and 30.292% respectively.

Table 2.1 Percent error of $\rho_\Psi - \rho_L$ (top) and $\nabla^2\rho_\Psi - \nabla^2\rho_L$ (bottom) at the given BCPs where ρ_L represents the nRDF construction up to $Lmax$.

	Lmax = 0	Lmax = 4	Lmax = 8	Lmax = 12	Lmax = 16
$H1$-$C2$	30.556	0.326	0.037	0.027	0.006
	87.034	3.263	0.357	0.659	0.120
$C2$-$N3$	16.662	0.024	0.208	0.037	0.006
	81.520	6.670	2.612	0.992	0.269
$C2$=$O6$	18.602	0.563	0.328	0.062	0.002
	356.727	112.137	49.102	14.025	0.426
$N3$-$H4$	22.370	1.152	0.170	0.025	0.000
	78.905	12.073	2.541	0.519	0.114
$N3$-$H5$	22.211	1.138	0.176	0.025	0.001
	78.401	12.607	2.643	0.540	0.101

2.5.2 Topological Properties Along the Bond Path

The fitting protocol minimizes the Mean Square Residual (MSR) of the nRDFs to the fitted aRDFs. A side effect of is that the fitted aRDFs tend to oscillate about the related nRDFs in the Laplacian as Figure 2.3 shows. The Laplacian is evaluated along the bond profile of C=O (top) and C-N (bottom) for ρ_Ψ, nRDF, MD-aRDF(20,10), and MI-aRDF(20,1) and the three PA approximations are compared to the target ED, $\nabla^2\rho_\Psi - \nabla^2\rho_L$.
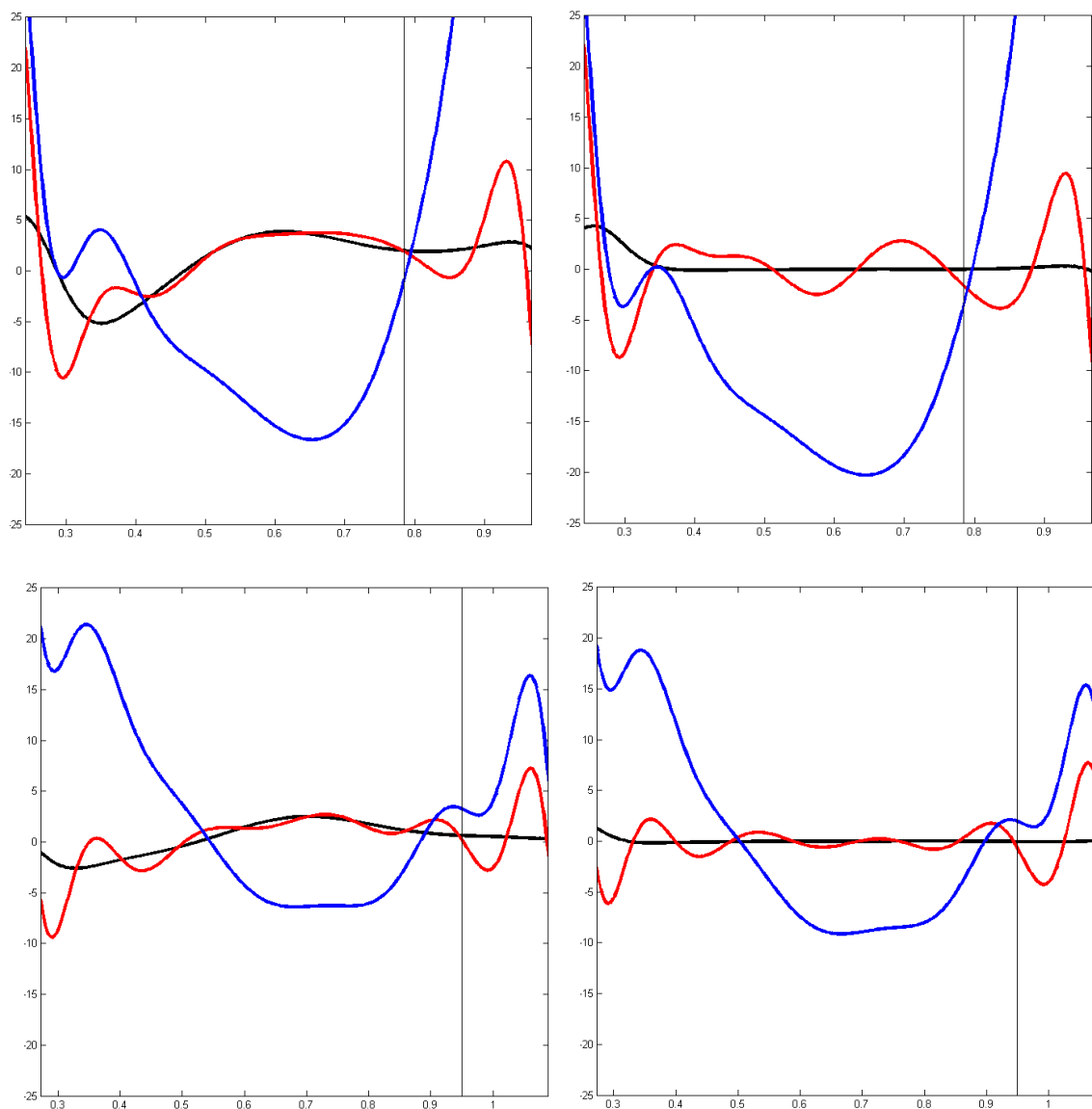
Figure 2.3 Errors in reconstructing $\nabla^2 \rho(\boldsymbol{r})$ ($e/\text{Å}^5$) in the profile of $C{=}O$ (top) and $C{-}N$ (bottom). $\nabla^2 \rho_\Psi - \nabla^2 \rho_L$ is shown with *Lmax* = 4 (left) and *Lmax* = 16 (right) for the nRDF (black), MD-aRDF(20,10) (red), and MI-aRDF(20,1) (blue). The black vertical line shows where the BCP is in relation to the profile.

2.5.3 Topological Properties in the Plane

To show that the non-convergent nature of the MI model is not simply due to a smaller basis set for higher order poles, we also analyze the MI-aRDF(20,10) model (with 10 Slaters for higher order poles) so that it is the exact same as the MD-aRDF(20,10) model in all but the dependence upon m. The difference ED maps, $\rho_\Psi - \rho_L$, are shown in Figure 2.4 for varying values of *Lmax* and different forms of ρ_L. Figures 2.4 (a) and (b), which display the errors in MI-aRDF(20,1) and MI-aRDF(20,10) respectively, show very little improvement as *Lmax* increases from left to right, indicating that there is no convergence. Furthermore, while there is qualitative differences between Fig. 2.4a and Fig. 2.4b, the MI-aRDF model with 10 Slaters (a) does not show significant improvement over the single Slater (b) model.

The MD-aRDF model, also compared to ρ_Ψ in Fig. 2.4c, does show signs of improvement as Lmax increases. The largest differences are near to the nuclei which is partly because the ED has extremely sharp peaks at the nucleus, and also because the aRDF model is using Slater functions to model a Gaussian peak of the target ED. Finally, the nRDF comparison (Fig. 2.4d) shows a very accurate approximation to the target ED, indicating that the m-dependent multipole expansion is truly convergent.
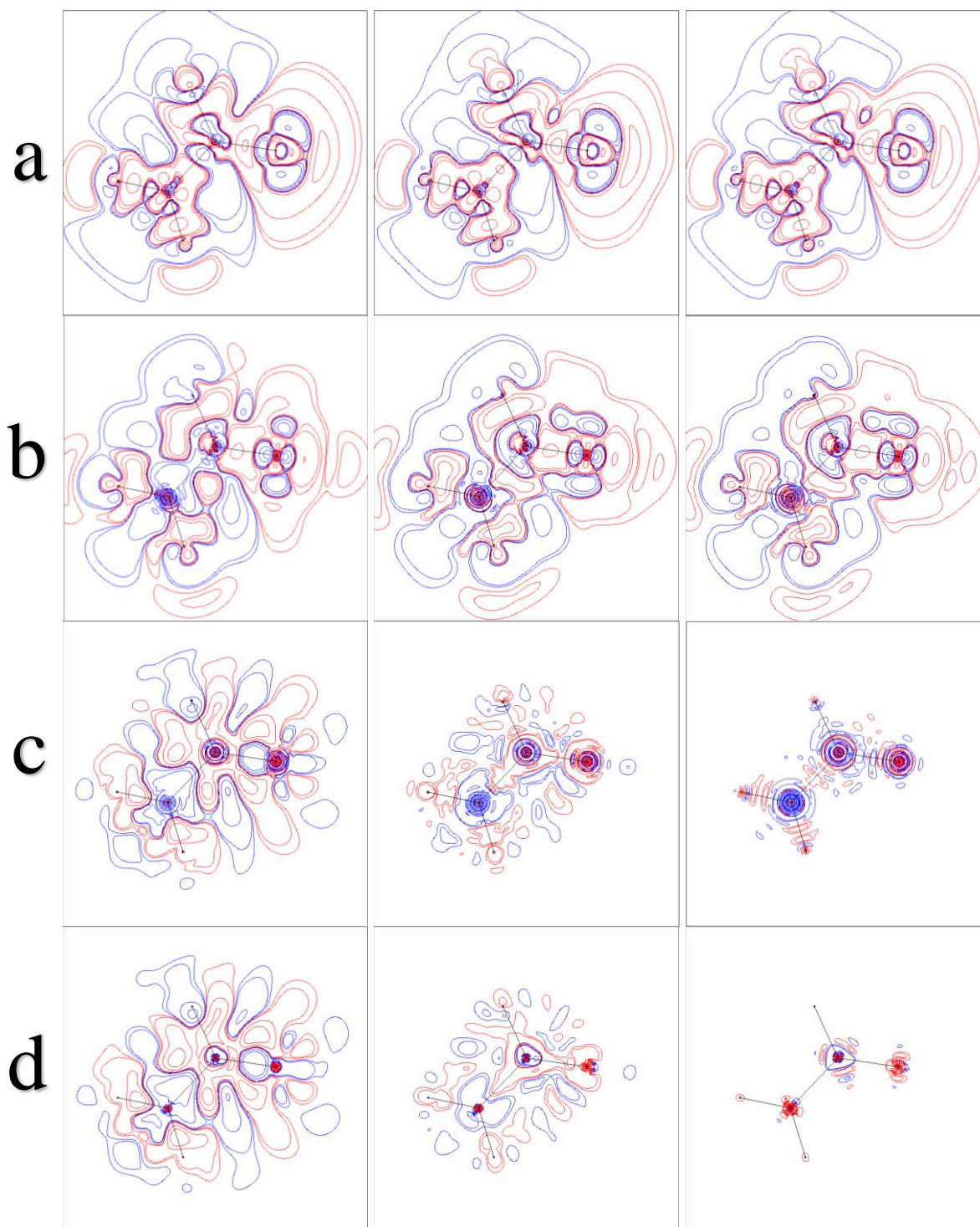
Figure 2.4 The convergence of $\rho_\Psi - \rho_L$ in the plane with Lmax values are (4,8,16) from left to right (a) MI-aRDF with 1 Slater for higher order poles (b) MI-aRDF with 10 Slaters for higher order poles (c) MD-aRDF with 10 Slaters for higher order poles (d) MD-nRDF. Contours are $\pm\{1,2\} \times \{10^{-1}, 10^{-2}, 10^{-3}\}$ $e/\text{Å}^3$ with positive / negative contours in red / blue.

3. Statistical Treatment of Dynamic Topological Properties

3.1 Relationships Between Topological Properties and Bond Lengths

Analysis of the relationship between TPs at the BCPs and the bond length has been studied in detail by Gibbs and others [44, 45, 46, 47] based on various molecules containing chemically analogous bonds (formed by the same pair of atoms but exhibiting different bond lengths). It has been found that the average ED at a BCP and the bond length are related via the power-law equation $R = 1.60 \times \langle \rho(r_c) \rangle^{-0.23}$ [45]. This analysis is extended in the current work by comparing the bond lengths and associated TPs at r_c for a specific molecule undergoing nuclear vibrations rather than for a set of molecules, each at equilibrium position.

For smearing a two-centered ED (see Eqn. 1.8), it becomes necessary to describe the motion of not only the nuclei, but the off-nucleus centered basis function products as well (a GTO product is a GTO centered between the nuclei). Scheringer et. al. [48, 29] have proposed a method for approximating the ADPs of inter-nuclear density units. We build on their model by applying an adapted method to represent the motion of not only inter-nuclear density units, but of the BCPs themselves.

3.2 Objectives

In this numerical experiment, we generate a large ensemble of wave functions, consistent with harmonic vibrations, for the formamide molecule. Performing a topological analysis on each member of the ensemble allows for a statistical treatment of thermally averaged topological properties. Of particular interest in this experiment are the

Topological Properties (TP) of the ED at the BCPs and the motion of the BCPs as they are "carried by the ED" while the nuclei are in thermal motion.

3.3 Methods

Performing an optimization of formamide (CH_3NO) using Gaussian09 [15] leads to the static representation of the formamide molecule, whose molecular structure is shown in Figure 3.1, at the B3LYP/6-311G** level of theory. Normal mode analysis [49] of the molecule at the equilibrium geometry allows for the creation of the MSDA matrix (Section 1.4) which can be represented, in Cartesian Space, as a $3N \times 3N$ matrix or, in internal coordinates, as a $(3N - 6) \times (3N - 6)$ diagonal matrix. The block diagonal elements of the Cartesian MSDA represent the Anisotropic Displacement Parameters (ADP) for each nucleus and can be used to define probability ellipsoids such as those used in the ORTEP program [50]. By invoking the harmonic approximation, we generate a large ensemble of 500,000 nuclear geometries which are consistent with thermal motion.
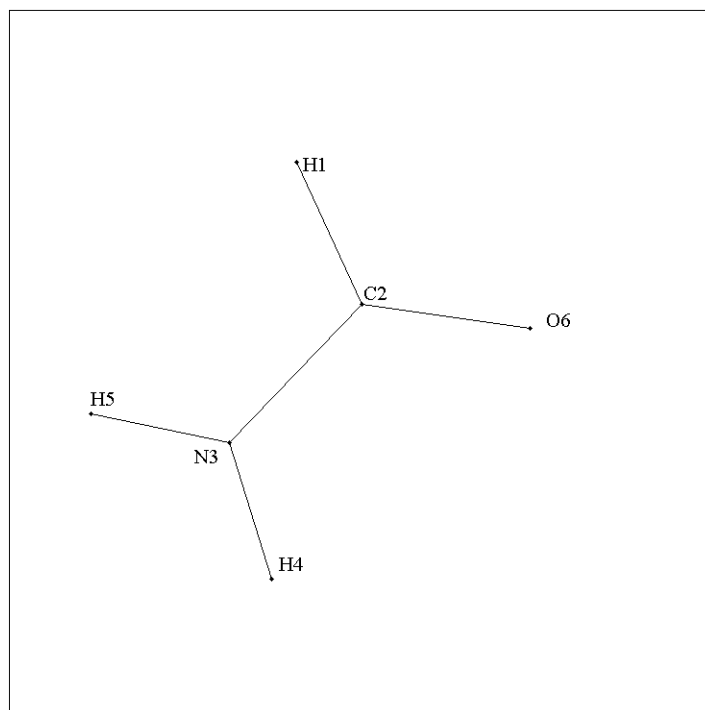
Figure 3.1 Molecular geometry of formamide

Each of these geometries is expressed as linear combinations of nuclear displacements added to the equilibrium geometry. For each normal mode of vibration, we have the unit displacements in Cartesian coordinates as well as the amplitude of displacement (which we regard as the standard deviation of a normal distribution). To create one of these perturbed geometries X_i (which is to be considered as one of many members in the ensemble $\{X_i\}$ exemplifying harmonic oscillations about the equilibrium geometry, X_{eq}, such that $\mu(\{X_i\}) = X_{eq}$, $\sigma^2(\{X_i\}) = U$), the normal modes can be sampled independently.

The perturbed geometry, X_i, is the sum of independent displacements from the equilibrium geometry by $3N - 6$ normal modes, each of which has a normalized vector of displacement and related mean square amplitude, (d_j, σ_j^2). The perturbed geometry can

then be expressed as a sum over these normal modes, $X_i = X_{eq} + \sum_{j=1}^{3N-6} r_\sigma^{(j)} d_j$ where $r_\sigma^{(j)}$ is a random variable coming from the Normal distribution $N(0, \sigma_j)$. By doing this for each normal mode, the perturbed geometry can be regarded as coming from a large sample which is normally distributed about X_{eq}. This technique is employed to create each geometry yielding the large ensemble which is consistent with harmonic vibrations.

For each perturbed molecular geometry, a wave function is calculated in Gaussian09 (without optimization) and a thorough topological analysis is performed with Denprop [51]. By taking the covariance of the large ensemble of BCPs, the ADPs for each BCP is found and a method is described for efficiently approximating ADPs of BCPs by using only the information from its bonding atoms. It is shown that the displacements of each BCP are normally distributed about the equilibrium position of each BCP, but that the distribution of TPs are not normally distributed about the TPs of the molecule in its equilibrium position.

3.4 Results and Discussion

3.4.1 Validation

To prove that the geometries are truly representative of harmonic vibrations, it is sufficient to show that the mean of the ensemble returns the equilibrium geometry and that the covariance of all geometries return the MSDA calculated from normal mode analysis. Table 3.1 shows the distance between the locations of the thermally averaged nuclei and the locations of the equilibrium positions.

Table 3.1 Distance from equilibrium (Eq) and average (Avg) nuclear positions.

Nucleus	$\|X_{Eq} - X_{Avg}\|$ in Å
$H1$	4.462e-04
$C2$	2.939e-05
$N3$	3.369e-05
$H4$	3.671e-04
$H5$	1.707e-04
$O6$	2.886e-05

The locations of the thermally averaged nuclei are found by taking the algebraic mean of the ensemble. By taking the standard deviation of this large dataset it is possible to show that the distribution of atomic displacements indeed follow a normal distribution. Figure 3.2 displays the data plotted along with the fitted Gaussian distributions for X,Y,Z coordinates of the $N3$ and $O6$ nuclei. The error for all coordinates of both fits remain under 0.5%.
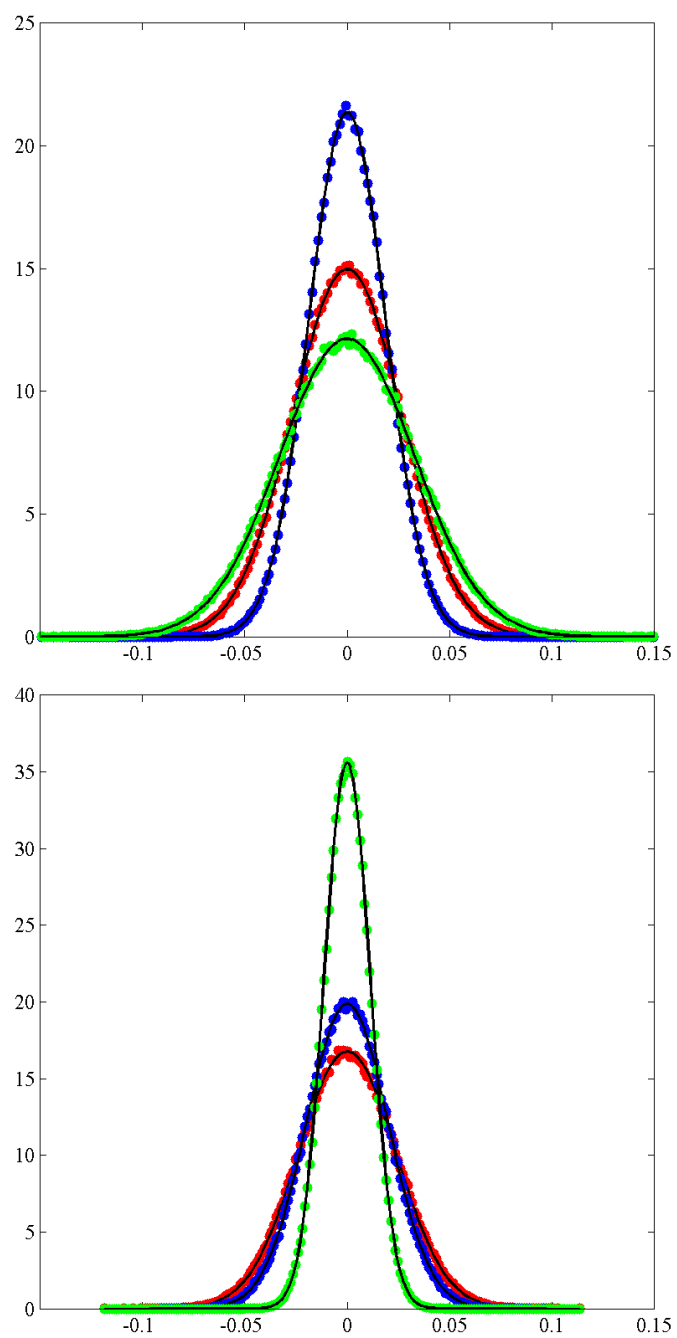
Figure 3.2 Cartesian displacements in Å (X,Y,Z = R,B,G) binned and plotted alongside Gaussian distributions (black) for *N3* (top) and *O6* (bottom) atoms.

Finally, it becomes necessary to make sure the MSDA is preserved. Having previously calculated the MSDA of the equilibrium structure it is possible compare the covariance of the ensemble of geometries and the two should be similar. Comparison of the covariance of the ensemble with the MSDA from normal mode analysis reveals that the largest error is 9.15e-05 Å^2 which occurs with the covariance between the Y and Z coordinates of the H5 nucleus. It is not possible to evaluate the percent error in the H5(Y,Z) case because the associated MSDA element is zero, but the largest difference between covariance and MSDA which is not targeting a zero valued MSDA element is only 2.70e-05 Å^2. This element represents the correlated motion of the H4 and H5 atoms along the Z axis (with a percent error of only 0.147%).

3.4.2 Normal Distribution of Critical Points

Each member of the ensemble has its own set of 5 CPs and, as such, it is possible to choose a specific CP and perform a statistical analysis on its properties throughout the ensemble. The distribution of the BCP locations can be shown to be Normal in much the same way that the distribution of nuclear positions was validated. The equilibrium and Mean positions of the BCPs are compared in Table 3.2 where it is again apparent that the Mean BCP locations are the same as the equilibrium BCP locations.

Table 3.2 Distance from equilibrium (Eq) and average (Avg) BCP locations.

Bond	$\left\| X_{Eq} - X_{Avg} \right\|$ in Å
$H1$-$C2$	1.155e-03
$C2$-$N3$	6.780e-03
$C2$=$O6$	1.452e-03
$N3$-$H4$	3.347e-03
$N3$-$H5$	5.090e-03

The standard deviations of these displacements are again found, binned, and each fitted to a univariate Gaussian to show that they are Normally distributed. The largest percent error (0.77%) is found for the X coordinate in the $C2$-$N3$ BCP location, while all other sum of squares of errors are below 0.17%. Figure 3.3 shows the fitted Gaussians to the $C2$-$N3$ and the $C2$=$O6$ BCPs.
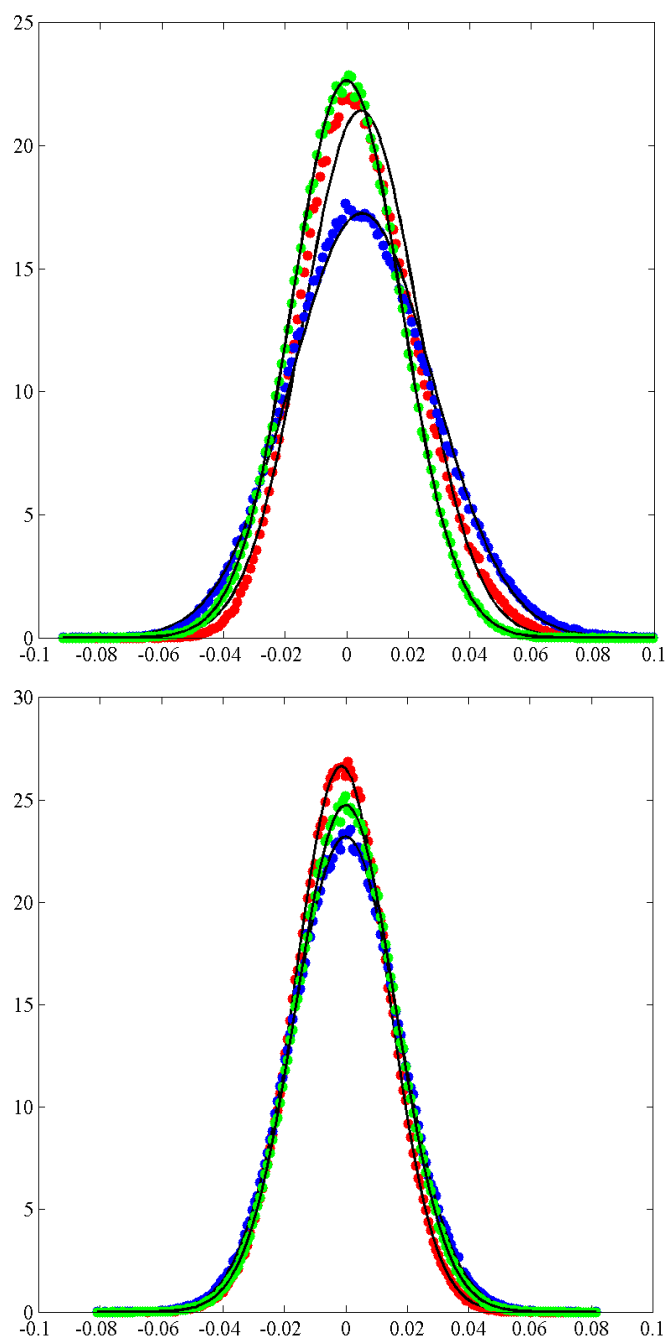
Figure 3.3 Cartesian displacements in Å (X,Y,Z = R,B,G) binned and plotted alongside Gaussian distributions (black) for $C2$-$N3$ (top) and $C2$=$O6$ (bottom) BCPs.

3.4.3 Reconstruction of Critical Point ADPs

As part of the validation of harmonicity of the nuclear motion, a comparison between the covariance of the nuclear positions with the MSDA was made and it was concluded that the two were nearly identical. The same method can be applied to the large ensemble of BCPs to construct the MSDA of BCPs. Presently the ability to construct the ADPs of BCPs requires substantial computational resources to generate a large ensemble of wave function based densities. As such, an effort is made to find a simple analytic approximation to the ADPs of BCPs to be compared with the available data.

To predict the unknown ADPs of a BCP from the known ADPs o nuclei in the bond, we developed a method similar to that of Scheringer [29, 48]. We construct the unknown BCP ADP, U^*, as a linear combination of nuclear ADPs, $U^{i,i}$, and their correlations, $U^{i,j}$, (these are all symmetric second rank tensors, $U^{i,i}$ representing the ADPs of the ith nucleus and $U^{i,j}$ representing the correlation between the ith and j^{th} nucleus).

We approximate the U^* tensor as

$$U^* = \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j U^{i,j} \tag{3.1}$$

subject to the constraint that $\sum_{i=1}^{N} \alpha_i = 1$ where $N = 2$ corresponding to the two nuclei taking part in bonding. The percent error for a tensor approximation is calculated as

$$Error = 100 * \frac{\sum\left(U_{i,j}^* - U_{i,j}^{CP}\right)^2}{\sum\left(U_{i,j}^{CP}\right)^2} \tag{3.2}$$

and is used to compare the results of different methods for finding α_i.

We analyze four methods for approximating the ADPs of these BCPs and assess their accuracy. Each method involves finding a single parameter, α_1, since $\alpha_2 = 1 - \alpha_1$.

Method 1 utilizes the generated ensemble data to find a best fit for α_1 to show the potential accuracy for the model. Method 2 defines α_1 as the percentage of the distance from nucleus 1 to the bond length, $\alpha_1 = \frac{|X_{CP}-X_1|}{|X_2-X_1|}$. Method 3 defines α_1 to be the same as in method 2, but ignores correlation in the construction of U^*. The difference between method 2 and 3 will show the necessity for including off diagonal elements of the MSDA in the construction of U^*. Finally, method 4 keeps $\alpha_1 = 0.5$ as a comparison to a rudimentary model where an average between the two nuclear ADPs [30]. Table 3.3 lists the calculated errors for each method.

Table 3.3 Errors for reproduction of thermal parameters of BCPs using various methods.

Bonds	Method 1 Error	Method 2 Error	Method 3 Error	Method 4 Error
$H1$-$C2$	3.22 %	4.75 %	10.02 %	12.66 %
$C2$-$N3$	7.79 %	17.05 %	9.23 %	27.68 %
$C2{=}O6$	2.12 %	3.43 %	22.20 %	35.43 %
$N3$-$H4$	0.26 %	0.37 %	2.06 %	33.76 %
$N3$-$H5$	0.09 %	0.18 %	1.56 %	33.52 %

3.4.4 Thermally Averaged Topological Properties

In the current work, we analyze how the bond TPs of the formamide molecule are affected by internal nuclear vibrations. To a given maximal value of probability of nuclear displacements, thermal ellipsoids can be constructed for each nucleus [50]. An ellipsoid is defined as $\vec{d}^T U^{-1} \vec{d} = c^2$ where $\vec{d}$ is the displacement vector, U, is the 3×3 symmetric ADP tensor and $c = 1.5382$ for a probability of 50%. In general, the probability of finding an atom inside an ellipsoid can be found by [50]

$$\text{Pr}(c) = \left(\frac{2}{\pi}\right)^{1/2} \int_0^c r^2 \text{Exp}[-r^2/2]\, dr = \sqrt{\frac{2}{\pi}}\left(-ce^{-\frac{c^2}{2}} + \sqrt{\frac{\pi}{2}}\,\text{Erf}\left(\frac{c}{\sqrt{2}}\right)\right) \quad (3.3)$$

One manner of quantifying the distribution of TPs resulting from nuclear vibrations is to compare this distribution to the accompanying distribution of bond lengths resulting from the same nuclear vibrations. In Figure 3.4 the distribution of ED (top) and Laplacian (bottom) at the $C=O$ BCP closely resembles experimental results by Gibbs et. al. [44] (a power law of decay). To get a better idea of the distributions, it is beneficial to classify each member of the ensemble into one of three groups based on the number of configurations for which the nuclei involved in bonding stay inside the 50% thermal ellipsoids.
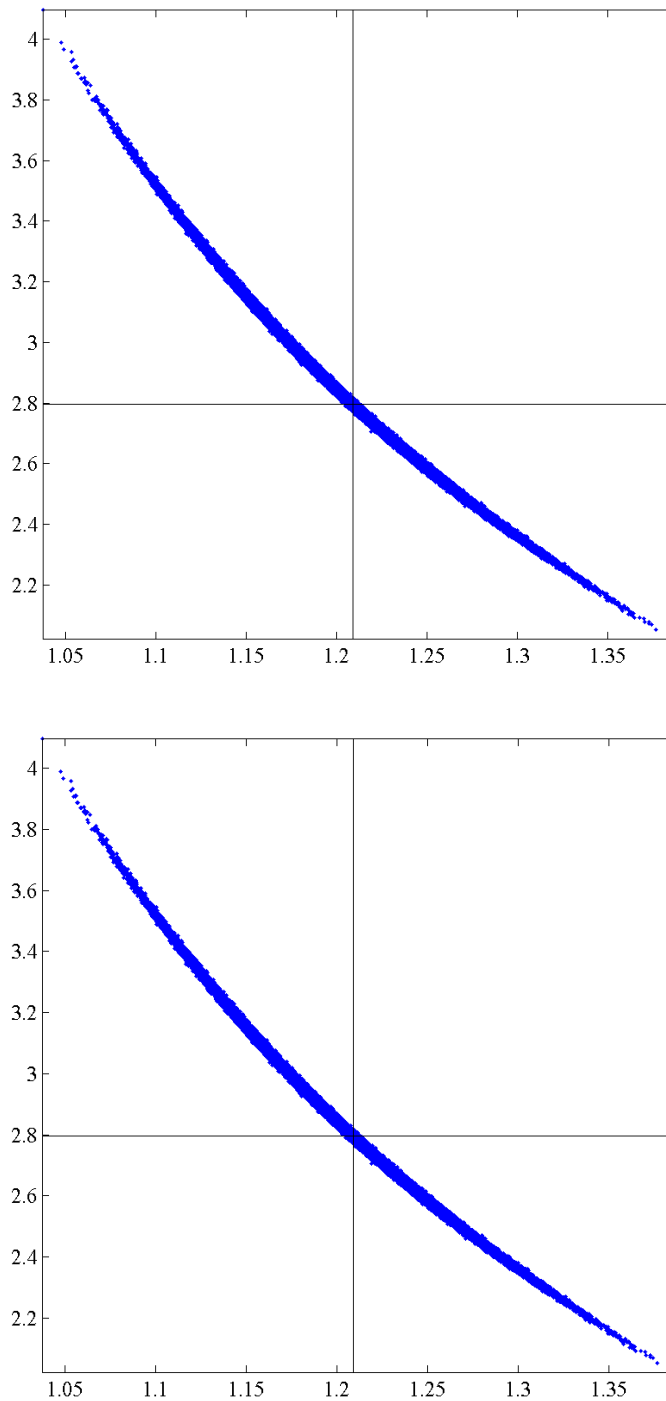
Figure 3.4 Distributions of TPs ($\rho(\boldsymbol{r_c})$; $e/\text{Å}^3$ and $\nabla^2\rho(\boldsymbol{r_c})$; $e/\text{Å}^5$) for the $C=O$ bond in the ensemble. Vertical and horizontal lines represent the bond distance and TP of the associated static ED.

Figure 3.5 plots the TPs for the $C{=}O$ BCP again with each point (member of the ensemble) designated by one of the three groups. Group 1 (red) has both nuclei located outside of their 50% ellipsoids, group 2 (green) has exactly one nucleus inside and one nucleus outside of their respective ellipsoids, and group 3 (blue) has both nuclei inside their ellipsoids.
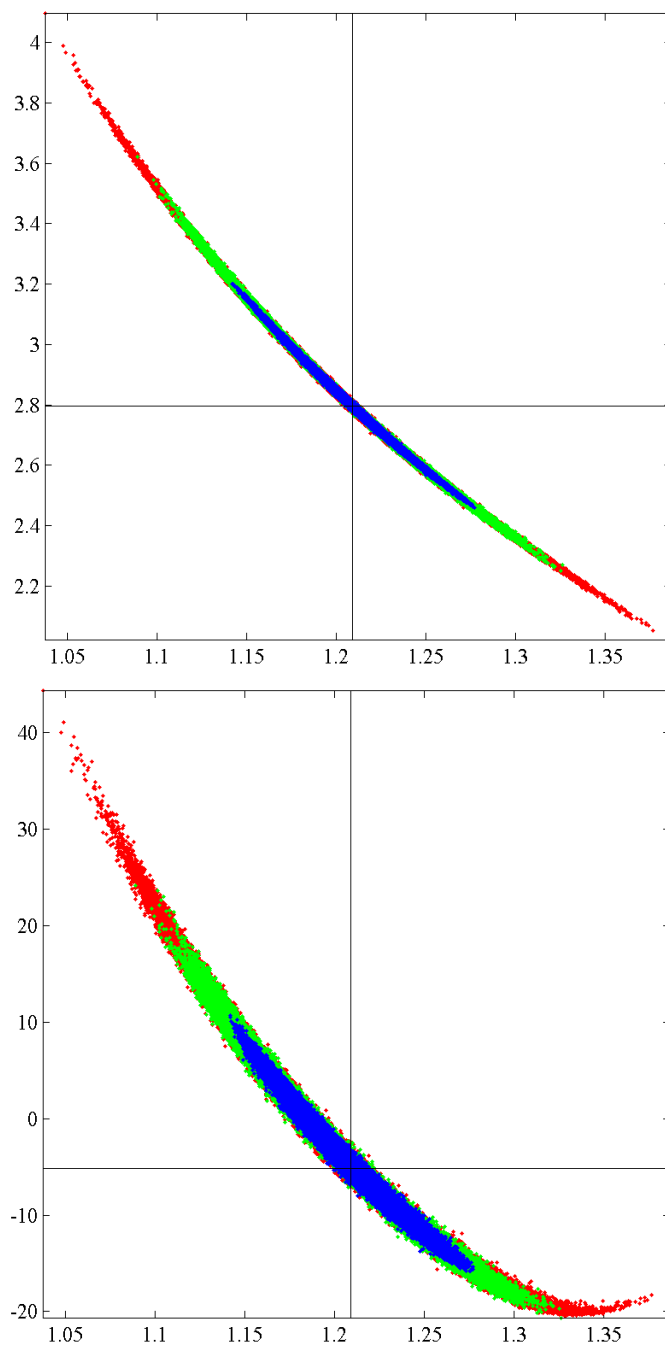
Figure 3.5 Distributions of TPs ($\rho(r_c)$; $e/\text{Å}^3$ and $\nabla^2\rho(r_c)$; $e/\text{Å}^5$) for the $C\!=\!O$ bond in the ensemble. Each member of the ensemble is plotted in either Red, Green or Blue depending on how many nuclei are inside their 50% probability ellipsoids.

The bond TP change is due, primarily, to bond stretching as evidenced by Figure 3.6. The effect of stretching the bond, alone, is plotted by the black dotted line and it is obvious that this trend is the main cause of the distribution. The 'width' of the distribution is markedly different for the $C=O$ and $C\text{-}N$ bonds; the $C\text{-}N$ TPs seem to have much higher variance than do the $C=O$ TPs.
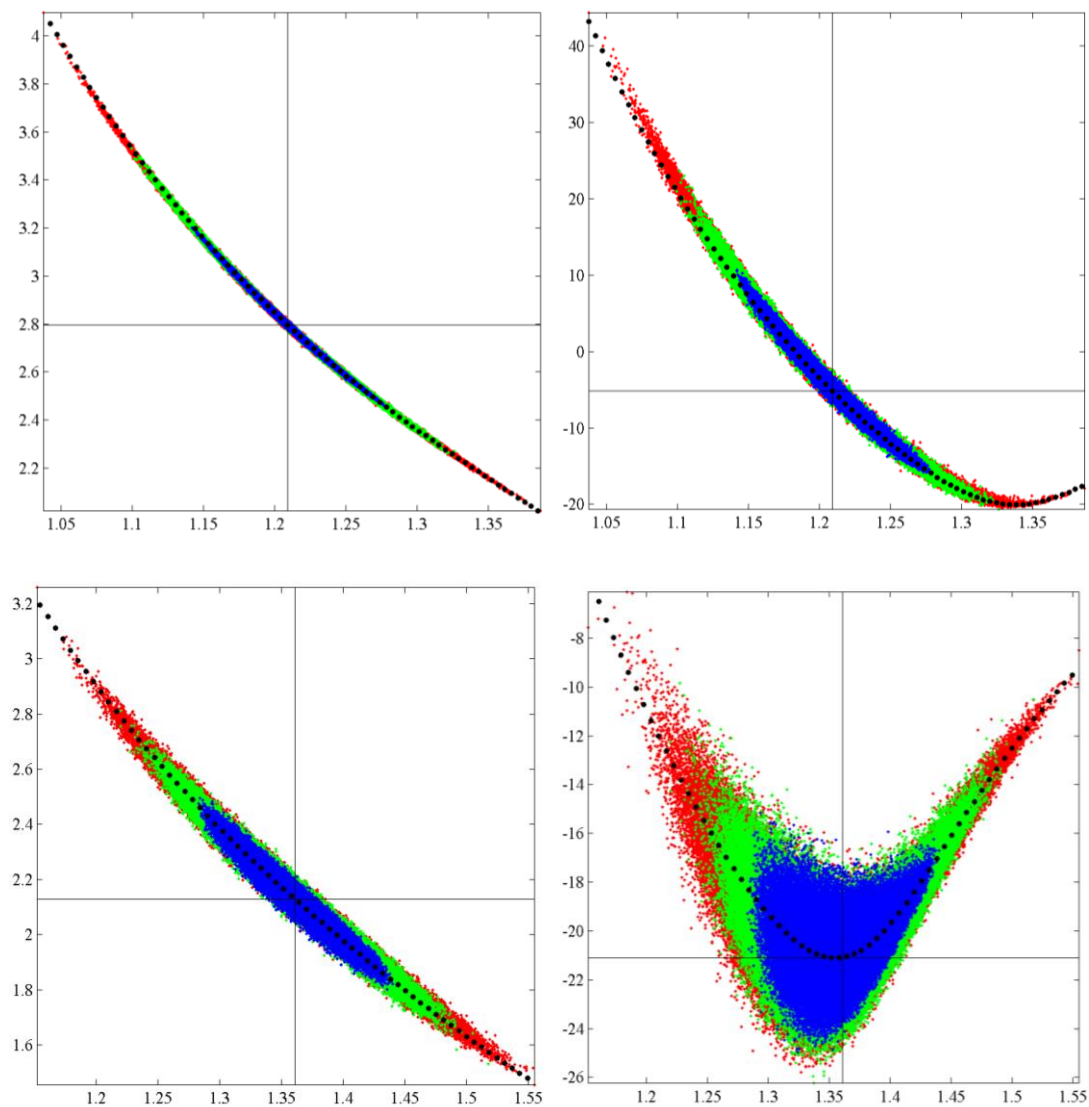
Figure 3.6 Distributions of $\rho(\boldsymbol{r_c})$ (left) and $\nabla^2\rho(\boldsymbol{r_c})$ (right) for C=O (top) and C-N (bottom) BCPs in the ensemble. Each point represents one of the 500,000 members of the ensemble and is classified by how many (0 - Red, 1 - Green, 2 - Blue) of the nuclei are inside the associated 50% probability ellipsoid. The black dotted line shows the effect that stretching the bond, alone, has on the TP.

The characteristic difference between the $C=O$ bond and the C-N bond in this molecule is that the bond TPs of the former are primarily affected by only one normal mode whereas those for the latter are affected by many normal modes. Figure 3.7 shows this difference by plotting the affect that each normal mode, acting in isolation, has on the distribution of TPs.
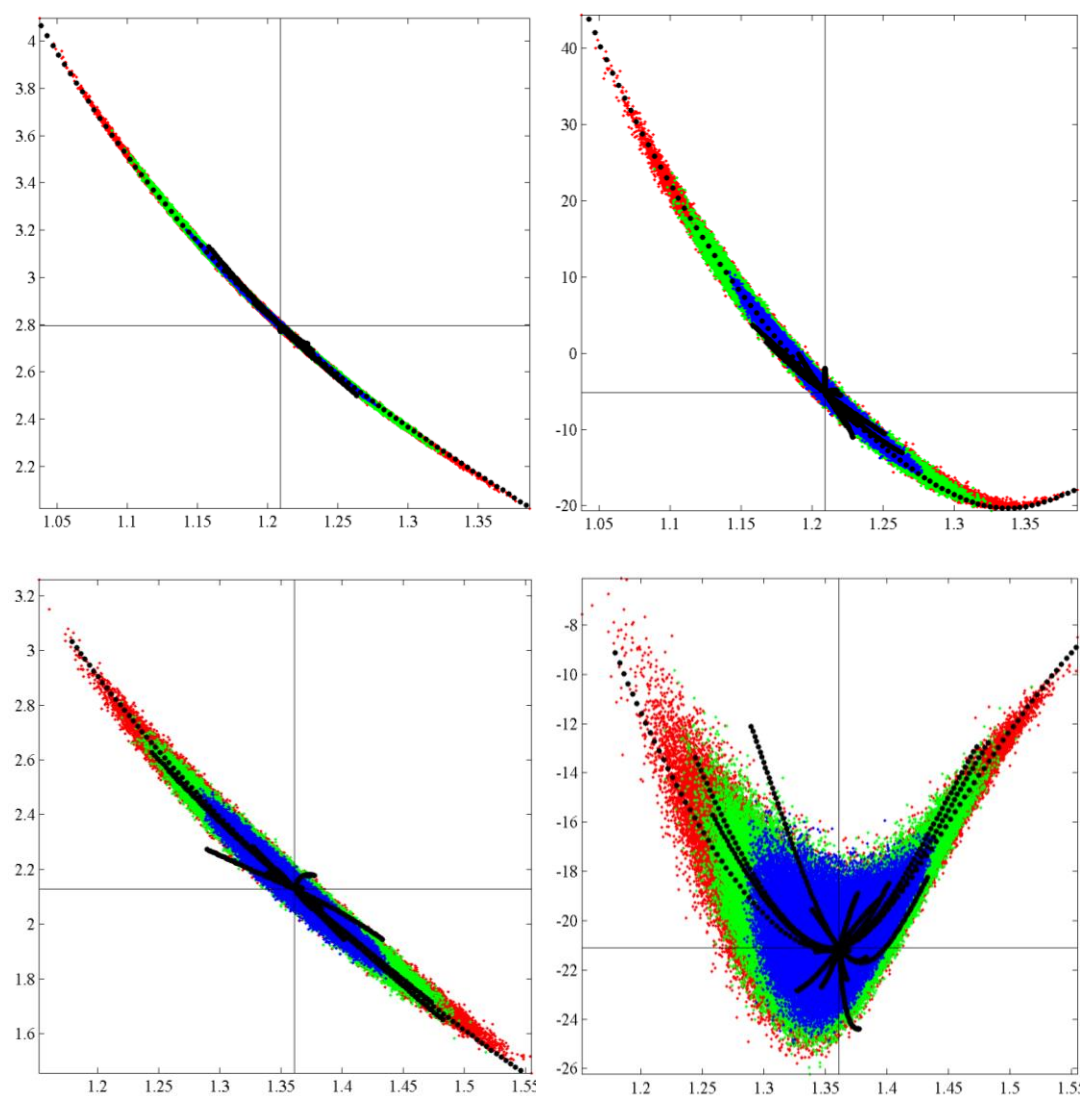
Figure 3.7 Distributions of $\rho(\boldsymbol{r_c})$ (left) and $\nabla^2\rho(\boldsymbol{r_c})$ (right) for $C=O$ (top) and $C\text{-}N$ (bottom) BCPs in the ensemble. Each point represents one of the 500,000 members of the ensemble and is classified by how many (0 - Red, 1 - Green, 2 - Blue) of the nuclei are inside the associated 50% probability ellipsoid. The black dotted lines shows the effect that each normal mode, in isolation, has on the TPs.

Visual inspection of the distributions above leads to the conclusion that the TPs are not normally distributed about the mean TP. To demonstrate this quantitatively, the first four statistical moments (mean, variance, skewness, kurtosis of the centered data) as well as the error for fitting the binned data to a Gaussian distribution are listed in Table 3.4.

Table 3.4 First statistical moments of distributions and % error for fitting the binned data to a Normal distribution.

		Mean	Variance	Skewness	Kurtosis	% Error
$(H1,C2)$	ED	-0.022	0.094	0.550	3.511	1.670%
	Laplacian	-0.059	53.111	-0.987	4.281	6.356%
$(C2,N3)$	ED	0.006	0.031	0.236	3.169	0.303%
	Laplacian	0.946	3.438	0.928	4.566	4.811%
$(C2,O6)$	ED	0.002	0.042	0.372	3.283	0.740%
	Laplacian	0.657	50.013	0.577	3.503	1.645%
$(N3,H4)$	ED	-0.107	0.158	0.601	3.744	1.811%
	Laplacian	2.64	229.692	-1.161	4.980	8.550%
$(N3,H4)$	ED	-0.160	0.169	0.494	3.631	1.294%
	Laplacian	4.327	228.254	-1.146	4.956	7.989%

From this statistical treatment of the TPs, we can easily conclude that the distributions are not Normal. Mathematically, this can be validated by combination of two facts: the distribution of the nuclei (and by extension the bond distances) are not normally distributed, and the TPs vary with bond distance according to a power law. The combination of these two statements indicates that the TPs are non-linear functions of a normal distribution which means that they cannot, themselves, be normally distributed.

4. Validation of the Convolution Approximation[2]

4.1 Applications of Convolution to Experimental Data

Analysis of high-resolution single-crystal X-Ray diffraction data has become a routine procedure to elucidate the ED of solids [52, 53]. This experimental route is considered by many investigators to be complementary or even alternative to the theoretical one offered by computational quantum chemistry, despite the vague link between the models as well as the driving principles utilized in the two methods. Nevertheless, for the lion's share of the experimental reports the focus has chiefly been on validation of the results accomplished by comparing theoretical and experimental EDs, commonly in terms of BCP properties [10]. This collation is conceivable in two substantially different ways. In the usual course of modern X-Ray studies an analytic SF model (dominantly the rigid pseudoatom model [22, 23]) is called upon that explicitly accounts for thermal smearing (density deformation due to nuclear vibrations) and whose static and dynamic parameters are jointly adjusted to the observed data using the standard least-squares (LS) protocol. Such a fitting procedure allows for a technically straight forward and apparently satisfactory, yet intractable, decoupling of thermal smearing effects from bonding effects [54, 55]. In other words, the 'experimental' static ED and the ADPs obtained through such an analysis is unavoidably biased due to the failure of the applied scattering formalism to explicitly and adequately account for the physics underlying the coupling between electronic and nuclear motion.

---

[2] The contents of this chapter also appears in The Journal of Mathematical Chemistry [79].

A far less common method nowadays, but critically reflected on by the pioneers of the field [56, 57], is to evaluate the vibration average of the theoretical ED using experimental ADPs. The efforts have been hampered by a number of issues / uncertainties; (a) the experimental ADPs have a limited physical significance, not only for the above mentioned reason, but also because these parameters are likely to absorb systematic errors and are affected by disorders (static and dynamic structural fluctuations); (b) internuclear correlations (vibration couplings) are not accessible from Bragg diffraction; (c) the lack of rigorous treatment of the vibration smearing of internuclear (two-center) density units. The last issue has been an indisputable source of bias in density matrix fitting [30, 58, 59] and wave function supported refinements (two-center ED models) of X-Ray data [17].

It is to be emphasized that both approaches (that is, decoupling the thermal motion from the experimental or smearing the theoretical ED) rely on the harmonic-convolution approximation which is at the heart of modeling coherent elastic diffraction [60], but fails to comply with the Born-Oppenheimer (BO) approximation [1], which, on the other hand, is at the heart of molecular quantum chemistry. This contradiction provokes the plain questions: are the theoretical and 'experimental' EDs comparable at all and what can we reasonably expect to learn from their agreement / disagreement after all?

Assessment of the convolution approximation has been the subject matter of computational studies on small molecules [61, 62] as well as temperature dependent diffraction analyses [63, 64]. The temperature independence of experimental static BCP properties has often been used to infer the 'thermal-decoupling capability' of a scattering / static ED model used for the analysis of the diffraction data [65].

In this study we evaluate the theoretical dynamic molecular ED derived as the exact convolution of the nuclear PDF for harmonic internal vibrations with the static molecular ED given within the LCAO-MO formalism in terms of Gaussian basis functions. The convoluted dynamic ED is compared with the mean ED obtained within the adiabatic approximation [66] as the average over a large number of static EDs corresponding to nuclear configurations consistent with harmonic vibrations. Both the electronic and nuclear distributions are derived according to standard computational chemistry protocols with the aid of the Gaussian09 program suite [15] and locally developed computer codes [51].

4.2 Convolution of Two-Center Gaussian Basis Products

Within the LCAO-MO approach, the ED can be expressed as a linear combination of basis function products:

$$\rho(\boldsymbol{r}) = \sum_{i,j}^{Nprims} C_{ij} g_i(\boldsymbol{r}) g_j(\boldsymbol{r}) \tag{4.1}$$

$$g_i(\boldsymbol{r}) = (x - X_i)^{n_i} (y - Y_i)^{m_i} (z - Z_i)^{l_i} Exp[-\alpha_i |\boldsymbol{r} - \boldsymbol{R}_i|^2] \tag{4.2}$$

where $g_i(\boldsymbol{r})$ is a (non-normalized) primitive Gaussian Type Orbital (GTO) centered at $\boldsymbol{R}_i = (X_i, Y_i, Z_i)$. The product of two primitive GTOs centered at (X_a, Y_a, Z_a) and (X_b, Y_b, Z_b) is also a GTO centered between the two [67].

Let $\boldsymbol{a} = (a_x, a_y, a_z) = (x - X_a, y - Y_a, z - Z_a)$ have a normal trivariate PDF with

zero expectation value ($\langle \boldsymbol{a} \rangle = (0,0,0)$) and covariance $\boldsymbol{U}_{aa} = \{Cov(a_i, a_j) = \langle \boldsymbol{a}^T \boldsymbol{a} \rangle\}$

(where i, j range over x, y, z), a 3×3 symmetric tensor, the ADPs of nucleus $\boldsymbol{R}_a$. If $\boldsymbol{b}$ is

likewise defined with $\boldsymbol{U}_{bb}$, then the displacement vector of the GTO at $\boldsymbol{c}$ defined by the

product of GTOs at $\boldsymbol{a}$ and $\boldsymbol{b}$ is

$$\boldsymbol{c} = \frac{\alpha \boldsymbol{a} + \beta \boldsymbol{b}}{\alpha + \beta} \tag{4.3}$$

where α and β are the exponents of the respective primitive GTOs of the product. The

covariance matrix of center $\boldsymbol{c}$ is

$$\boldsymbol{U}_{cc} = \langle \boldsymbol{c}^T \boldsymbol{c} \rangle = \frac{\alpha^2 \langle \boldsymbol{a}^T \boldsymbol{a} \rangle + \alpha \beta (\langle \boldsymbol{a}^T \boldsymbol{b} \rangle + \langle \boldsymbol{b}^T \boldsymbol{a} \rangle) + \beta^2 \langle \boldsymbol{b}^T \boldsymbol{b} \rangle}{(\alpha + \beta)^2}$$

$$= \frac{\alpha^2 \boldsymbol{U}_{aa} + \alpha \beta (\boldsymbol{U}_{ab} + \boldsymbol{U}_{ba}) + \beta^2 \boldsymbol{U}_{bb}}{(\alpha + \beta)^2} \tag{4.4}$$

where $\boldsymbol{U}_{ba} = (\boldsymbol{U}_{ab})^T$ are the off diagonal blocks of $\boldsymbol{U}$ representing correlation between the

vibration of nuclei $\boldsymbol{a}$ and $\boldsymbol{b}$. Eqn. 4.4 provides an exact representation for the ADPs of a

product of primitive GTOs. To reproduce a situation in which the off diagonals may be

unknown, it is sufficient to set both $\boldsymbol{U}_{ab}$ and $\boldsymbol{U}_{ba}$ to zero. The methods described in this

section are implemented in the current version of the Denprop [51] software package.

4.3 Results and Discussion

The convolution formalism detailed above is applied to the LCAO-MO EDs of the

formamide and octasulfur molecules at the B3LYP/6-311G** [43] and MP2/cc-PVTZ [68]

levels of theory, respectively, calculated at the equilibrium geometries using the

Gaussian09 program suite [15]. The MSDAs corresponding to a temperature of 23 K were

derived from the harmonic vibration frequencies and normal modes at the same levels of theory. The optimized geometries were obtained using the 'very tight' convergence criteria leading to small frequencies for the external modes. For the sake of simplicity, we imposed planarity for formamide, as this structure has been the subject of a vibration smearing study [69] based on Fourier expansion and external ADPs. The lowest frequency (highest amplitude) mode is the NH_2 wagging ($\omega(NH_2)$) which displaces all three atoms out of the molecular (XY) plane. Since we consider only internal modes of vibration, the MSDA matrix has a rank of $3N - 6$. Nevertheless, the block diagonal matrices (the ADPs) are positive definite (with condition number greater than 8) thus defining valid trivariate normal PDFs.

Two alternative dynamic EDs are evaluated for formamide. One is also a convoluted ED but without correlations in the nuclear vibrations, that is, only the block diagonal elements of the total MSDA matrix are included in the convolution ($\langle\rho(\boldsymbol{r})_{block}\rangle_T$). This approach closely resembles the formalism used to model Bragg diffraction data from which no covariance information can be retrieved. The other one is obtained as an average over a large ensemble of static EDs (sample size; $M = 5 \cdot 10^5$) generated by sampling the nuclear configuration space (using the nuclear PDFs due to the B3LYP/6-311G** normal modes) and mapping each member of this ensemble onto an equivalent set of single-point electronic wave functions. The sample mean of the corresponding EDs ($\overline{\rho(\boldsymbol{r})} = M^{-1}\sum \rho_i(\boldsymbol{r}, \boldsymbol{R}_i)$) can be considered as the 'correct' thermal-average ED (within the time independent BO approach), since its evaluation does not rely on the convolution approximation; each member of the ED set is consistent with one and only one nuclear configuration [70].

One of the striking results of this simulation is that the static and related dynamic EDs exhibit an overall fair agreement, except for the NH_2 group of the formamide molecule. Figure 4.1 displays difference ED contour maps $((\rho(\boldsymbol{r})) - \langle\rho(\boldsymbol{r})\rangle_T)$ in the molecular plane of formamide (1a) and in the plane defined by three consecutive S-atoms of the octasulfur molecule (1b) whose molecular structure is displayed in Figure 4.2. These maps well demonstrate the effect of nuclear vibrations; charge migrations from near-nuclear regions toward internuclear regions. The BCP properties of the two EDs (Table 4.1) are very close in value. Even the largest differences, obtained for the positive eigenvalues of the Hessian at the BCP (λ_3), remain below 10% for bonds formed between massive atoms. However, no BCP is found for any of the $N\text{-}H$ bonds on the dynamic EDs; neither for the convoluted $(\langle\rho(\boldsymbol{r})\rangle_T, \langle\rho(\boldsymbol{r})_{block}\rangle_T)$ nor for the averaged $(\overline{\rho})$. These results are consistent with those found for the 'experimental' ED of trialanine [71] by the maximum entropy method which reconstructs the dynamic ED from diffraction data without having resource to any thermal smearing and static ED model.
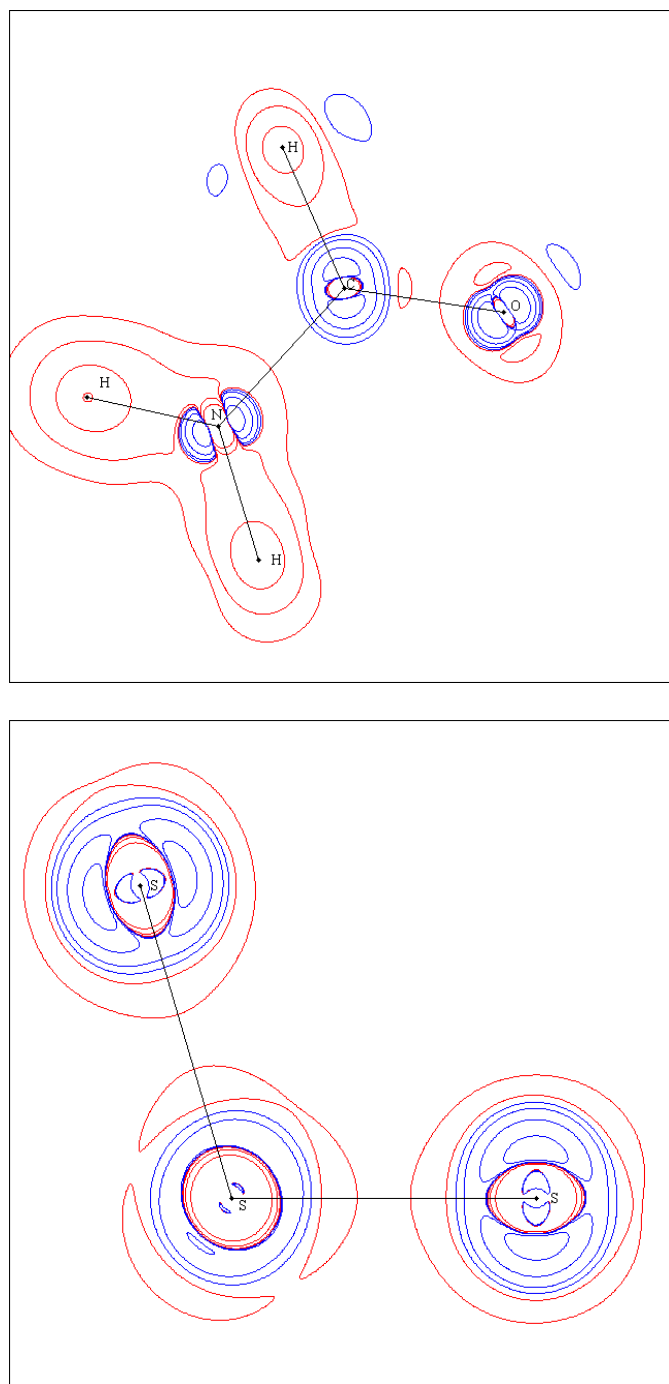
Figure 4.1 Difference ED $(\rho(\boldsymbol{r}) - \langle\rho(\boldsymbol{r})\rangle_T)$ contour maps: (a, top) formamide in the molecular plane, (b, bottom) octasulfur in the plane defined by three S-atoms. Contour levels are $\pm(0.01, 0.05, 0.25, 1.25)\ e/\text{Å}^3$, red / blue lines are postive / negative contours.
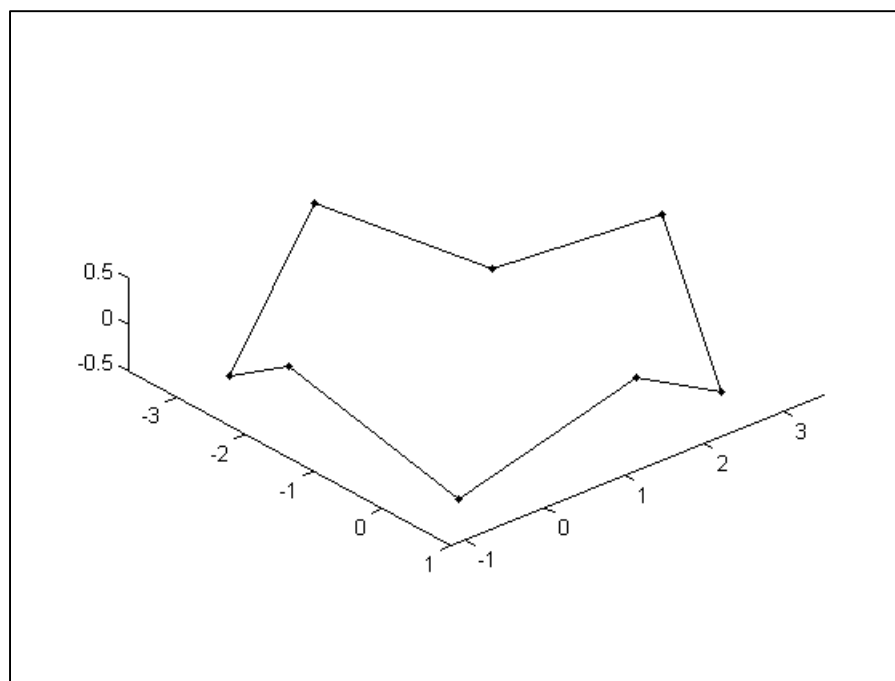
Figure 4.2 Molecular geometry of octasulfur in Å

On Figure 4.3a the difference density $(\langle\rho(\boldsymbol{r})_{block}\rangle_T - \langle\rho(\boldsymbol{r})\rangle_T)$ in the plane of the formamide molecule reveals that the independent nuclear vibration model (block diagonal representation) is in a close agreement with that obtained by the correlated model using the full MSDA. The maximum error is only $0.015\ e/\text{Å}^3$, which occurs at the location of the Nitrogen nucleus. Figure 4.3b shows the difference in $(\overline{\rho(\boldsymbol{r})} - \langle\rho(\boldsymbol{r})\rangle_T)$ where a quantitative comparison highlights that the maximum absolute error for formamide is only $1.45\ e/\text{Å}^3$ found at the nucleus of the oxygen atom.

Table 4.1 Topological Properties at BCPs of Static and Dynamic EDs.
Comparison of ρ ($e/\text{Å}^3$; top line), $\nabla^2\rho$ ($e/\text{Å}^5$; second line)
and λ_i ($e/\text{Å}^5$; lines 3-5) for BCPs in formamide and octasulfur.

Bond		Static	Convolution	Average
C-H	ρ	1.861	1.791	1.798
	$\nabla^2\rho$	-22.582	-22.903	-23.387
	λ_1	-18.003	-16.553	-16.867
	λ_2	-17.570	-16.290	-16.572
	λ_3	12.991	9.939	10.052
C-N	ρ	2.128	2.119	2.132
	$\nabla^2\rho$	-21.090	-20.069	-20.152
	λ_1	-16.376	-16.220	-16.371
	λ_2	-14.910	-14.825	-15.106
	λ_3	10.196	10.976	11.325
$C{=}O$	ρ	2.796	2.785	2.797
	$\nabla^2\rho$	-5.200	-2.894	-2.629
	λ_1	-25.971	-25.760	-25.978
	λ_2	-24.254	-24.180	-24.395
	λ_3	45.025	47.046	47.744
S-S	ρ	0.995	0.988	
	$\nabla^2\rho$	-2.693	-2.777	
	λ_1	-4.043	-3.983	
	λ_2	-4.000	-3.931	
	λ_3	5.350	5.136	
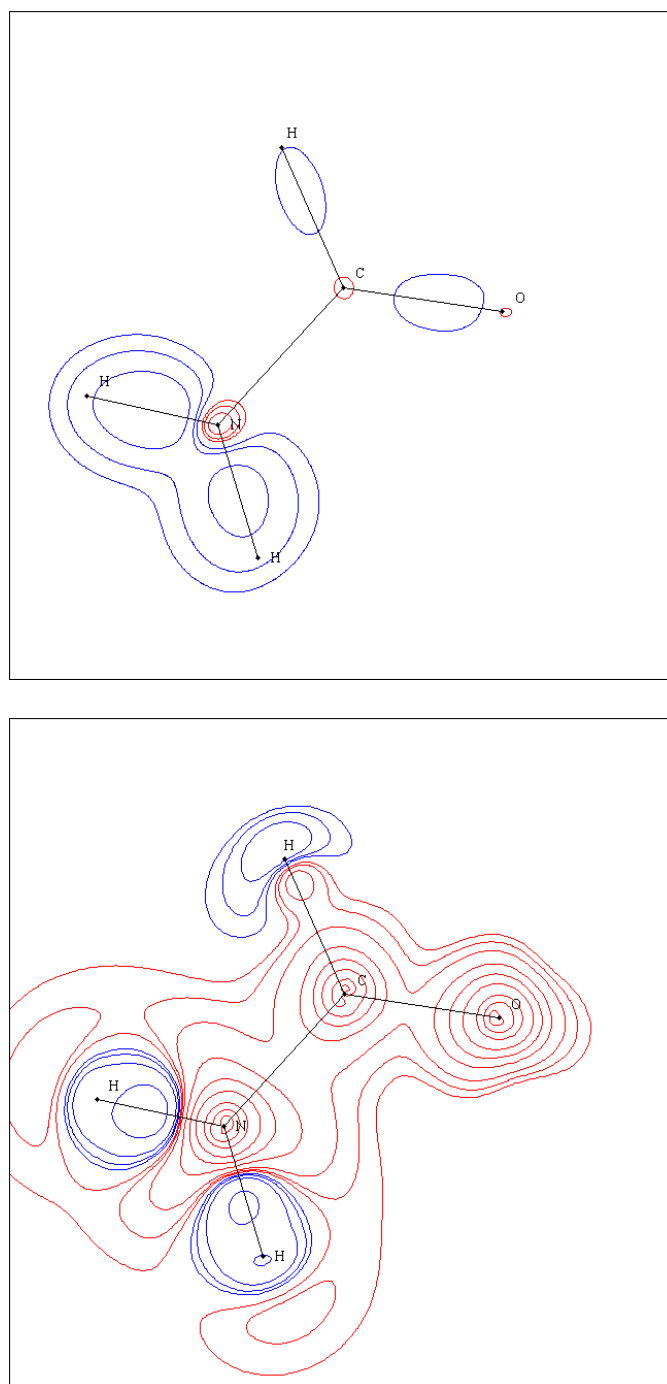
Figure 4.3 The differences (a, top; $\langle \rho(\boldsymbol{r})_{block} \rangle_T - \langle \rho(\boldsymbol{r}) \rangle_T$) and (b, bottom; $\overline{\rho(\boldsymbol{r})} - \langle \rho(\boldsymbol{r}) \rangle_T$) for formamide in the plane. Contour levels are $\{\pm 2^n\}_{n=1,2,3} \{10^{-m}\}_{m=1,2,3}$ $e/\text{Å}^3$ expressed by red / blue lines for positive / negative values.

Figure 4.4 compares the Laplacian of the static $\nabla^2\rho(\boldsymbol{r})$ and dynamic $(\nabla^2\langle\rho(\boldsymbol{r})\rangle_T)$ EDs along the $C{=}O$ (4.4a) and S-S (4.4b) bond paths. While a surprisingly good agreement is found in the near-BCP region for both bonds, the residual functions $(\nabla^2\langle\rho(\boldsymbol{r})\rangle_T - \nabla^2\rho(\boldsymbol{r}))$ exhibit pronounced differences in the vicinity of the valence-shell charge concentrations (VSSC), especially for the polar $C{=}O$ bond. These findings correlate well with the experimental irreproducibility of the profiles of $C{=}O$ bonds [52] which might suggest incomplete decoupling of thermal smearing effects and static density topology from X-Ray structure factors [55].
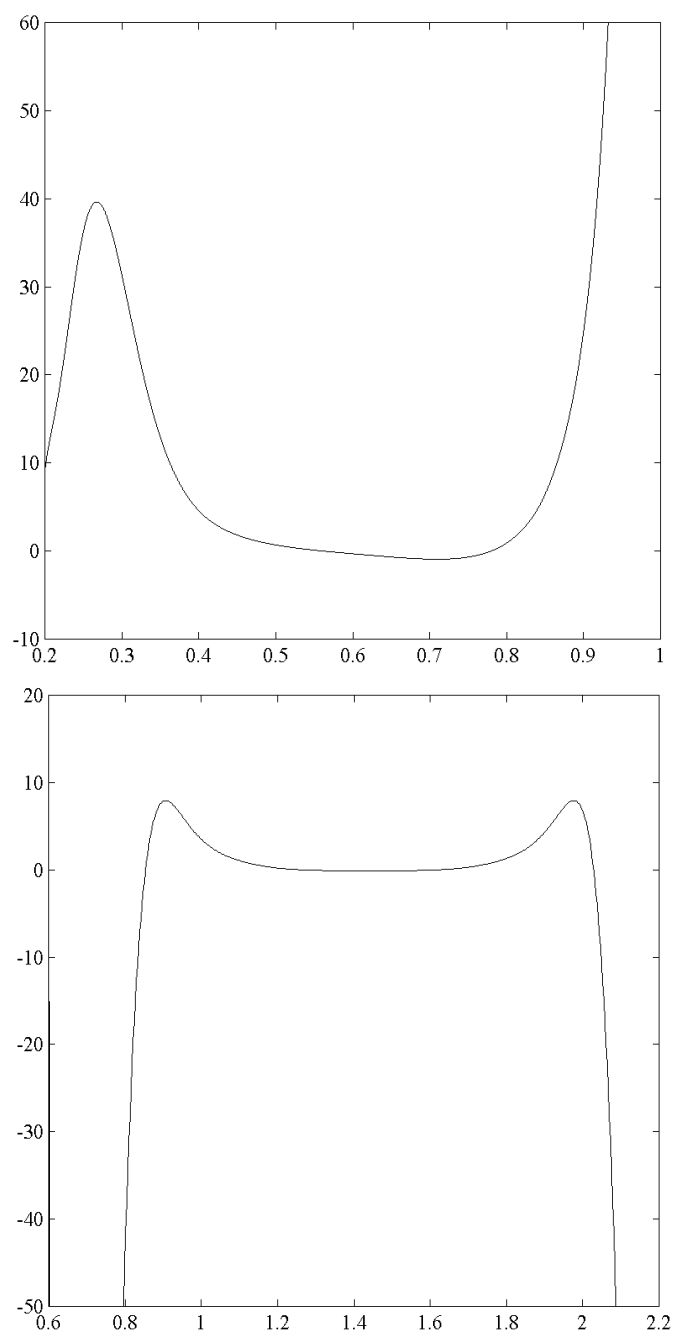
Figure 4.4 The dynamic Laplacian with reference to that of the static $(\nabla^2 \langle \rho(\boldsymbol{r}) \rangle_T - \nabla^2 \rho(\boldsymbol{r}))$ along (a, top) the $C=O$ bond for the formamide molecule and (b, bottom) the $S\text{-}S$ bond in the octasulfur molecule. Units are $e/\text{Å}^3$.

5. The Effect of Dynamic Parameters on the Static Electron Density

5.1 Parameter Estimates of the Electron Density

Due to the Least Squares method in the X-Ray Diffraction refinement program XD [35] and the fact that the refined parameters are highly correlated, their standard uncertainties can affect the resulting model density. As the model density is used to glean physical information about the underlying chemical structure, the uncertainties in the model parameters will implicitly cause uncertainties in the related topological structure. For this reason it is important to quantify the uncertainty in the TPs of the ED due to the uncertainties in the model parameters.

Recall that there are primarily two types of model parameters under investigation. The *dynamic parameters* ('thermal parameters'; ADPs) account for the thermal smearing of the ED and thus the observed experimental data, but they are not seen again in the analysis of the static TPs. The *static parameters* ('populations'; P_{lm}s, parameters of the RDFs and the nuclear mean positions) are explicitly involved in the static representation of the molecular ED and have a direct effect on the related topological analysis. There is, however, an implicit connection between the dynamic and static parameters of the PA model since their least squares estimates are unavoidably correlated.

To analyze the standard uncertainty of a function, error propagation [34, 35, 72] can be implemented to map the uncertainties from the independent variables of the function to the function value itself. The standard uncertainty (SU) of a function $f(x_1, \ldots, x_n)$ can be expressed in terms of the SUs of its variables $\{x_i\}$ as [34]

$$\sigma^2(f) = \boldsymbol{g}^T \boldsymbol{V} \boldsymbol{g} \tag{5.1}$$

where $g_i = \frac{\partial f}{\partial x_i}$ and V_{ij} is the covariance between parameters x_i and x_j. This method is primarily invoked [35] to monitor the SU of the static PA ED. Implicit in this method of error propagation is the fact that only the SUs of the static parameters are involved (the reason for this is clear: the dynamic parameters are not, formally, included in the expression of the static ED and thus, $\frac{\partial f}{\partial U_{ij}} = 0$ so the related elements in the vector $\boldsymbol{g}$ from Eqn. 5.1 are also zero).

Error propagation (Eqn. 5.1) can exactly reproduce the SU of a function with linear dependence on parameters (and linearization can be employed to work with functions including dependence upon non-linear parameters). In the expression of scalar TPs $(\rho(\boldsymbol{r}), \nabla^2 \rho(\boldsymbol{r}))$, the above error propagation can perfectly model the uncertainty in the ED due to the linear parameters, such as multipole populations. This naïve approach cannot, however, accomplish two fundamental things.

First, it cannot show how the SUs of the dynamic parameters affect the SUs of the TPs (due, again, to the fact that these U_{ij}s are not formally included in the expression of the static ED). Secondly, it can only model the SU of scalar properties. An important aspect of computing the TP at a CP is finding the 'exact' location of the CP. Because finding a CP is a numerical iterative process, the location of a CP cannot be expressed concisely in

terms of the static parameters and thus the error propagation scheme above cannot properly assess the uncertainty in the location of the CP.

In the current work, we implement methods to perform uncertainty quantification (UQ) on scalar and non-scalar properties of the ED which incorporate both static and dynamic parameters.

5.2 Methods

Upon successful refinement of experimental XRD data using the XDLSM program suite, the user is presented with a set of refined parameters, X, and the covariance matrix, V, both of which are involved when performing UQ on the resulting model density. Details for the refinement process will be given in more depth in section 5.3, but for now it is sufficient to describe the two types of parameters being refined: the static parameters, P_{lm}s, and the dynamic parameters, U_{ij}s.

The resulting parameter-estimate vector X is thought of as a N-vector which can be blocked into N_d and N_s sized vectors: $X = [X_d : X_s]$, where the number of total parameters refined, N, is then the sum of the number of dynamic, N_d, and static parameters, N_s. The covariance matrix can be blocked accordingly:

$$V = \begin{bmatrix} V_d & V_{ds} \\ (V_{ds})^T & V_s \end{bmatrix} \tag{5.2}$$

We perform UQ by three different ways: Error Propagation (EP), Static Parameter Sampling (SPS), and Dynamic Parameter Sampling (DPS).

To simulate the analysis of experimental XRD data, dynamic SFs were calculated based on the Gaussian smearing approach described in Chapter 4. Accurate reconstruction of TPs at BCPs, as described in section 5.3, will be further validation that the model

employed for convolution accurately represents a dynamic system in the harmonic approximation.

5.2.1 Error Propagation

In the EP method, we are reproducing methods which are currently employed for UQ of scalar properties of the ED. We perform this method to find the SU of $\rho(r)$ and $\nabla^2 \rho(r)$ at various points in the plane of formamide involving only X_s and V_s to perform the analysis.

5.2.2 Static Parameter Sampling

In the SPS method we still take only X_s and V_s, but we use them to create a 25,000 member ensemble of densities, $\{\rho_i\}$, sampled from the multivariate normal distribution $N(X_s, V_s)$. This method is similar to the method invoked in previous chapters to analyze thermal motion effects, but it is now being employed to consider the effects of parameter uncertainties on the static ED. The SPS method should duplicate the results of the EP method but can also estimate the uncertainty in the location of BCPs, which the EP method cannot.

5.2.3 Dynamic Parameter Sampling

Finally, the DPS method takes into account the full X and V parameter sets to manifest uncertainties of the dynamic parameters in the static ED. The 25,000 member ensemble for the DPS data set is constructed differently than the SPS data set and consists of two phases in the construction.

In Phase 1, (X_d, V_d) is used to sample 250 sets of dynamic parameters from the normal distribution $N(X_d, V_d)$ yielding an ensemble of dynamic parameters, $\{D^j\}$. For each D^j, a LS refinement is performed on the static parameters yielding $(X_s, V_s)^j$ which

are static parameters and their covariance for a specific set of dynamic parameters, D^j. In Phase 2, SPS is performed on $(X_s, V_s)^j$ with an ensemble size of 100. The resulting ensemble $\{\rho_i^j\}$, the cardinality of which is 25,000, now includes SUs from both static and dynamic parameters.

5.3 Results and Discussion

The refinement is performed on theoretical SFs generated by the Gaussian smearing method established in Chapter 4. Theoretical SFs are generated up to $Sin(\theta)/\lambda = 1.2$ such that $(-40 \leq h \leq 40, 0 \leq k \leq 40, 0 \leq l \leq 40)$ yielding 15,343 reflexions, giving the number of reflexions, $Nref$, relative to the number of variables, Nv, as $Nref/Nv = 151.9$. The process consists of refining a series of independent parameter sets with the final refinement carried out on non-zero (symmetry allowed) P_{lms} up to $L = 4$ for all atoms and U_{ij}s of heavy atoms simultaneously yielding a very reasonable goodness of fit parameter (R-value) at $R = 0.13\%$. As evidenced by Table 5.1, the refined U_{ij}s reflect the true nuclear motion derived from normal mode analysis.

Table 5.1 True (top) vs refined (bottom) ADPs ($\text{Å}^2 \times 10^6$) for nuclei in formamide.

		U_{11}	U_{22}	U_{33}	U_{12}
$O6$	True	567	403	125	236
	Refined	579	379	92	242
$N3$	True	710	348	1081	200
	Refined	681	340	1037	200
$C2$	True	592	997	693	-72
	Refined	543	942	665	-72
$H1$	True	10401	7150	10636	1993
	Refined	9812	4157	6085	3163
$H4$	True	14315	5750	30386	3475
	Refined	11681	6529	22164	2180
$H5$	True	5334	12758	52542	1920
	Refined	7408	10730	37554	1512

Furthermore, the total ED defined by refined parameters, ρ_{RES}, reproduces the target ED, ρ_Ψ, reasonably well as the difference map in Figure 5.1, $\rho_\Psi - \rho_{RES}$, reveals. The largest absolute error, occurring at the site of the Nitrogen nucleus, is 13.81 $e/\text{Å}^3$ which is only a 1.13% error. The largest contribution to this error is due to the fact that ρ_Ψ is expressed with GTOs and the refined ED, ρ_{RES}, utilizes Slater functions which exhibit much sharper peaks at the nucleus.
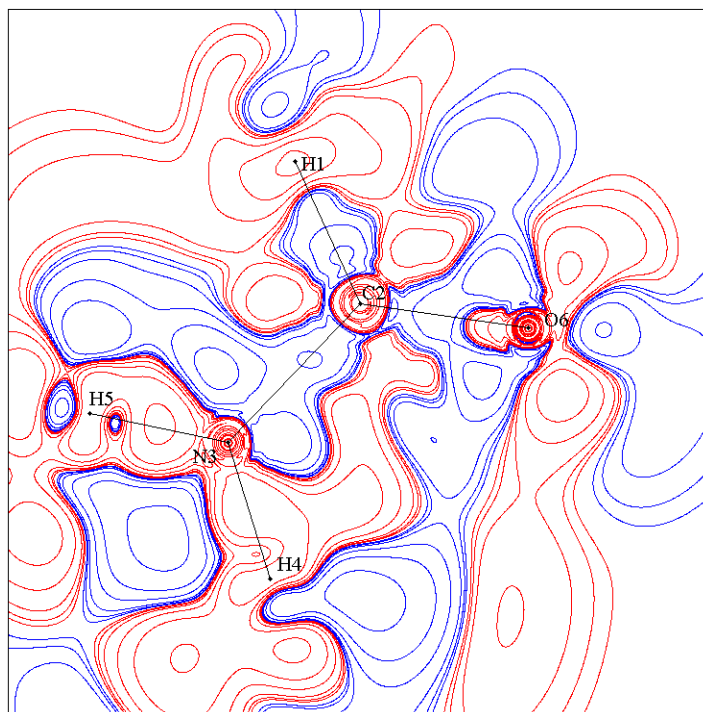
Figure 5.1 Difference map of $\rho_\Psi - \rho_{RES}$ for the formamide molecule in the plane. Contour lines are $\{\pm 2^n\}_{n=1,2,3}\{10^{-m}\}_{m=1,2,3}$ $e/\text{Å}^3$ with positive / negative values expressed as red / blue contours.

5.3.1 Contributions of Uncertainty

The final refinement parameters and covariance matrix $(\boldsymbol{X}, \boldsymbol{V})$ were involved in the UQ according to the methods above. Not surprisingly, EP and SPS methods show almost indistinguishable SUs. What is quite surprising however, is that the DPS method agrees with the two former methods to a remarkably high degree. The remaining part of this discussion compares and contrasts results from the three methods and show that they are all in agreement.

Recall that the SUs of the parameters are used for sampling and the resulting distribution of TPs is involved in the UQ. The reason for the agreement between the EP and SPS methods seems to be obvious; the EP method exactly represents the SU of a linear

function while the SPS method measures the SU in the distribution of a function which is comprised normally distributed linear parameters, so they should (and do) agree. What remains to be explained is why the DPS and SPS methods agree. The reason for this is due to the fact that the SU of the static parameters are orders of magnitudes larger than the SU of the dynamic parameters; dynamic parameters have a (max, mean) SU of $(2.9 \times 10^{-6}, 2.1 \times 10^{-6})$ while static parameters have an (max, mean) SU of $(2.7 \times 10^{-3}, 6.6 \times 10^{-4})$ indicating that the SU of the static parameters "washes out" any SU due to dynamic parameters.

Much of the reasoning behind the larger SUs of the P_{lm}s comes from the high correlation between the parameters in the LS fitting procedure. An analysis of the correlation matrix shows that the largest 12 (P_{lm}, P_{lm}) correlations all have an absolute correlation coefficient of $|r| > 0.5$, the highest of which is between $O_6(P_{3,3})$ and $N_3(P_{3,3})$ at $r = 0.949$. On the contrary, the largest (U_{ij}, U_{ij}) correlation is only the 25th largest in the entire parameter set at $r = -0.482$, the second largest is the 90th in the entire parameter set at $r = -0.315$.

5.3.2 Uncertainty of Topological Properties

The target ED, the ED used in the Gaussian smearing creation of SFs, is a wave function (WFN) representation optimized with Gaussian and is then fitted in direct space by the numerical projection methods and fitting protocol described in Chapter 2 resulting in a $L = 4$ m-dependent PA representation (FPR). These parameters are included as initial values in the refinement and the resulting (RES) density is used in the UQ. To measure the accuracy of the EP method (as well as to validate the SPS method), the uncertainty of ρ

and $\nabla^2\rho$ were evaluated at known BCP locations according to RES expression. Table 5.2 shows the SU of the two TPs using the EP, SPS, and DPS methods evaluated at the BCP locations according to the RES model. The striking similarities between the SUs of different methods show that the EP is quite useful in classifying uncertainties of scalar properties of the ED, even if thermal parameters are ignored.

Table 5.2 Standard Uncertainty of ED ($\times 10^2$, top) and Laplacian (bottom) at each BCP location.

		EP	SPS	DPS
$(H1,C2)$	ED	0.281	0.280	0.281
	Laplacian	0.178	0.179	0.179
$(C2,N3)$	ED	0.641	0.631	0.641
	Laplacian	0.232	0.228	0.232
$(C2,O6)$	ED	0.558	0.552	0.559
	Laplacian	0.161	0.160	0.161
$(N3,H4)$	ED	0.805	0.802	0.795
	Laplacian	1.988	1.987	1.970
$(N3,H4)$	ED	0.373	0.369	0.374
	Laplacian	1.967	1.964	1.960

To quantify the uncertainty of the fitted model due to the location of BCP, a topological analysis was performed on each member of the SPS and DPS data sets. Table 5.3 compares the TPs evaluated at the BCP location by contrasting the WFN, FPR, and RES models as well as the mean and SU of TPs evaluated at BCP locations for each member of the SPS and DPS data sets. One important aspect to note here is that the RES model is competitive with the FPR model for representing the target WFN properties, especially at the $C=O$ BCP location where the FPR Laplacian struggles to accurately represent the target WFN value.

Table 5.3 ED ($e/\text{Å}^3$) and Laplacian ($e/\text{Å}^5$) for WFN, FPR, and RES models. Averaged properties are given for the SPS and DPS methods. SUs are listed in parentheses with SU for ED scaled by $\times 10^2$.

		WFN	FPR	RES	SPS		DPS	
$(H1,C2)$	ED	1.861	1.852	1.878	1.878	(0.28)	1.878	(0.28)
	Laplacian	-22.582	-21.872	-23.702	-23.701	(0.20)	-23.694	(0.20)
$(C2,N3)$	ED	2.128	2.126	2.170	2.170	(0.63)	2.170	(0.64)
	Laplacian	-21.090	-19.302	-23.200	-23.193	(0.44)	-23.172	(0.45)
$(C2,O6)$	ED	2.796	2.770	2.819	2.819	(0.55)	2.818	(0.56)
	Laplacian	-5.200	-1.109	-4.760	-4.760	(0.18)	-4.765	(0.19)
$(N3,H4)$	ED	2.280	2.249	2.216	2.216	(0.81)	2.215	(0.81)
	Laplacian	-39.197	-33.445	-35.881	-35.738	(0.59)	-35.736	(0.57)
$(N3,H4)$	ED	2.297	2.266	2.221	2.219	(0.50)	2.219	(0.51)
	Laplacian	-39.134	-33.250	-32.009	-31.754	(2.80)	-31.918	(2.70)

The difference between the data presented in Table 5.2 and the SUs presented in Table 5.3 is that the former is related to the TPs evaluated at a specific location (the *actual* BCP location of the RES model) whereas the latter is representative of the distribution of TPs according to the BCP location of each member in the SPS and DPS datasets. It is to be mentioned that the direct space fit (FPR) and the reciprocal space fit (RES) seem to perform equally well in reconstructing the target TPs at the BCP locations. For example, the FPR parameters seem to reconstruct the C-N BCP very well, but the RES parameters do a much better job at estimating the properties at the C=O BCP.

The major differences between the data presented in Table 5.2 and that of Table 5.3 appear for the SUs of TPs at the N-H BCPs. This is largely due to the SU in the Cartesian coordinates of the BCPs. Table 5.4 lists the Cartesian coordinates of the BCPs for each model and the sampling methods for UQ along with SUs. Once again, it is to be noted that

the reciprocal space fit (RES) seems competitive with the direct space fit (FPR) as evidenced by the approximated BCP locations.

Table 5.4 Cartesian Coordinates of BCP locations (Å) for WFN, FPR, and RES models. Mean BCP locations are reported for SPS and DPS methods of UQ with standard deviations (reported in parentheses). Z coordinates are omitted because there is no deviation from zero in any of the models due to the planarity of the molecule.

		WFN	FPR	RES	SPS		DPS	
$(H1,C2)$	X	-0.303	-0.300	-0.302	-0.302	(5.4e-04)	-0.302	(5.4e-04)
	Y	1.067	1.054	1.065	1.065	(1.1e-03)	1.065	(1.1e-03)
$(C2,N3)$	X	-0.346	-0.338	-0.340	-0.340	(1.1e-03)	-0.340	(1.1e-03)
	Y	0.054	0.065	0.057	0.057	(1.9e-03)	0.057	(1.9e-03)
$(C2,O6)$	X	0.410	0.407	0.408	0.408	(1.1e-04)	0.408	(1.1e-04)
	Y	0.352	0.353	0.352	0.352	(1.0e-03)	0.352	(1.0e-03)
$(N3,H4)$	X	-0.716	-0.720	-0.734	-0.734	(2.2e-03)	-0.734	(2.2e-03)
	Y	-1.277	-1.269	-1.246	-1.246	(6.9e-03)	-1.247	(6.8e-03)
$(N3,H5)$	X	-1.657	-1.650	-1.602	-1.602	(1.3e-02)	-1.603	(1.3e-02)
	Y	-0.417	-0.419	-0.437	-0.437	(2.9e-03)	-0.436	(3.0e-03)

The SUs listed in Table 5.4 can be a bit misleading because they do not include any information about chemical connectivity. To properly demonstrate how the SUs in BCP locations have an effect on the TPs at the BCPs, each member of the SPS and DPS data sets were projected into a local coordinate system to find the SU along and perpendicular to the bond in which the CP exists. Table 5.5 lists these uncertainties showing that in all but the $C=O$ bond, the SU of the BCP location in the direction of the bond path is greater than it is in a direction perpendicular to the bond path.

Table 5.5 Standard Uncertainties of BCP locations parallel and perpendicular to the bond path in question.

		SPS	DPS
$(H1,C2)$	Parallel	1.184e-03	1.182e-03
	Perpendicular	2.695e-04	2.700e-04
$(C2,N3)$	Parallel	2.077e-03	2.097e-03
	Perpendicular	7.690e-04	7.731e-04
$(C2,O6)$	Parallel	1.315e-04	1.319e-04
	Perpendicular	1.018e-03	1.022e-03
$(N3,H4)$	Parallel	7.174e-03	7.116e-03
	Perpendicular	9.180e-04	9.259e-04
$(N3,H5)$	Parallel	1.366e-02	1.373e-02
	Perpendicular	1.302e-03	1.317e-03

6. Conclusion[3]

An improvement to the standard PA model is found to more accurately reproduce a target ED by means of higher expansion in *Lmax* for multipoles, increased size of the basis function for higher order multipoles and most importantly, introduction of *m*-dependent functions for higher order multipoles.

The isolated atom is spherical but becomes deformed upon bonding which makes RSHs a natural choice for describing the deformation and properly redistributes the charge density to account for bonding effects. The RSHs form a complete orthogonal basis on the surface of the unit sphere so that any function can be expressed as their linear combination in this 2-dimensional space:

$$f(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} c_{lm} d_{lm}(\theta, \phi) \tag{6.1}$$

When the radial dimension is added to describe a function in 3-space, the mixing coefficients, c_{lm}, become functions of r

$$\rho(r, \theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} c_{lm}(r) d_{lm}(\theta, \phi) \tag{6.2}$$

which are manifested as Radial Density Functions, $\rho_{lm}(r) = \langle \rho_{tot} | d_{lm} \rangle_{\Omega}$, in the PA model. Even in the 2-dimensional case, they form a complete basis only in the infinite expansion so any set of basis functions which are truncated too soon introduce non-negligible error. Finally, the terms which describe the deformation are quite complex and cannot be accurately described with a single Slater function.

---

[3] Portions of this conclusion also appear in The Journal of Mathematical Chemistry [79].

A new formalism has been put forward [6] to account for these inadequacies and the sources of error (the finite expansion in *Lmax m*-dependence and the inability to perfectly fit with analytic functions) has been analyzed to partition/evaluate the error. It is shown that *m*-dependent RDFs are necessary and can, in the infinite limit of the expansion, perfectly reconstruct any target ED but that great care must be taken in fitting these *m*-dependent RDFs with analytic functions so as not to introduce further artificial error.

While the above seeks to amend issues in representation of the static ED, a computational framework has been built for the analysis of thermal motion effects resulting in the dynamic ED. The nuclear ADPs are used to approximate the ADPs of bond critical points and show that the approximations are very sensitive to correlation effects between atoms. From these computational experiments we see that the location of BCPs are normally distributed (much like the nuclear coordinates) but that the topological properties are not normally distributed. Moreover, we find that the thermally averaged topological property is not necessarily the same as the static topological property and that any discrepancy may be caused by the sensitivity of the property to more than one normal mode of vibration.

A straightforward formalism is also presented to calculate the dynamic molecular ED from that of the stationary ED within the harmonic convolution approximation and the LCAO-MO formalism utilizing Gaussian basis functions. In spite of its simplicity, the method turns out to provide a very accurate estimation of the thermally smeared ED ($\overline{\rho}$) obtained by a more rigorous and computationally much more demanding procedure that involves averaging over a large ensemble of single-point static EDs. This is well substantiated by the closeness of the BCP values obtained for the two dynamic EDs in

Table 4.1. Even more surprisingly, the static and convoluted dynamic EDs are found to exhibit a high degree of similarity in terms of BCP properties. However, for molecules containing hydrogen atoms, the two EDs do not necessarily reveal a topological equivalence. The convoluted ED for the formamide molecule, for example, exhibits no BCP for the *N-H* bonds because the hydrogen density peaks (the (3,+3) critical points of the static ED) are completely 'washed out' by thermal vibrations (became saddle points) even at such low temperature as 23 K at which only the ground vibration states are populated. This result is of significance to the reliability of X-Ray charge density based topological analysis of hydrogen bonds. The majority of these studies rely on higher-temperature diffraction data (typically around 100K) and the data interpretation utilizes either isotropic temperature factors for the H-atoms [73, 74] or (preferably) ADPs from independent observations / calculations [75, 76], making thus the experimental static ED for both covalent and non-covalent interactions involving H-atoms decisively model dependent. While this issue is quite widely recognized, the invisibility of the *N-H* bond path for the convoluted theoretical ED, as shown in this study, indirectly implies that the results of topological analysis of experimental EDs of bonding involving H-atoms is not just simply supported, but entirely settled by the choice of H-atom ADPs and mean positions. To be more precise, the 'experimental static' topology of H-bonding is a direct manifestation of parameters not observable by X-Ray diffraction. To reach this conclusion one should just consider what it would take to solve the inverse problem, that is, to reconstruct fine topological details of the static ED for the *N-H* bonds from that of the convoluted dynamic ED for formamide.

It is to be emphasized that our analysis is restricted to internal (intra-molecular) vibrations of relatively high frequencies and thus low amplitudes. The inclusion of external modes (translation and rotation within the molecular mean field model [77]) or acoustic modes (within lattice dynamics [3]) is most likely to further demolish characteristic features of the static topology. A more general conclusion is that smearing the theoretical ED to compare it with the experimental ED deserves more attentions than it has received during the most recent X-Ray charge density era. After all, the molecular static ED is not a direct observable.

In the practice of fitting a static ED model to observed dynamic SFs, it is standard procedure in the modern era of Charge Density analysis to invoke the method of error propagation. Error propagation is applied to static parameters of the multipole model to quantify uncertainty of TPs of the static ED. It has not been investigated, until now, what effect the SUs of the dynamic parameters of the associated dynamic ED play in the SU of the static ED. The contribution of the SU of the dynamic parameters to the static ED has been almost impossible to diagnose as the dynamic parameters do not appear in the static ED.

The only way to elucidate the effect of the dynamic parameters on the static ED is to perform a computational experiment where dynamic parameters are sampled from their standard uncertainties and their effect seen in the static parameters. Upon successful completion of the experiment it is found that, indeed, the uncertainty of the dynamic parameters can be ignored in the UQ of the static ED.

7. Contributions and Achievements

The research presented in this dissertation made extensive use of computational software to achieve results which would not have been possible in the absence of a High Performance Computing (HPC) framework. In the course of this research substantially high cost computation were involved which pushed the available hardware to their limits. This hardware includes a personal laptop ("Flexo") which boasts a quad core Intel i7 2.4 GHz cpu with 20 GB of 1600 MHz RAM, 2 TB of HD space (7200 RPM) and an Nvidia GeForce 770M GPU with 3 GB GDDR5 RAM which was primarily used for prototyping Fortran, Python, and C / C++.

Upon successful compilation and prototyping, the source code was executed on a desktop ("Pauli") which includes a 6 core 3.0 GHz AMD Phenom II cpu, and 8 GB of RAM. When the source code required it, parallelism was employed via "Supernode" on the Babbage cluster which utilizes a 48 core 2.2 GHz AMD Opteron cpu and 256 GB RAM, as well as "MrWhite" on the Rainbow cluster which has available a hyperthreading 16 core 3.1 GHz Intel Xeon cpu and 64 GB of RAM.

As ongoing research in this field continues at the university others will continue to make use of algorithms, source code, and discoveries from those who have gone before them. Contributions to source code undergoing development at the university are listed here.

7.1 Efit

The Efit application (which fits aRDFs to nRDFs as described in Chapter 2) is a standalone program written in Fortran (5,000 lines of code) which makes use of the output

of a previously written program 'Radfun' (written by Dr. Anatoliy Volkov and others). The Efit application performs an efficient search in parallel and writes fitted aRDFs to a file in a format which can be read by Denprop.

7.2 Denprop

Denprop has many contributing authors and in recent years additions have been made to further the analysis of static and dynamic EDs. The current version of Denprop's source code is 171,000 lines of which 110,000 have been contributed for the sake of the research discussed in this dissertation.

7.2.1 Normalization Coefficients for Real Spherical Harmonics

As described in [23, 37], density normalized Spherical Harmonics are needed for utilizing normalized basis functions in the PA expansion since the RDFs are only radial in nature and rely on (normalized) RSHs for accurate description of the static ED. Normalization coefficients for RSHs have been computed numerically using Mathematica [78] as well as in Fortran using a Gauss-Legendre quadrature scheme for pre-calculation of coefficients up to $Lmax = 40$. These coefficients, in collaboration of Dr. Anatoliy Volkov, were written to subroutines used by Denprop contained in the 'Ylm_JRM_lmax06.f90' and 'Ylm_JRM_lmax20.f90' Fortran files which together comprise about 95,000 lines of code.

7.2.2 Electron Density Calculations via Numerical Radial Functions

The projection of a target ED onto RSHs yielding nRDFs takes place in a program called 'Radfun' which has been written in Fortran by Dr. Anatoliy Volkov and others. The output file of this program expresses nRDFs on a fine grid. New subroutines, contained in the file name 'nradfun.f90', have been written to calculate the ED and related TPs from

these grid based RDFs by interpolation and are now a part of the Denprop package. These subroutines use cubic spline interpolation to compute the ED between predefined numerical gridpoints. For points outside the domain of the radial grid, linear extrapolation is used. Cubic spline interpolation involves the calculation of numerical derivatives which are calculated with $O(h^4)$ accuracy.

7.2.3 Dynamic Electron Density Calculations

A new input file containing information about the MSDA of a molecular structure is needed to calculate the dynamic ED of a wave function based ED model. Via implementation of methods described in Chapter 4, the Gaussian basis functions are smeared so as to yield a dynamic ED which Denprop is then able to perform a thorough topological analysis on. These methods are also applied to analytic calculation of theoretical SFs which can then be used in the XD program suite. The files which are involved in these calculation include 'dynamic.f90' and 'g_grad.f90', containing a combined 12,000 lines of source code, which make use of precalculated derivatives of thermally smeared GTOs as computed analytically by Mathematica.

7.2.4 Uncertainty Quantification

Upon input of a new file containing information about the covariance matrix of the P_{lm}s for the PA model, error propagation is implemented in Denprop so that SUs of ρ and $\nabla^2 \rho$ can be printed. Currently the SUs are printed for calculation at a point and for a BCP search, but subroutines in 'uncertainty.f90' include the ability to calculate the SUs for any point. As such, further development can make use of this subroutine so as to print SUs for any type of search (profile, 2d / 3d grid calculations, etc).

7.2.5 Fourier Space Convolution

Denprop is able to produce Structure Factors (SF) analytically when a GTO based model density is desired. The new methods described in Chapter 4 are implemented in the software both for direct and reciprocal space dynamic density within the convolution approximation. To validate that the convoluted dynamic SFs, $\langle F_{hkl} \rangle_{\circledast}$, are representative of the statistical average, a separate Fortran program, "smear_SFs", was written to calculate the statistical dynamic SFs, $\langle F_{hkl} \rangle_{\mu}$, as the average magnitudes over all 500,000 members of the ensemble for a given reflection, hkl, along with their standard uncertaintites.

We wish to calculate the Sum of Squares Error (SSE) for the difference between $\langle F_{hkl} \rangle_{\circledast}$ and $\langle F_{hkl} \rangle_{\mu}$ where the summation takes place over reflections, hkl. For a given reflection $\langle F \rangle_{\mu} := \langle F_{hkl} \rangle_{\mu}$ is not uniquely defined by the members of the ensemble $\{F_i\}$ since $\langle F \rangle_{\mu}$ can be regarded as the average over magnitudes $F_1 = M^{-1} \sum F_i$ or as the magnitude of the average complex number $F_2 = M^{-1} \sqrt{(\sum A_i)^2 + (\sum B_i)^2}$ which are not equal. As it turns out, the SSE is smaller for F_2 (0.11), than it is for F_1 (186.5) but the generation of F_1, F_2, and their standard uncertainties are computationally intensive.

For each of the 500,000 members in the ensemble of direct space EDs, $\{\rho_i\}$, the analytic SFs were calculated with Denprop. The resulting reciprocal space ensemble, $\{F_i\}$, was 1.2 TB of HD space as there were 13,943 reflections needed in the analysis. Extrapolation on a naïve implementation of the smear_SFs application showed that the generation of F_2 would take 88 years. The bottleneck in the application was file I/O, so data reduction on the reciprocal space ensemble reduced the hard disk space to 350 GB. Babbage's "Supernode" has 256 GB of RAM available, so the internal floating point

variables were converted from double to single precision, reducing the volatile memory footprint to only 105 GB resulting in only a 79 hour execution time.

7.3 Publications and Presentations

7.3.1 Gordon Research Conference

June, 2013: Les Diablerets, Switzerland

Oral Presentation: "A Topological Analysis of Thermally Averaged Density"

7.3.2 European Charge Density Meeting

September, 2012: High Tatras, Slovakia

Poster Presentation: "Direct Space Reconstruction of Electron Density"

7.3.3 Synchrotron Charge Density School

March, 2013: Argonne National Labe, Lemont, IL

Oral Presentation: "Radfun and Efit: A Convergent Expansion"

7.3.4 Tennessee Academy of Science Meeting

November, 2013: Tullahoma, TN

Poster Presentation: "A Novel Approach to the Pseudoatom Expansion"

7.3.5 Publications

J. R. Michael and T. Koritsanszky, "Validation of convolution approximation to the thermal-average electron density," *J. Math. Chem.,* 2014. DOI: 10.1007/s10910-014-0425-y [79]
J. R. Michael and T. Koritsanszky, "Computational Study of Uncertainties of Topological properties of the Thermally Smeared Molecular Electron Density," *Acta Cryst. (to be published)* [70]
J. R. Michael and A. Volkov, "Density and wavefunction normalized spherical harmonics in Cartesian frame for $0 \leq l \leq 20$", *Acta Cryst. (to be published)* [80]

8. References

[1] M. Born and J. R. Oppenheimer, "Zur Quantentheorie der Molekeln," *Ann. Physik,* vol. 389, no. 20, pp. 457-484, 1927.

[2] C. Pisani, A. Erba, M. Ferrabone and R. Dovesi, "Nuclear motion effects on the density matrix of crystals: An ab initio Monte Carlo harmonic approach," *J. Chem. Phys.,* vol. 137, p. 044114, 2012.

[3] A. O. Madsen, B. Civalleri, M. Ferrabone, F. Pascale and A. Erba, "Anisotropic displacement parameters for molecular crystals from periodic Hartree-Fock and density functional theory calculations," *Acta. Cryst. A,* vol. 69, no. 3, pp. 309-321, 2013.

[4] J. M. Gillet and T. Koritsanszky, "Past Present and Future of Charge Density and Density Matrix Refinements," in *Modern Charge-Density Analysis*, New York, Springer, 2012, pp. 181-211.

[5] D. Jayatilaka and B. Dittrich, "X-ray structure refinement using aspherical atomic density functions obtained from quantum-mechanical calculations," *Acta. Cryst. A,* vol. 64, no. 3, pp. 383-393, 2008.

[6] T. Koritsanszky, A. Volkov and M. Chodkiewicz, "New Directions in Pseudoatom-Based X-Ray Charge Density Analysis," *Struct. Bond.,* vol. 147, pp. 1-25, 2012.

[7] A. Altomare, C. Cuocci, C. Giacovazzo, A. Moliterni and R. Rizzi, "Covariance and correlation estimation in electron-density maps," *Acta. Cryst. A,* vol. 68, no. 2, pp. 244-255, 2012.

[8] P. Parois and M. Lutz, "Linear transformations of variance/covariance matrices," *Acta. Cryst. A,* vol. 67, no. 4, pp. 383-390, 2011.

[9] C. Gatti and P. Macchi, "A Guided Tour Through Modern Charge Density Analysis," in *Modern Charge-Density Analysis*, C. Gatti and P. Macchi, Eds., New York, Springer Science+Business Media B.V., 2012, pp. 1-77.

[10] R. F. Bader, *Atoms in Molecules*, A Quantum Theory, New York: Oxford University Press, Inc., 1990.

[11] C. D. Sherrill, *The Born-Oppenheimer Approximation,* Georgia Institute of Technology: School of Chemistry and Biochemistry, 2005.

[12] P. Coppens, *X-Ray Charge Densities and Chemical Bonding*, New York: Oxford University Press, Inc., 1997.

[13] A. O. Madsen, "Modeling and Analysing Thermal Motion in Experimental Charge Density Studies," in *Modern Charge Density Analysis*, New York, Springer, 2012, pp. 133-163.

[14] G. H. Stout and L. H. Jensen, *X-Ray Structure Determination*, 2nd Ed., New York: John Wiley & Sons, Inc., 1989.

[15] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, *Gaussian09, Revision A.1,* Wallingford: Gaussian, Inc., 2009.

[16] W. J. Hehre, L. Radom, P. v. Schleyer and J. A. Pople, *Ab Initio Molecular Orbital Theory*, New York: Wiley-Interscience, 1986.

[17] D. Jayatilaka and D. J. Grimwood, "Wavefunctions derived from experiment. I. Motivation and theory," *Acta Cryst. A,* vol. 57, no. 1, pp. 76-86, 2001.

[18] D. Grimwood and D. Jayatilaka, "Wavefunctions derived from experiment. II. A wavefunction for oxalic acid dihydrate," *Acta Cryst. A,* vol. 57, no. 1, pp. 87-100, 2001.

[19] I. Bytheway, D. Grimwood and D. Jayatilaka, "Wavefunctions derived from experiment. III. Topological analysis of crystal fragments," *Acta Cryst. A,* vol. 58, no. 3, pp. 232-243, 2002.

[20] I. Bytheway, D. J. Grimwood, B. N. Figgis, G. S. Chandler and D. Jayatilaka, "Wavefunctions derived from experiment. IV. Investigation of the crystal environment of ammonia," *Acta Cryst. A,* vol. 58, no. 3, pp. 244-251, 2002.

[21] C. Giacovazzo, "Crystallographic Computing," in *Fundamentals of Crystallography*, New York, Oxford University Press Inc., 1992, pp. 61-140.

[22] R. F. Stewart, "One-electron density functions and many-centered finite multipole expansions," *Isr. J. Chem,* vol. 16, no. 2-3, pp. 124-132, 1977.

[23] N. K. Hansen and P. Coppens, "Testing aspherical atom refinements on small-molecule data sets," *Acta Cryst. A,* vol. 34, no. 6, pp. 909-921, 1978.

[24] R. F. Stewart and D. Feil, "A theoretical study of elastic X-ray scattering," *Acta Cryst. A,* vol. 36, no. 4, pp. 503-509, 1980.

[25] P. W. Higgs, "Vibrational modifications of the electron distribution in molecular crystals. II. Mean square amplitudes of thermal motion," *Acta Cryst.,* vol. 8, no. 2, pp. 99-104, 1955.

[26] R. F. Stewart, "Vibrational averaging of X-ray scattering intensities," *Isr. J. Chem.,* vol. 16, no. 2-3, pp. 137-143, 1997.

[27] P. Debye, "Interferenz von Röntgenstrahlen und Wärmebewegung," *Ann. Phys.,* vol. 348, no. 1, pp. 49-92, 1913.

[28] I. Waller, "Zur Frage der Einwirkung der Wärmebewegung auf die Interferenz von Röntgenstrahlen," *Z. Phys. A,* vol. 17, no. 1, pp. 398-408, 1923.

[29] C. Scheringer, "Temperature factors for internuclear density units. I. Theory in the harmonic approximatio," *Acta. Cryst. A,* vol. 33, no. 3, pp. 426-429, 1977.

[30] P. Coppens, T. V. Willoughby and L. N. Csonka, "Electronic population analysis of accurate diffraction data. I. Formalism and restrictions," *Acta Cryst. A,* vol. 27, no. 3, pp. 248-256, 1971.

[31] K. Tanaka, "X-ray analysis of wavefunctions by the least-squares method incorporating orthonormality. I. General formalism," *Acta Cryst. A,* vol. 44, no. 6, pp. 1002-1008, 1988.

[32] R. F. Stewart, "Generalized X-ray scattering factors," *J. Chem. Phys.,* vol. 51, pp. 4569-4576, 1968.

[33] A. Volkov, P. Macchi, L. Farrugia, C. Gatti, P. Mallinson, T. Richter and T. Koritsanszky, "XD 2006 - A Computer Program Package for Multipole Refinement, Topological Analysis of Charge Densities and Evaluation of Intermolecular Energies from Experimental and Theoretical Structure Factors," 2006.

[34] J. Tellinghuisen, "Statistical error propagation," *J. Phys. Chem. A,* vol. 105, pp. 3917-3921, 2001.

[35] T. Koritsanzksy, S. Howard, P. Macchi, C. Gatti, L. J. Farrugia, P. R. Mallinson, A. Volkov, Z. Su, T. Richter and N. K. Hansen, "XD Manual," 2004.

[36] P. R. Mallinson, T. Koritsanzky, E. Elkaim, N. Li and P. Coppens, "The Gram-Charlier and multipole expansions in accurate X-ray diffraction studies: Can they be distinguished?," *Acta. Cryst. A,* vol. 44, no. 3, pp. 336-342, 1987.

[37] A. Paturle and P. Coppens, "Normalization factors for spherical harmonic density functions," *Acta Cryst. A,* vol. 44, no. 1, pp. 6-8, 1988.

[38] J. R. Michael, *Prequalifying Exam,* Murfreesboro: Unpublished, 2010.

[39] E. Clementi and C. Roetti, "Roothaan-Hartree-Fock atomic wavefunctions: Basis functions and their coefficients for ground and certain excited states of neutral and ionized atoms," *At. Data Nuc. Data Tables,* vol. 14, no. 3-4, pp. 177-478, 1974.

[40] J. C. Slater, "Atomic shielding constants," *Phys. Rev.,* vol. 36, pp. 57-74, 1930.

[41] A. Volkov, M. Chodkiewicz and T. Koritsanszky, "Radfun", 2012.

[42] F. L. Hirshfeld, "Bonded-atom fragments for describing molecular charge densities," *Theo. Chem. Acc.,* vol. 44, no. 2, pp. 129-138, 1977.

[43] A. D. Becke, "Density-functional exchange-energy approximation with correct asymptotic behavior," *Phys. Rev. A,* vol. 38, pp. 3098-3100, 1988.

[44] G. V. Gibbs, R. T. Downs, D. F. Cox, K. M. Rosso, N. L. Ross, A. Kirfel, T. Lippmann, W. Morgenroth and T. D. Crawford, "Experimental bond critical point and local energy density properties determined for Mn-O, Fe-O, and Co-O bonded interactions for tephroite, Mn2SiO4, fayalite, Fe2SiO4, and Co2SiO4 olivine and selected organic metal complexes: Comparison with properties calculated for non-transition and transition metal M-O bonded interactions for silicates and oxides," *J. Phys. Chem. A,* vol. 112, no. 37, pp. 8811-8823, 2008.

[45] G. V. Gibbs, F. C. Hill, M. B. Boisen and R. T. Downs, "Power law relationships between bond length, bond strength, and electron density distributions," *Phys. Chem. Miner.,* vol. 25, no. 8, pp. 585-590, 1998.

[46] G. V. Gibbs, A. E. Whitten, M. A. Spackman, M. Stimpfl, R. T. Downs and M. D. Carducci, "An exploration of theoretical and experimental electron density distributions and SiO bonded interactions for the silica polymorph coesite," *J. Phys. Chem. B,* vol. 107, no. 47, pp. 12996-13006, 2003.

[47] I. D. Brown and R. D. Shannon, "Empirical bond-strength-bond-length curves for oxides," *Acta Cryst. A,* vol. 29, no. 3, pp. 266-282, 1973.

[48] C. Scheringer, "Temperature factors for internuclear density units. II. Considerations with respect to experimental accuracy," *Acta Cryst. A,* vol. 33, no. 3, pp. 430-433, 1977.

[49] J. W. Ochterski, *Vibrational Analysis in Gaussian,* Gaussian, Inc., 1999.

[50] M. N. Burnett and C. K. Johnson, "ORTEP-III: Oak Ridge Thermal Ellipsoid Plot Program for Crystal Structure Illustrations," Oak Ridge National Laboratory Report ORNL-6895, Oak Ridge, 1996.

[51] A. Volkov, T. Koritsanszky, M. Chodkiewicz and H. F. King, "On the basis-set dependence of local and integrated electron density properties: Application of a new computer program for quantum-chemical density analysis," *J. Comp. Chem.,* vol. 30, no. 9, pp. 1379-1391, 2009.

[52] T. S. Koritsanszky and P. Coppens, "Chemical applications of X-ray charge-density analysis," *Chem. Rev.,* vol. 101, no. 6, pp. 1583-1628, 2001.

[53] C. Gatti, "Chemical bonding in crystals: New directions," *Z. f. Krist.,* vol. 220, no. 5-6, pp. 339-487, 2005.

[54] F. L. Hirshfeld, "Charge deformation and vibrational smearing," *Isr. J. Chem,* vol. 16, no. 2-3, pp. 168-174, 1977.

[55] R. Flaig, T. Koritsanszky, D. Zobel and P. Luger, "Topological analysis of experimental electron densities of amino acids: 1. D,L-Aspartic Acid at 20K," *J. Am. Chem. Soc,* vol. 120, no. 10, pp. 2227-2236, 1998.

[56] E. D. Stevens, J. Rys and P. Coppens, "Calculation of dynamic electron distributions from static molecular wavefunctions," *Acta Cryst. A,* vol. 33, no. 2, pp. 333-338, 1977.

[57] E. D. Stevens, J. Rys and P. Coppens, "Electron density distribution of the azide ion. Quantitative comparison of theoretical calculations with experimental measurements," *J. Am. Chem. Soc.,* vol. 99, no. 1, pp. 265-267, 1977.

[58] S. Howard, J. P. Huke, P. R. Mallinson and C. S. Frampton, "Density-matrix refinement for molecular crystals," *Phys. Rev. B,* vol. 49, no. 11, pp. 7124-7136, 1994.

[59] H. Schmider, V. H. Smith Jr. and W. Weyrich, "Reconstruction of the one-particle density matrix from expectation values in position and momentum space," *J. Chem. Phys.,* vol. 96, pp. 8486-8994, 1992.

[60] M. Born and K. Huang, *Dynamical Theory of Crystal Lattices*, London: Oxford Universty Press, 1954.

[61] C. A. Coulson and M. W. Thomas, "The effect of molecular vibrations on apparent bond lengths," *Acta Cryst. B,* vol. 27, no. 7, pp. 1354-1359, 1971.

[62] A. F. J. Ruysink and A. Vos, "Theoretical calculation of the time-averaged electron density distribution for vibrating ethyne molecules in a model crystal structure," *Acta Cryst. A,* vol. 30, no. 4, pp. 497-502, 1974.

[63] M. Messeschmidt, S. Scheins and P. Luger, "Charge density of (-)-strychnine from 100 to 15 K, a comparison of four data sets," *Acta Cryst. B,* vol. 61, no. 1, pp. 115-121, 2005.

[64] J. Oddershede and S. Larsen, "Charge density study of naphthalene based on X-ray diffraction data at four different temperatures and theoretical calculations," *J. Phys. Chem. A,* vol. 108, no. 6, pp. 1057-1063, 2004.

[65] R. Destro, L. L. Presit, R. Soave and A. E. Goeta, "Multi-temperature Electron Density Studies," in *Modern Charge Density Analysis*, C. Gatti and P. Macchi, Eds., New York, Springer Science+Business Media B.V. 2012, 2012, pp. 659-696.

[66] M. Born and V. A. Fock, "Beweis des Adiabatensatzes," *Z. f. Phys. A,* vol. 51, no. 3-4, pp. 165-180, 1928.

[67] S. F. Boys, "Electronic wave functions. I. A general method of calculation for the stationary states of any molecular system," *Proc. R. Soc. Lond. A,* vol. 200, pp. 542-554, 1950.

[68] C. Møller and M. S. Plesset, "Note on an approximation treatment for many-electron systems," *Phys. Rev,* vol. 46, pp. 618-622, 1934.

[69] E. D. Stevens, J. Rys and P. Coppens, "Quantitative comparison of theoretical calculations with the experimentally determined electron density distribution of formamide," *J. Am. Chem. Soc.,* vol. 100, no. 8, pp. 2324-2328, 1978.

[70] J. R. Michael and T. Koritsanszky, "Computational study of uncertainties of topological properties of the thermally smeared molecular electron density," *Acta Cryst. (to be published).*

[71] A. Hofmann, J. Netzel and S. van Smaalen, "Accurate charge density of trialanine: A comparison of multipole formalism and the maximum entropy method," *Acta Cryst. B,* vol. 63, no. 2, pp. 285-295, 2007.

[72] K. O. Arras, "An introduction to error propagation: derivation, meaning and examples," Swiss Federal Insitute of Technology Lausanne, Zurich, 1998.

[73] E. Espiosa, E. Mollins and C. Lecomte, "Hydrogen bond strengths revealed by topological analyses of experimentally observed electron densities," *Chem. Phys. Lett.,* vol. 285, no. 3-4, pp. 170-173, 1988.

[74] P. Munshi and T. N. Guru Row, "Intra- and intermolecular interactions in small bioactive molecules: Cooperative features from experimental and theoretical charge-density analysis," *Acta Cryst. B,* vol. 62, no. 4, pp. 612-626, 2006.

[75] A. O. Madsen, H. O. Sorensen, C. Fensburg, R. Stewart and S. Larsen, "Modeling of the nuclear parameters of H atoms in X-ray charge density studies," *Acta Cryst. A,* vol. 60, no. 6, pp. 550-561, 2004.

[76] P. Roversi and R. Destro, "Approximate anisotropic displacement parameters for H atoms in molecular crystals," *Chem. Phys. Lett.,* vol. 386, no. 4-6, pp. 472-478, 2004.

[77] H. B. Bürgi and S. C. Capelli, "Dynamics of molecules in crystals from multi-temperature anisotropic parameters. I. Theory," *Acta Cryst. A,* vol. 56, no. 5, pp. 403-412, 2000.

[78] Wolfram Research, Inc., *Mathematica,* Champaign: Wolfram Research, Inc., 2014.

[79] J. R. Michael and T. Koritsanszky, "Validation of convolution approximation to the thermal-average electron density," *J. Math. Chem.,* 2014. DOI: 10.1007/s10910-014-0425-y

[80] J. R. Michael and A. Volkov, "Density and wavefunction normalized spherical harmonics in Cartesian frame for $0 \leq l \leq 20$", *Acta Cryst. (to be published)*