A QUANTUM CRYSTALLOGRAPHIC APPROACH TO STUDY PROPERTIES OF MOLECULES IN CRYSTALS

Inaugural dissertation
of the Faculty of Science,
University of Bern

presented by

Michelle Ernst
from Winterthur

Supervisor of the doctoral thesis:

Prof. Piero Macchi
Department of Chemistry and Biochemistry

A QUANTUM CRYSTALLOGRAPHIC APPROACH TO STUDY PROPERTIES OF MOLECULES IN CRYSTALS

Inaugural dissertation
of the Faculty of Science,
University of Bern

presented by

Michelle Ernst
from Winterthur

Supervisor of the doctoral thesis:

Prof. Dr. Piero Macchi
Department of Chemistry and Biochemistry

Accepted by the Faculty of Science.

Bern, 14.02.2020

The Dean
Prof. Dr. Zoltan Balogh

Acknowledgements

I would like to thank:

- **Piero Macchi** for everything you taught me and the support you gave me during my PhD but also in finding the best way for my scientific future.
- **Rebecca Scatena** and **Fabio Montisci** for all the amazing times inside and outside our office, your support and your sense of (in)justice. You are great!
- **Jürg Hauser** for the support with computers and for being a role model in terms of work-life balance.
- **Tomasz Poręba** for all your funny and crazy questions and **Michał Andrzejewski** for your uncomplicatedness.
- **Stefano Racioppi** for convincing people to stay at parties and **Arianna Lanza** for your thoughtfulness.
- **Leonardo Dos Santos** for introducing me to this research field and everything I learnt from you.
- **Alessandro Genoni** for the very warm welcome in Metz and spending a lot of time for me.
- **Simon Grabowsky** and **Florian Kleemiß** for hosting me in their group during the last months of my doctorate and for arriving in Bern with a lot of openness.
- **Georgia Cametti** for your helpfulness and the wonderful cards. **René Schliemann** for your friendliness and humor.
- All the people from the enormous Italian-Bernese community: **Francesco**, **Giuseppe**, **Fulvio**, **Valentine**, **Pavel**, **Giovanni**, **Matteo**, **Alessandro**, etc. for causing that I learnt a new language during my PhD and all the wonderful times.

- My coworkers: **Anna Krawczuk**, **Niccolò Riccardi**, **Tomasz Wesołowski**, and **Nicola Casati** for collaborating.
- **Paulina Dominiak** for reading and judging this dissertation and for demonstrating that in science there is also space for calm women.
- **Samuel Leutwyler** for accepting to chair the defence session of this dissertation and for piquing my interest in quantum chemistry.
- And last but absolutely not least, my family for being on my side and **Lars Gnägi** for your support concerning my PhD and life in general.

Bern, towards the end of 2019 Michelle Ernst

Our group during a typical activity

Abstract

English
In this dissertation, the behaviour of atoms, bonds, functional groups and molecules *in vacuo* but especially also in the crystal is studied using quantum crystallographic methods. The goal is to deepen the understanding of the properties of these building blocks as well as of the interactions among them, because good comprehension of the microscopic units and their interplay also enables us to explain the macroscopic properties of crystals.

The first part (chapters 1-3) and second part (chapter 4) of this dissertation contain theoretical introductions about *quantum crystallography*. On the one hand, this expression contains the term *quantum* referring to quantum chemistry. Therefore, the very first chapter gives a brief overview about this field. The second chapter addresses different options to partition quantum chemical entities, such as the electron density or the bonding energy, into their components. On the other hand, quantum crystallography consists obviously of the *crystallographic* part and chapter 3 covers these aspects focusing predominantly on X-ray diffraction. A more detailed introduction to quantum crystallography itself is presented in the second part (chapter 4).
The third part (chapters 5-9) starts with an overview of the goals of this work followed by the results organized in four chapters.

The goal is to deepen the understanding of properties of crystals by theoretically analysing their building block. It is for example studied how electrons and orbitals rearrange due to the electric field in a crystal or how high pressure leads to the formation of new bonds. Ultimately, these findings shall help to rationally design materials with desired properties such as high refractive index or semiconductivity.

Deutsch
Mithilfe quantenkristallografischer Methoden werden Atome, Bindungen, funktionellen Gruppen und Moleküle *in vacuo* aber vor allem auch in Kristallen untersucht. Das Ziel ist es die Eigenschaften dieser Bestandteile zu verstehen und wie sie miteinander interagieren. Das Verständnis der Verhaltensweise der einzelnen Bausteine sowie deren Zusammenspiel auf mikroskopischer Ebene kann auch die makroskopischen Eigenschaften von Kristallen erklären.

Abstract

Der erste Teil dieser Doktorarbeit (Kapitel 1-3) beinhaltet eine theoretische Einleitung in die verschiedenen Bereiche der Quantenkristallografie. Wie der Name *Quantenkristallografie* besagt, besteht diese zum einen aus dem *quantenchemischen* Teil, weswegen das erste Kapitel eine kurze Einführung in die Quantenchemie gibt. Das zweite Kapitel widmet sich den verschiedenen Möglichkeiten quantenchemische Grössen wie zum Beispiel die Elektronendichte oder Bindungsenergien in Einzelteile zu zerlegen. Zum anderen trägt der *kristallografische* Teil zur Quantenkristallografie bei. Kapitel drei besteht daher aus einem kurzen Überblick über die Kristallografie mit Fokus auf der Röntgenbeugung.
Anschliessend folgt im zweiten Teil (Kapitel 4) eine ausführlichere Einleitung in die Quantenkristallografie selbst.
Der dritte Teil (Kapitel 5-9) beginnt mit einer kurzen Übersicht über die Ziele dieser Arbeit worauf die Resultate, gegliedert in vier verschiedene Kapitel, folgen.

Das Ziel dieser Arbeit ist es die Eigenschaften von Kristallen besser zu verstehen, indem man ihre Einzelteile theoretisch analysiert und mit verschiedenen Methoden rationalisiert. Beispielsweise wird untersucht wie sich Elektronen und Orbitale aufgrund des elektrischen Feldes in Kristallen neu anordnen oder wie unter hohem Druck Bindungen neu geformt werden. Schlussendlich können all diese Erkenntnisse helfen, Materialien mit spezifischen gewünschten Eigenschaften herzustellen.

Français
Les atomes, les liaisons entre eux, les groupes fonctionnels et les molécules sont examinés en utilisant des méthodes de la cristallographie quantique. Le but est de comprendre les propriétés de ces composants et comment ils interagissent *in vacuo* mais surtout aussi dans les cristaux. En comprenant leurs caractéristiques et interactions au niveau microscopique, on peut aussi rationaliser les propriétés macroscopiques des cristaux.

La première partie (chapitres 1-3) de cette thèse de doctorat contient une introduction brève à la *cristallographie quantique*. Comme le nom l'indique, ce domaine de recherche est composé de la chimie quantique et la cristallographie. Pour cette raison le premier chapitre donne une introduction à la chimie quantique. Le deuxième chapitre présente quelques méthodes de décomposition des quantités de la chimie quantique comme la densité électronique ou l'énergie de liaison. Le troisième chapitre couvre la partie cristallographique.
Ensuite dans la deuxième partie (chapitre 4) une introduction plus détaillée sur la cristallographie quantique elle-même est donnée.
La troisième partie (chapitres 5-9) commence par un aperçu des objectives de cette dissertation suivis des résultats structurés en quatre chapitres.

Le but est de comprendre les propriétés des cristaux en analysant leurs *building blocks* avec différentes méthodes théoriques. Il était par example examiné comment les électrons et les orbitales se réorganisent dans un cristal à cause du champ électrique ou comment des nouvelles liaisons sont formées sous pression. Finalement on peut utiliser ces conclusions pour modeler des matériaux avec des propriétés désirées.

Contents

Acknowledgements i

Abstract iii

I Introduction: Theory of Quantum Chemistry and Crystallography 1

1 Quantum Chemistry: Wavefunctions and Density Matrices 3
- 1.1 Calculation of wavefunctions using basis functions . . . 3
- 1.2 Reduced density matrices . . . 11
- 1.3 Density functional theory (DFT) . . . 12
- 1.4 Periodic ab-initio calculations . . . 14

2 Partitionings of Quantum Chemical Entities 17
- 2.1 Quantum Theory of Atoms in Molecules . . . 17
  - 2.1.1 Topology of the electron density . . . 18
  - 2.1.2 Atomic properties and bond properties . . . 19
  - 2.1.3 Software . . . 20
  - 2.1.4 Interaction analysis with other methods . . . 20
- 2.2 Interaction energies and their decomposition into classic and non-classic contributions . . . 21
  - 2.2.1 Overview . . . 21
  - 2.2.2 IQA - Energy terms from partitioned electron density . . . 22
  - 2.2.3 EDA - Energy terms from partitioned Hamiltonian . . . 27
  - 2.2.4 Exemplary calculation for water . . . 28
- 2.3 Atomic polarizabilities . . . 29

3 Crystallography and X-ray Diffraction 31
- 3.1 Crystallography . . . 31
- 3.2 The relationship between structure factors and electron density . . . 32
- 3.3 Modelling the structure and the electron density . . . 34
- 3.4 Structure factor simulation . . . 35

II Introduction: Theory of Quantum Crystallography **37**

4 Quantum Crystallography **39**
4.1 History of combining X-ray diffraction and quantum chemistry 39
4.1.1 From the first quantum mechanical scattering factors to more elaborated ones based on atom-centred aspherical functions 42
4.1.2 Density matrix and wavefunction refinement 44
4.2 Hansen-Coppens multipolar model 47
4.3 Databases of multipolar pseudoatoms 50
4.3.1 ELMAM 51
4.3.2 UBDB 52
4.3.3 Invariom 52
4.3.4 Comparison 52
4.4 X-ray constrained wavefunctions 53
4.4.1 Alternative minimization criteria 53
4.4.2 Advantages and disadvantages of XCW 55
4.4.3 Procedure for X-ray constrained wavefunction calculations 56
4.5 Hirshfeld atom refinement and X-ray wavefunction refinement 58
4.6 Other scattering techniques than X-ray diffraction 60
4.6.1 Compton scattering 60
4.6.2 Neutron diffraction 61
4.6.3 Polarized neutron diffraction 61
4.7 Quantum crystallography from other measurements than diffraction 61

III Results **63**

5 Outline and Goals **65**

6 Towards a Generalized Database of Atomic Polarizabilities **69**
6.1 Introduction 69
6.2 Theoretical Background 73
6.2.1 Earlier Atomic Polarizability Databases and the Need for a New One . . . 73
6.2.2 Distributed Atomic Polarizabilities 76
6.3 Constructing the Database 78
6.3.1 Computational Details 78
6.3.2 The Local Coordinate System 79
6.3.3 Multivariate Data Analysis and Clustering 83
6.3.4 Recognizing a functional group 85
6.4 Results 85
6.4.1 Clustering the CH_2 Polarizabilities 85
6.4.2 Clustering all Functional Groups 89
6.4.3 Using the Database to Compute Polarizabilities 89
6.5 Conclusions 96

6.6 Addendum . . . 98
6.6.1 Comments and additions to the published article . . . 98
6.6.2 Related work of other researchers . . . 100

7 Pressure-Induced Polymerization and Electrical Conductivity of a Polyiodide 103
7.1 Publication . . . 103
7.2 Addendum . . . 112

8 Different Aspects of X-ray Constrained Wavefunctions 117
8.1 X-ray constrained wavefunction fitting using the self-electrostatic energy . . . 117
8.1.1 Theoretical background . . . 117
8.1.2 Calculation details . . . 119
8.1.3 Results . . . 119
8.1.4 Discussion . . . 124
8.2 Electronic energy from structure factors and from wavefunctions . . . 128
8.3 How well can XCW calculations retrieve electric field effects? . . . 131
8.3.1 The effect of linear electric fields on the electron density from XCW calculations . . . 132
8.3.2 The effect of point charges around urea simulating neighbouring molecules 135
8.3.3 Comparison of electric field and correlation effects on the electron density from XCW calculations . . . 138
8.3.4 Effect of the crystal field on the electron density from XCW calculations 140
8.4 Frozen density embedding using X-ray constrained wavefunctions . . . 143

9 Analysis of Crystal Field Effects and Interactions with X-ray Restrained ELMOs 145
9.1 Introduction . . . 146
9.2 Theory and Methods . . . 148
9.2.1 The ELMO and XR-ELMO Techniques . . . 148
9.2.2 Computational protocol . . . 150
9.3 Results . . . 152
9.3.1 HCN . . . 152
9.3.2 Oxalic acid dihydrate . . . 159
9.4 Conclusions . . . 163
9.5 Addendum . . . 165
9.5.1 Boron-nitrilotriacetate high pressure results . . . 165
9.5.2 XRW and XR-ELMO calculations on simulated and measured structure factors of NTA-B . . . 166

10 Conclusions 171

List of Acronyms 173

A Appendix to Chapter 7 **175**
A.1 Powder diffraction data of TEAI collected in epoxy-nanoalumina 4:1 175
A.2 Determination of the correct space group of TEAI 175
A.3 Structural motifs of iodine chains in TEAI 176
A.4 Bending of the structural motif in TEAI under high-pressure 177
A.5 Bonding analysis of I_3^-, I_5^- and I_7^- optimized in vacuo 178
A.6 Further IQA interaction terms for I_5^- 179
A.7 IQA and EDA terms for tri-iodide I_3^- 180
A.8 Temperature effect on the lattice parameters and interatomic distances in TEAI under ambient pressure 180
A.9 Electrical resistivity measurements under high pressure and variable temperature 181

B Appendix to Chapter 8 **183**
B.1 Implementation of new formula in *Tonto* 183
B.2 Point charges from *Gaussian* and *Tonto* 183

C Appendix to Chapter 9 **185**
C.1 Dipole moment of HCN 185
C.2 Electric field of 0.1 a.u. along HCN 186
C.3 Electron density difference for oxalic acid with and without water 186
C.4 Periodic DFT calculations for NTA-B 186
C.5 Electron density differences of NTA-B for structure factors simulated at different pressures 187
C.6 Change of carbonyl ELMOs with point charge 188

Bibliography **189**

Declaration of consent **215**

Curriculum Vitae **216**

Part I

Introduction: Theory of Quantum Chemistry and Crystallography

1 Quantum Chemistry: Wavefunctions and Density Matrices

The many-body problem describes, among other things, the issue of exactly calculating quantum systems with more than two particles. It is, in general, not possible to analytically solve equations describing these systems, and since atoms and molecules (except for the hydrogen atom) consist of more than two particles, we can consequently not exactly calculate their behaviours.
However, various approaches have been developed to find approximate solutions to this problem, some of which are introduced in this chapter. They allow us to, at least heuristically, determine chemical quantities like the energy or other properties of atoms, molecules, and materials. This, in turn, is the basis to compute the behaviour of molecules in solutions to understand chemical reactions; the catalytic activity of surfaces important for efficiently and cheaply storing energy; the activity and toxicity of drugs; mechanical, optical, electrical, magnetic, and thermal properties of solids; and many more things, which are important for various fields of chemistry. This dissertation is exploiting and developing some of those approximate solutions to answer questions from crystallography and materials science.

1.1 Calculation of wavefunctions using basis functions

This chapter provides a brief introduction to the calculation of wavefunctions with the Hartree-Fock (HF) method. [1, 2]

Schrödinger equation

The Schrödinger equation describes the behaviour of particles in a quantum system. In general, it cannot be solved analytically (except for the hydrogen atom). Thus, the goal is to find numerical solutions, which are as close as possible to the exact solution. In order to do that, certain approximations can be made. The first one is that it is often sufficient to consider the time-independent Schrödinger equation:

$$\hat{H}_{tot}|\Psi\rangle = E_{tot}|\Psi\rangle \tag{1.1}$$

Where $\hat{H}_{tot}$ is the Hamilton operator, Φ the wavefunction and E_{tot} the total energy of the system. The Hamilton operator contains five parts: an operator for the kinetic energy of the nuclei $\hat{T}_{nucl}$, one for the potential energy of the nuclei $\hat{V}_{nucl}$, one for the nucleus-electron potential $\hat{V}_{nucl-el}$, one for the kinetic energy of the electrons $\hat{T}_{el}$, and one for the potential energy of the electrons $\hat{V}_{el-el}$.

$$\hat{H}_{tot} = \hat{T}_{nucl} + \hat{V}_{nucl} + \hat{V}_{nucl-el} + \hat{T}_{el} + \hat{V}_{el-el} \tag{1.2}$$

The second approximation is that one is typically dealing with light elements and therefore the systems can be treated non-relativistically. Thirdly, within the Born-Oppenheimer approximation nuclei and electrons can be treated separately thanks to their different masses owing to which the electrons move much faster than the nuclei. Thus, the electrons can be described in the static field of the nuclei. Therefore, the kinetic energy of the nuclei is neglected and one deals with a constant value for their potential energy. Obviously, a change of the nuclear coordinates goes along with a change of the nuclear potential energy term (as well as the electronic terms) leading to very different energies depending on the configuration, which can be depicted with a potential energy surface as shown in Figure 1.1. For a specific point on the energy surface, the operators acting on the electrons ($\hat{V}_{nucl-el}$, $\hat{T}_{el}$, $\hat{V}_{el-el}$) can be applied to describe the motion of the electrons in the field of the fixed nuclei.

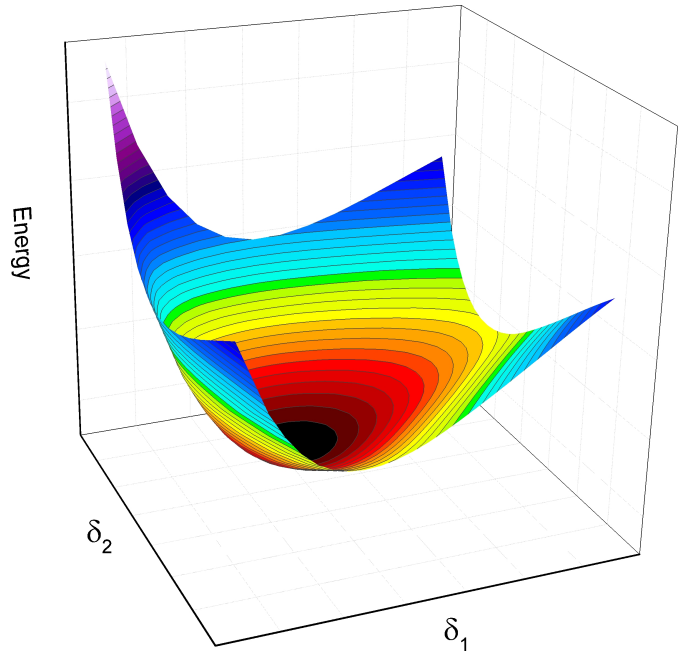


Figure 1.1 – Potential energy surface involving two distances. This particular example was done with I_5^- used in chapter 7, where δ_1 and δ_2 are the bond distances to the central iodine atom and the most stable point corresponds to I_5^- with C_{2v} symmetry.

Pauli exclusion principle

Besides fulfilling the Schrödinger equation, a wavefunction for fermions needs to be antisymmetric with respect to the exchange of two particles.

$$\Psi(x_1, x_2) = -\Psi(x_2, x_1) \tag{1.3}$$

If two particles have identical coordinates $x_1 = x_2$, zero is the only solution to equation 1.3:

$$\Psi(x_1, x_2) = -\Psi(x_2, x_1) = -\Psi(x_1, x_2) = 0 \tag{1.4}$$

From this follows the Pauli exclusion principle, which states that two fermions can never have the same quantum state. Besides the charge, electrons also have a spin, which can take two values, α and β. Since they are Fermions, they have to obey the Pauli exclusion principle and therefore either the spatial or the spin part of a wavefunction describing electrons needs to be antisymmetric, whereas the other part needs to be symmetric.

(Spin-)orbitals

In chemistry, the wavefunction is usually expressed in terms of orbitals, which are one-electron wavefunctions. Spatial orbitals describe the probability density of finding electrons in space. This means that the square of the spatial part ψ of the orbital integrated over a given volume $|\psi(\mathbf{r})|^2 d\mathbf{r}$ describes the probability of an electron to be in this volume. Spin orbitals $\chi(\mathbf{x})$ additionally inform about the spin of the electron (α or β).

Many-electron wavefunctions and Slater determinants

A multi-electron wavefunction, which is a solution of the electronic Hamiltonian could be simply constructed by multiplying the orbitals of all electrons:

$$\Psi^{(HP)} = \chi_1^{(HP)} * \chi_2^{(HP)} * \chi_3^{(HP)} * ... \chi_N^{(HP)} \tag{1.5}$$

This wavefunction, called Hartree-product, is a solution of the Schrödinger equation, in the sense that the orbitals are eigenvectors and the orbital energies eigenvalues of the Hamilton operator. However, such a wavefunction is not antisymmetrized. This can be achieved by taking the correct linear combinations of spin and spatial functions for the orbitals in such a way that the Pauli exclusion principle is fulfilled.
The two-electron function $\Psi_{12}^{HP} = \chi_i(x_1)\chi_j(x_2)$, for example, is an eigenvector of the Hamiltonian but not a correct wavefunction since it is not antisymmetric (and also not symmetric). A linear combination of this product and the same product with exchanged coordinates, however, is antisymmetric.

$$\Psi(x_1, x_2) = \frac{1}{\sqrt{2}}\{\chi_i(x_1)\chi_j(x_2) - \chi_j(x_1)\chi_i(x_2)\} \tag{1.6}$$

This can be generalized to N-electrons and in order to express this linear combination, a short-hand notation, the Slater determinant, was introduced:

$$\Psi(x_1, x_2, ..., x_N) = \frac{1}{\sqrt{N!}} = \begin{vmatrix} \chi_i(x_1) & \chi_j(x_1) & \dots & \chi_k(x_1) \\ \chi_i(x_2) & \chi_j(x_2) & \dots & \chi_k(x_2) \\ \dots & \dots & \dots & \dots \\ \chi_i(x_N) & \chi_j(x_N) & \dots & \chi_k(x_N) \end{vmatrix} \tag{1.7}$$

A wavefunction written as Slater determinant is automatically antisymmetrized and normalized.

Hartree-Fock method

The variational principle states that the best wavefunction is the one producing the lowest energy. Hartree was the first one to make use of this principle and he proposed a procedure for solving the Schrödinger equation with atomic orbitals (AOs). However, his method was based on a simple product of atomic orbitals and not on a correctly antisymmetrized one. Fock extended the approach to correctly antisymmetrized wavefunctions.

Left-multiplication of equation 1.1 with Ψ gives:

$$E = \langle \Psi | \hat{H}_{el} | \Psi \rangle \tag{1.8}$$

The electronic Hamiltonian $\hat{H}_{el}$ contains the kinetic energy operator acting on all electrons individually (one-electron part), the Coulomb attraction part between the electrons and the nuclei (also one-electron part) and the electron-electron repulsion (two-electron term).

$$\hat{H}_{el} = -\frac{1}{2}\sum_{i=1}^{N} \nabla_i^2 - \sum_{i=1}^{N} \frac{Z}{r_i} + \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{1}{r_{ij}} \tag{1.9}$$

Inserting orbitals χ_i into equation 1.8 and writing out the operators like in equation 1.9 leads to:

$$E = \sum_{i=1}^{n_{occ}} \langle \chi_i(1) | \hat{h}_i | \chi_i(1) \rangle + \frac{1}{2} \sum_{i=1}^{n_{occ}} \sum_{j=1}^{n_{occ}} \{ \langle \chi_i(1)\chi_i(1) \left| \frac{1}{r_{ij}} \right| \chi_j(2)\chi_j(2) \rangle - \langle \chi_i(1)\chi_j(1) \left| \frac{1}{r_{ij}} \right| \chi_j(2)\chi_i(2) \rangle \} \tag{1.10}$$

$$= \sum_{i=1}^{n_{occ}} \epsilon_i + \frac{1}{2} \sum_{i=1}^{n_{occ}} \sum_{j=1}^{n_{occ}} \{J_{ij} - K_{ij}\} \tag{1.11}$$

The first term is the one-electron term and gives rise to the orbital energies ϵ_i; the sum goes over all occupied spin orbitals. This term contains the kinetic energy of the electrons and the potential energy due to their attraction to the nuclei. The orbital energies are negative (stabilizing) for the occupied spin-orbitals. The second term is the Coulomb repulsion energy J_{ij}, a two-electron term which is always destabilizing. The third one is also a two-electron

term and gives rise to the exchange energy K_{ij}, which is always stabilizing. The exchange energy would not be present in a Hartree product, it arises due to the antisymmetrization requirement for the wavefunction. The sum goes over all pairs of occupied spin-orbitals and the factor 1/2 is needed to compensate for double-counting.

In chemistry, molecular orbitals (MOs) are usually described as linear combinations of atomic orbitals (LCAO method). Thus, the problem consists in finding the coefficients in the linear combination that produce the lowest total electronic energy.
The procedure starts with a first approximate wavefunction consisting of MOs χ_i. Only one electron is considered at once, which moves in the mean field produced by the nuclei and all other electrons. Thus, the effective potential can be obtained from the orbitals of all other electrons of the first approximation, which gives rise to the effective one-electron Fock operator $\hat{f}$ and the Fock equation for one electron can be written as:

$$\hat{f}(i)\chi_i(x_i) = \epsilon_i \chi_i(x_i) \tag{1.12}$$

where the operator is:

$$\hat{f}(i) = -\frac{1}{2}\nabla_i^2 + \sum_A \frac{Z_A}{r_{iA}} + \sum_B^{n_{occ}} \left\{\hat{J}_b(i) - \hat{K}_b(i)\right\} = \hat{h}(i) + v_i^{HF} \tag{1.13}$$

The one-electron part (kinetic energy of the electrons and nucleus-electron interaction) is summarized in $\hat{h}(i)$ while the two electron part consists of the Coulomb and exchange interaction and is $v_i^{HF} = \sum_b^{nocc}\{\hat{J}_b(i) - \hat{K}_b(i)\}$, where v_i^{HF} is the average electron-electron potential that electron 1 feels due to all other electrons. Therefore, the solution depends on all other spin-orbitals, which is why the Hartree-Fock equations can only be solved iteratively. By solving the one-electron Hartree-Fock equation, an improved orbital χ_1 and corresponding energy ϵ is obtained.
Afterwards, the same equation is solved for electron 2 and so on for all other electrons. After all electrons were calculated an improved set of spin orbitals is obtained. Now, one starts with a next cycle of one-electron Hartree-Fock equations and goes on until the spin-orbitals and the corresponding energies do not change anymore. Therefore, a given potential resulting from the effect of all other electrons does not change a particular spin-orbital anymore and vice versa. At this point the solution is self-consistent and this is why the procedure is called self-consistent field approach. In practice, the iterative cycles are repeated until the change is below a specified convergence criterion. The so-obtained set of spin-orbitals is the best solution for the Hartree-Fock case. The n orbitals with the lowest energy are the occupied ones and the higher ones are unoccupied (virtual orbitals).

Roothaan-Hall equations

Different functions are used to describe atomic orbitals, most common are Slater-type orbitals or linear combinations of Gaussian-type orbitals. These functions form the basis set, and

their linear combinations are not optimized in a Hartree-Fock calculation. What is optimized, instead, are the linear combinations of atomic orbitals forming the spatial molecular orbitals (MO-LCAO): $\psi_i = \sum_\mu C_{\mu i}\Phi_\mu$.
The Fock equation with a basis set becomes:

$$\hat{f}(i)\sum_{\nu=1}^{N} C_{\nu i}|\phi_\nu(i)\rangle = \epsilon_i \sum_{\nu=1}^{N} C_{\nu i}|\phi_\nu(i)\rangle \tag{1.14}$$

Left-multiplication with $\phi_\mu^*(i)$ and integration over all space leads to:

$$\sum_\nu C_{\nu i}\int_{-\infty}^{+\infty}\phi_\mu^*(i)\hat{f}(i)\phi_\nu(i)d\mathbf{r} = \epsilon_i\sum_\nu C_{\nu i}\int_{-\infty}^{+\infty}\phi_\mu^*(i)\phi_\nu(i)d\mathbf{r} \tag{1.15}$$

The overlap matrix **S** is defined as:

$$S_{\mu\nu} = \int_{-\infty}^{+\infty}\phi_\mu^*(i)\phi_\nu(i)d\mathbf{r} \tag{1.16}$$

Equation 1.15 can be expressed in matrix form:

$$\mathbf{FC} = \mathbf{SC}\epsilon \tag{1.17}$$

F is the Fock matrix, **C** is the coefficient matrix, and **S** is the overlap matrix arising due to the fact that the atomic orbitals are not generally orthogonal to each other. The first column in **C** specifies MO 1, the second MO 2, and so on. ϵ is a diagonal matrix, which specifies the orbital energies. Equation 1.17 is called the Roothaan-Hall equation.

The density matrix is defined as:

$$P_{\mu\nu} = 2\sum_a^{N/2} C_{\mu a}C_{\nu a}^* \tag{1.18}$$

Together with the basis functions, the density matrix defines the electron density:

$$\rho(\mathbf{r}) = \sum_{\mu\nu} P_{\mu\nu}\Phi_\mu(\mathbf{r})\Phi_\nu^*(\mathbf{r}) \tag{1.19}$$

As written in equation 1.13 the Fock operator is:

$$\hat{f}(i) = \hat{h}(i) + \sum_a^{N/2}\{2\hat{J}_a(i) - \hat{K}_a(i)\} \tag{1.20}$$

where the variable a designates all other molecular orbitals. The one-electron operator contains the kinetic and the nucleus-electron attraction parts:

$$\hat{h}(i) = -\frac{1}{2}\nabla_i^2 - \sum_A \frac{Z_A}{|r_i - R_A|} \tag{1.21}$$

The Fock operator in a basis $\{\Phi_\nu\}$ can be expressed in matrix representation:

$$\begin{aligned} F_{\mu\nu} &= \int d\mathbf{r}_i \Phi_\mu^*(i) \hat{f}(i) \Phi_\nu(i) \\ &= \int d\mathbf{r}_i \Phi_\mu^*(i) \hat{h}(i) \Phi_\nu(i) + \sum_a^{N/2} \int d\mathbf{r}_i \Phi_\mu^*(i) \{2\hat{J}_a(i) - \hat{K}_a(i)\} \Phi_\nu(i) \end{aligned} \tag{1.22}$$

For clarity it makes sense to define the core-Hamiltonian matrix, which is:

$$H_{\mu\nu}^{core} = \int d\mathbf{r}_i \Phi_\mu^*(i) \hat{h}(i) \Phi_\nu(i) \tag{1.23}$$

$H_{\mu\nu}^{core}$ does not depend on the expansion coefficients and therefore it is sufficient to calculate these integrals only once for a given basis set.

Equation 1.21 can be re-written in short-hand notation:

$$\begin{aligned} F_{\mu\nu} &= H_{\mu\nu}^{core} + \sum_a^{N/2} \left\{ 2 \left\langle \Phi_\mu \Phi_\nu \left| \frac{1}{r_{ij}} \right| \Psi_a \Psi_a \right\rangle - \left\langle \Phi_\mu \Psi_a \left| \frac{1}{r_{ij}} \right| \Psi_a \Phi_\nu \right\rangle \right\} \\ &= H_{\mu\nu}^{core} + \sum_a^{N/2} \left\{ 2\left(\mu\nu|aa\right) - \left(\mu a|a\nu\right) \right\} \end{aligned} \tag{1.24}$$

Inserting the linear combination of atomic orbitals $\Psi_a = \sum_\lambda C_{\lambda a} \Phi_\lambda$ results in:

$$\begin{aligned} F_{\mu\nu} &= H_{\mu\nu}^{core} + \sum_a^{N/2} \left\{ 2 \left\langle \Phi_\mu \Phi_\nu \left| \frac{1}{r_{ij}} \right| \sum_\sigma C_{\sigma a} \Phi_\sigma \sum_\lambda C_{\lambda a} \Phi_\lambda \right\rangle - \left\langle \Phi_\mu \sum_\lambda C_{\lambda a} \Phi_\lambda \left| \frac{1}{r_{ij}} \right| \sum_\sigma C_{\sigma a} \Phi_\sigma \Phi_\nu \right\rangle \right\} \\ &= H_{\mu\nu}^{core} + \sum_a^{N/2} \sum_\lambda \sum_\sigma C_{\lambda a} C_{\sigma a}^* \left[2\left(\mu\nu|\sigma\lambda\right) - \left(\mu\lambda|\sigma\nu\right) \right] \end{aligned} \tag{1.25}$$

Replacing the sum over the expansion coefficients with the density matrix **P** gives:

$$F_{\mu\nu} = H_{\mu\nu}^{core} + \sum_\lambda \sum_\sigma P_{\lambda\sigma} \left[\left(\mu\nu|\sigma\lambda\right) - \frac{1}{2} \left(\mu\lambda|\sigma\nu\right) \right] \tag{1.26}$$

The expressions in parenthesis are four-center two-electron integrals:

$$\left(\mu\nu|\sigma\lambda\right) = \int d\mathbf{r}_i d\mathbf{r}_j \; \Phi_\mu^*(i) \Phi_\nu(i) \left| \frac{1}{r_{ij}} \right| \Phi_\sigma^*(j) \Phi_\lambda(j) \tag{1.27}$$

Thus, the Fock matrix depends on the expansion coefficients **C** and so the Roothaan equations need to be solved in an iterative way until self-consistency is reached as previously described. An additional complication arises from the fact that the basis functions are not orthonormal. In order to solve the Roothaan equation, they need to be orthonormalized corresponding to a transformation of the overlap matrix **S** into a unit matrix. Then the Roothaan equation

becomes:

$$\mathbf{F'C'} = \mathbf{C'}\,\epsilon \tag{1.28}$$

This equation has only non-trivial solutions if the determinant is zero:

$$|\mathbf{F'} - \epsilon| = 0 \tag{1.29}$$

Solving this equation results in the eigenvectors **C**' corresponding to the molecular orbitals and the eigenvalues ϵ, which are the orbital energies. The coefficients **C**' can be back-transformed to **C** corresponding to the coefficients of the non-orthogonal basis set.

Post-Hartree-Fock methods

The more basis functions (AOs) are used, the better is the approximation of the calculated MOs. With an infinite basis set one reaches hypothetically the Hartree-Fock limit, which is the lowest possible energy one can obtain with one determinant. However, this is still not a very good solution because electron correlation is completely neglected. For this reason various post-HF techniques were developed.

It has been anticipated in the previous section, that in the HF ground state the lowest orbitals are filled with two electrons each. They are called the occupied orbitals, while all others are higher in energy and unoccupied (also called virtual). The variationally best single Slater determinant is the one constructed from these approximate orbitals. In excited states, the electrons also move to the virtual orbitals. When one electron is in an energetically higher orbital, it is a singly excited state, for two electrons it is doubly excited, and so on. In these cases, the Slater determinant contains other orbitals than the energetically lowest ones. The configuration interaction method makes use of these excited states by forming the best linear combination of them such that the total energy is lowered with respect to the HF energy. With this procedure electron correlation is introduced. Usually doubly or triple excited states are mixed into the HF ground state determinant because with more excited states the calculation becomes too complex. However, in principle one can make determinants with all possible excitations up to all N electrons. Such a calculation is called full configuration interaction (full CI). In principle the use of all possible determinants with an infinite basis set gives rise to an exact solution of the Schrödinger equation as one can see in Figure 1.2.

Another approach is the Møller-Plesset method where electron correlation is introduced by adding perturbing operators in addition to the standard Hamilton-operator. Generally post-HF methods such as CI and MP2 are time-consuming. A different approach where electron correlation is also not neglected is the Density Functional Theory (DFT) method, shortly explained in section 1.3.

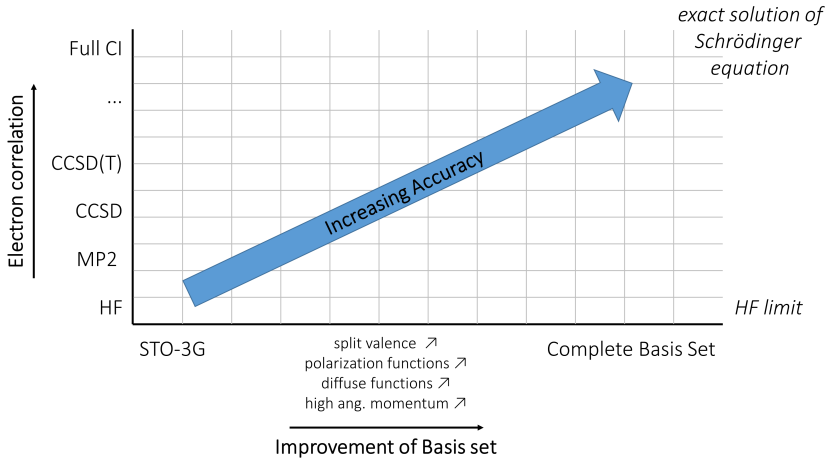


Figure 1.2 – Accuracy of quantum chemical calculations

1.2 Reduced density matrices

Wavefunctions depend on the electronic and spin coordinates of all electrons they describe and as a consequence, all properties of the system can be obtained by applying the corresponding operator. However, most of the properties in an N-electron system depend on only one electron or maximum two electrons at the same time. Therefore, very often more simplified expressions for a given system can be used, like for example the electron density or reduced density matrices (RDMs) [3, 4]. RDMs depend only on the coordinates of 1, 2, ... p electrons at the same time, where p is smaller than or equal to the total number of electrons N.

For a wavefunction with N electrons, the RDM of order p (p-electron-RDM, Γ^p) can be obtained by squaring the wavefunction (WFN) and integrating over all other coordinates but the ones until p. [5]

$$\Gamma^p(x_1, x_2, ..., x_p; x_1', x_2', ..., x_p') =$$
$$\binom{N}{p} \int \Psi(x_1, x_2, ...x_p, x_{p+1}, ..., x_N) \times \Psi^*(x_1', x_2', x_p', x_{p+1}, ..., x_N) dx_{p+1} ... dx_N$$

The 1-electron RDM Γ^1 depends on the coordinates of only one electron, the 2-electron RDM Γ^2 on the coordinates of two, and so on. Therefore, the 1-electron RDM can be written as:

$$\Gamma^1(x_1; x_1') = N \int \Psi(x_1, x_2, ..., x_N) \times \Psi^*(x_1', x_2, ..., x_N) dx_2 ... dx_N \tag{1.30}$$

When the spin is not considered, the spinless 1-electron RDM $\Gamma^1(r_1; r_1')$ depends on the three spatial coordinates of one electron only, however on the unprimed and the primed variables of this coordinate. The diagonal part ($r_1 = r_1'$) of the spinless 1e-RDM corresponds to the electron density and depends only on three spatial coordinates. The electron density is an

observable and can be obtained for example from X-ray diffraction as described in chapter 3. The off-diagonal part of the spinless 1e-RDM ($r_1 \neq r_1'$) can be obtained from Compton-scattering, as for example described by Gillet *et al.* [6]. Within the LCAO approximation the spinless 1e-RDM can be expressed with atomic orbitals ϕ_μ as: [7]

$$\Gamma^1(x_1; x_1') = \sum_{\mu\nu} P_{\mu\nu}\phi_\mu(r)\phi_\nu^*(r') \tag{1.31}$$

The 2e-RDM, given in equation 1.32, depends on the coordinates of two electrons at the same time like for example the Hamilton operator also does. Therefore, there were attempts to use simply this matrix instead of the full WFN for solving the Schrödinger equation. However, like also the 1e-RDM, the 2e-RDM needs to fulfill an additional condition called N-representability ensuring that the RDM is indeed representing a proper wavefunction [8, 9] and this is complicating the situation.

$$\Gamma^2(x_1, x_2; x_1', x_2') = \frac{N(N-1)}{2} \int \Psi(x_1, x_2, ..., x_N) \times \Psi^*(x_1', x_2', ..., x_N) dx_3 ... dx_N \tag{1.32}$$

Reduced density matrices can be used to obtain physical properties by applying the appropriate operators on them. One has to consider a given order when applying operators on RDMs. First, the operators act only on the unprimed coordinates. Then, the primed and unprimed coordinates are set equal $x_i' = x_i$. Lastly, the integration is carried out. Therefore, obtaining the expectation value of a one-electron operator requires only integration over spatial and spin coordinates of one electron:

$$\langle \hat{O}_1 \rangle = \int \left[\hat{O}_1 \Gamma^1(x_1; x_1') \right]_{x1=x'1} dr_1 \tag{1.33}$$

1.3 Density functional theory (DFT)

The essential concept of DFT is that the ground state energy of a system of electrons is determined by the electron density. This relationship was initially proposed by Thomas [10] and Fermi [11] already in the 1920ies. Therefore, one deals with only three coordinates in the DFT framework, in contrast to the Hartree-Fock and post-Hartree-Fock approaches where the N-electron wavefunctions with three spatial and one spin coordinate per electron is the basic tool. In 1964, Hohenberg and Kohn proved that the ground state energy E is indeed uniquely determined by the electron density $\rho(\mathbf{r})$ and thus is a functional of the density $E[\rho]$. [12] However, they only demonstrated that there is such a relationship but not how to find it. Their second theorem is the equivalent to the variational principle in HF theory. They showed that the energy of a trial density ρ_{trial} is always higher or equal to the energy of the exact ground state. In principle, DFT is based on the density and does not require the concept of orbitals, however, in 1965 Kohn and Sham [13] reintroduced orbitals for practical reasons. Their concept states that every system of interacting electrons moving in an external potential $v(\mathbf{r})$ can equivalently be described by a system of non-interacting electrons moving

in the potential $v_s(\mathbf{r})$ and these two systems have the same electron density. $v_s(\mathbf{r})$ is called the Kohn-Sham potential and it describes the potential in which the hypothetical non-interacting electrons are moving. The potential stems from the sum of the effect of the nuclei and all other electrons. [14]

Therefore, the "Kohn-Sham" operator applied on the Kohn-Sham orbitals consists of the kinetic operator and the potential operator

$$\left[-\frac{1}{2}\nabla^2 + v_s(\mathbf{r})\right]\phi_i = \epsilon_i\phi_i \tag{1.34}$$

The density of this non-interacting reference system is obtained in the usual way from the wavefunction $\rho(\mathbf{r}) = \sum \phi_i^* \phi_i$. What remains to be determined is the Kohn-Sham potential and if this was possible, equation 1.34 would result in the exact orbitals. The Kohn-Sham potential consists of three parts

$$v_s(\mathbf{r}) = v(\mathbf{r}) + V_{Coul}(\mathbf{r}) + v_{xc}(\mathbf{r}) \tag{1.35}$$

Where $v(\mathbf{r})$ is the attractive potential of the nuclei, $V_{Coul}(\mathbf{r})$ is the Coulomb repulsion between the electrons and $v_{xc}(\mathbf{r})$ the exchange correlation potential. The energy is therefore

$$E = T_S + V + V_{Coul} + E_{xc} \tag{1.36}$$

Where T_S is the kinetic energy, V the electron-nuclear attraction energy , V_{Coul} the Coulombic electron-electron interaction and E_{xc} the exchange-correlation energy. The potential terms V can be calculated from the Kohn-Sham density, which is identical to the density of the interacting system and thus these terms are exact. The kinetic energy is calculated based on the Kohn-Sham orbitals but the exact exchange-correlation potential is not known based on $\rho(\mathbf{r})$ (and of course also not based on the formulation of the density in terms of non-interacting orbitals). Contrariwise, if the energy was known exactly, the exchange-correlation potential could be obtained as the difference between this energy and the known parts (kinetic and one-electron potential terms).
The Kohn-Sham equations are solved in an SCF procedure like the HF equation. The resulting orbitals represent a system of non-interacting electrons and therefore do not have an obvious physical interpretation like the HF orbitals. However, the density derived from them corresponds to the density that interacting electrons of the system have. The advantage of this method is that a single Slater determinant is obtained, which keeps the computational costs low but nevertheless includes part of the electron correlation.

Different approximations for $v_{xc}(\mathbf{r})$ were made such as the local density approximation (LDA), a functional which depends only on the density itself and was already used by Kohn and Sham [13]. The generalized gradient approximation (GGA) depends in addition also on the gradient of the density [15]. The exchange-correlation can be split in an exchange and a correlation part $E_{xc} = E_x + E_c$. The exchange energy can also be calculated in the Hartree-Fock way but using the Kohn-Sham orbitals. Mixing this in leads to the so-called hybrid functionals.

Some functionals are obtained without any fitting to reference data like for example the widely used PBE [16] one and others use reference data like for example the most commonly used functional B3LYP consisting of Becke's three parameter exchange correlation functional and the LYP functional for correlation.

1.4 Periodic ab-initio calculations

Within the scope of this thesis, not only the calculation of molecules is important but also of crystals. A crystal is an extended system whose properties depend on the whole structure. It is obviously not possible to do calculations including all electrons of the crystal. The matrices in the Fock equation FC=SCE would have almost infinite dimensions. Luckily enough, crystals are, in fact, characterized by three-dimensional periodicity. By making use of this periodic behaviour one can circumvent the problem and calculate only one unit cell and then copy the solution for the other cells. The same assumptions as in molecular calculations can be made: only the time-independent, non-relativistic Schrödinger equation is considered and the Born-Oppenheimer approximation is used.

Subsequently, a simplified picture of periodic calculations is presented according to Dovesi *et al.* [17] where much more about this topic can be read.
The vector $\mathbf{g} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3$ is called lattice vector and connects two lattice points in the crystal. It is an integer multiple of the basis vectors of the primitive unit cell $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$.
The corresponding vector in the reciprocal lattice is $\mathbf{K} = K_1\mathbf{b}_1 + K_2\mathbf{b}_2 + K_3\mathbf{b}_3$ with K_j integers. $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ are the reciprocal lattice basis vectors. Vectors $\mathbf{a}$ and $\mathbf{b}$ are mutually orthogonal: $b_i \cdot a_j = 2\pi\delta_{ij}$

The Wigner-Seitz cell is defined as the portion of space being closer to one lattice point than to all others. The corresponding entity in reciprocal space is the first Brillouin zone: all points which are closer to one reciprocal lattice points than to all others belong to it.

The Hamilton operator can be written as $\hat{H} = \hat{K} + \hat{V}$ where $\hat{K}$ is the kinetic operator and $\hat{V}$ is a periodic potential due to the periodicity of the crystal. Therefore, the potential at a point $\mathbf{r}$ in the crystal has to be equal to the potential at $\mathbf{r} + \mathbf{g}$:

$$V(\mathbf{r} + \mathbf{g}) = V(\mathbf{r}) \tag{1.37}$$

The Schrödinger equation also needs to be invariant with respect to translation of $\mathbf{g}$.

$$\hat{H}(\mathbf{r} - \mathbf{g})\Psi(\mathbf{r} - \mathbf{g}) = E\Psi(\mathbf{r} - \mathbf{g}) \tag{1.38}$$

Solutions of this equation are called Bloch functions [18] and have the form

$$\Phi(\mathbf{r} + g; \mathbf{k}) = e^{i\mathbf{k}\cdot\mathbf{g}}\Phi(\mathbf{r}; \mathbf{k}) \tag{1.39}$$

where $\Phi(\mathbf{r}; \mathbf{k})$ is the Bloch function, $\mathbf{k}$ is the wavevector and $e^{i\mathbf{k}*g}$ is the Bloch phase factor.

The solutions can be limited to k in the first Brioullin zone because if $\mathbf{k}' = \mathbf{k} + \mathbf{K}$, then

$$e^{i\mathbf{k}'*\mathbf{g}} = e^{i\mathbf{k}*\mathbf{g}} + e^{i\mathbf{K}*\mathbf{g}} = e^{i\mathbf{k}*\mathbf{g}} \tag{1.40}$$

since $\mathbf{K} * g = 2\pi * "integer"$.

The Bloch function is itself a product of a plane wave and a periodic function $u(\mathbf{r};\mathbf{k})$:

$$\Phi(\mathbf{r};\mathbf{k}) = e^{i\mathbf{k}*\mathbf{r}} u(\mathbf{r};\mathbf{k}) \tag{1.41}$$

In principle the Bloch functions go over an infinite crystal. This is solved by introducing periodic boundary conditions. The easiest assumption is that the considered crystal is a finite part of an infinite crystal and therefore no surface effects are to be considered. The crystal is a parallelepiped containing $N = N_1 \cdot N_2 \cdot N_3$ unit cells with sides $N_j \cdot a_j$ (j=1,3) and N goes to infinity. Therefore:

$$\Phi(\mathbf{r} + mN_j a_j;\mathbf{k}) = \Phi(\mathbf{r};\mathbf{k}) \tag{1.42}$$

where m is an integer number.

With these conditions, the theorem of the Bloch function becomes:

$$\Phi(\mathbf{r} + mN_j a_j;\mathbf{k}) = e^{imN_j\mathbf{k}*a_j}\Phi(\mathbf{r};\mathbf{k}) \tag{1.43}$$

From equations 1.42 and 1.43 one sees that the phase factor must be one:

$$e^{imN_j\mathbf{k}*a_j} = 1 \tag{1.44}$$

The component $\mathbf{k}_j$ of the wave vector can be defined as:

$$\mathbf{k}_j = \frac{n_j}{N_j}\mathbf{b}_j \tag{1.45}$$

with n_j being integer. Therefore, **k** is any point in the reciprocal lattice. Again due to translational symmetry and as written in equation 1.40 the solutions can be limited to the first Brillouin zone.
There exist N of **k** points in every reciprocal cell:

$$\mathbf{k} = \frac{n_1}{N_1}\mathbf{b}_1 + \frac{n_2}{N_2}\mathbf{b}_2 + \frac{n_3}{N_3}\mathbf{b}_3 \tag{1.46}$$

When N_1, N_2, N_3 go to infinity, k becomes a continuous variable.

An important property of Bloch functions is that for two points k and k' in the first Brillouin zone all integrals that need to be calculated in the periodic Schrödinger equation are zero except k=k'. Bloch functions of different wave vectors are mutually orthogonal.
As a consequence, the Hamiltonian matrix represented with Bloch functions is block-diagonal where every block refers to a specific k and is independent of the others. Every block consists

of $n_f \times n_f$ Bloch functions.
An infinite number of blocks exists but since k is continuous it can be sampled at a finite amount of points.

In equivalence to molecular calculations, crystalline orbitals are expressed as linear combinations of Bloch functions

$$\Psi_n(\mathbf{r};\mathbf{k}) = \sum_j c_{jn}(\mathbf{k})\Phi_j(\mathbf{r};\mathbf{k}) \tag{1.47}$$

The solution lies in finding the optimum coefficients C(k) in the matrix equation $H(k)C(k) = S(k)C(k)E(k)$ in order to minimize the energy.
Therefore, the Schrödinger equation becomes:

$$h_{eff}\Psi^{\mathbf{k}}_{\mathbf{k}}(\mathbf{r}) = \epsilon^{\mathbf{k}}_i\Psi^{\mathbf{k}}_i(\mathbf{r}) \tag{1.48}$$

where h_{eff} contains the kinetic operator and a periodic potential operator, $\Psi^k_k(r)$ are crystalline orbitals, ϵ^k_i are the eigenvalues corresponding to the electronic bands, and i is the band index. From the solution of this equation follows the electronic band structure of the crystals. E_F is the Fermi level which corresponds to the highest energy of a populated state. The resulting crystalline orbitals are identically replicated over all primitive unit cells.

Bloch function can either be plane waves or localized functions. In most software packages like *Quantum Espresso* [19], *VASP* [20], *WIEN2k* [21], or *Turbomole* [22] plane waves are used because they are more favourable for the description of metals and delocalized electrons. However, plane waves have the disadvantage that they poorly describe the electron density cusps at nuclear positions. Therefore, pseudo-potentials are often required with this software. In *CRYSTAL17* [23] periodic atomic orbitals constructed from Gaussian type functions are taken, which is computationally more expensive than plane waves but has the advantage that standard basis sets like for molecular calculations can be processed.

Since periodic calculations are cost-intensive, very often DFT calculations are done but the approach is not limited to that. Also periodic HF, MP2, etc. calculations are possible.

In *CRYSTAL17* geometry optimizations as well as single point calculations can be performed. For geometry optimizations an external isotropic pressure can be defined and instead of minimizing just the electronic energy a term consisting of the electronic energy and the PV (pressure∗volume) term is minimized. On the one hand, this can be employed to simulate the measurement of crystals under pressure for comparison with experiment or to exclude experimental biases. On the other hand, it serves to extend the pressure range within which information is available beyond the experimentally reachable one in order to simulate properties of materials also under these conditions.

2 Partitionings of Quantum Chemical Entities

2.1 Quantum Theory of Atoms in Molecules

From a chemist's point of view the *microscopic* building blocks of matter are atoms, functional groups, and molecules. Contrariwise, the goal of a chemist is usually to obtain substances with certain *macroscopic* properties such as optical, electrical, magnetic or thermal properties of solids, given reactivity of surfaces or molecules in solutions, and much more. Therefore, it is crucial to correlate the macroscopic properties with the microscopic building blocks of a material. In fact, a lot of chemistry is explainable in terms of specific properties of building block like atoms and functional groups, because these entities keep similar characteristics in different chemical environments. An oxygen atom is usually nucleophilic and a carboxyl group acidic. Therefore, functional groups have prototypical characteristics, and it would be helpful to understand and predict them in order to rationally design materials with specific macroscopic properties.
Unfortunately, much of the quantum chemistry is based on canonical molecular orbital, which are spread over the entire molecule or system. This makes it difficult to extract local information specific to functional groups and atoms. One option is to make some *a-priori* constraints to the calculation, which allow the MOs to be only on given fragments (see chapter 9 about extremely localized molecular orbitals (ELMOs)) or to localize them *a-posteriori*. Another approach is based on the partitioning of the calculated (or experimental) electron density into atomic contributions. Different methods exist to do so, such as the Hirshfeld atom partitioning [24] or the Quantum Theory of Atoms in Molecules (QTAIM). The latter was used for various parts of this thesis and will thus shortly be introduced in this section.
The method was invented by Richard Bader and co-workers in the 1980ies [25] when they tried to find a quantum mechanical definition of atoms in molecules and therefore an approach that would make a link between experimental and theoretical chemistry.

QTAIM partitions the electron density based on its topology into so-called atomic basins (Ω). The entire space until a chosen limiting value of electrons per cubic Ångström is uniquely partitioned in a way that every point in space belongs to only one atom. Therefore, QTAIM

is independent of a reference model. Atomic properties are obtained by integrating over the atomic basins, which allows for studying the effect of different chemical environments on the atoms or larger building block, like in the polarizability database approach described in chapter 6. Another advantage of QTAIM is that properties are additive, signifying that properties of all individual atoms, like the charge or dipole, sum up to the value of the molecular property. Since the properties are prototypical for atoms and functional groups, they can also be transferred. Provided that a suitable local coordinate system is chosen, it is possible to calculate properties of a small and simple system and to transfer them to a larger system where explicit calculations are not possible or too time-consuming. This was initially used by Bader to justify his method with the example of estimating the heat of formation in a homologous series of hydrocarbons [26].

2.1.1 Topology of the electron density

QTAIM is based on a topological analysis of the electron density [27]. The main features are the critical points (CPs) which are points where the gradient of the electron density is zero. These can be maxima, minima or saddle points. Different CPs are distinguished by their curvature which is the sign of the second derivative of the electron density at this point. The number of non-zero curvatures is called the rank (ω) and is 3 at CPs. When the curvature is positive, it is assigned +1; when it is negative, it is assigned the value -1. The signature σ is the sum of the three curvatures. Therefore, CPs are defined by rank 3 and by their signature.

The electron density is dominated by the nuclear attractors which cause that the electron density has a maximum at these points called nuclear critical point (NCP). Every atomic basin and its content belongs to one NCP. For a NCP all curvatures are negative and therefore the signature is -3. It might happen that NCP are present at positions where there is no atomic nucleus. These are called non-nuclear maxima [28] and they are rare. A signature of +3 designates a cage critical point (CCP) and a signature of +1 a ring critical point (RCP) because it has a maximum in two directions and a minimum in one direction. These points are found inside rings of bonded atoms. The most important type of CPs for the analysis of bonding are the bond critical points (BCPs). These saddle points have two negative curvatures and one positive. The electron density has a minimum in the direction towards the neighbouring nuclear critical points which is defined as the bond path and it is maximal in all other directions. Therefore, the possible types of critical points are:
- (3,-3): NCP
- (3,-1): BCP
- (3,+1): RCP
- (3,+3): CCP

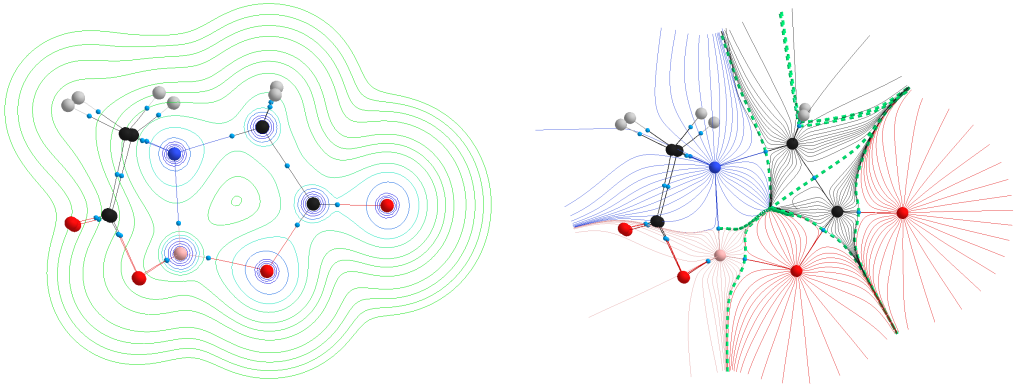
The total number and type of critical points follows the Poincaré-Hopf relationship, which states that for isolated molecules the number n of critical points are linked by

$$n_{NCP} - n_{BCP} + n_{RCP} - n_{CCP} = 1 \tag{2.1}$$

The separation into atomic basins is done with the zero-flux criterion. The borders of the basins are where the gradient vector field of the electron density has zero flux:

$$\nabla\rho * \mathbf{n} = 0 \tag{2.2}$$

Where **n** is a vector normal to the surface of the atomic basin. Basins are well defined and have one nucleus or non-nuclear maximum each. Figure 2.1 shows the electron density in the ring of a molecule used in chapter 9 of this thesis as well as its gradient vector field. The dashed green lines in Figure 2.1b are the zero-flux surfaces that separate the atomic basins in the plane of the ring.



(a) Electron density in the plane of a ring of the NTA-B molecule used in chapter 9

(b) Gradient vector field and some selected inter-atomic surfaces for the atoms of one ring of NTA-B.

Figure 2.1 – Electron density and QTAIM partitioning into atomic basins

2.1.2 Atomic properties and bond properties

Atomic properties as well as bond properties can be retrieved from QTAIM. Atomic properties such as the atomic charge, dipole moment, higher order moments, or energy are obtained by integration over the basin Ω. Due to the fact that QTAIM partitions the entire molecular space in a unique way, the atomic properties can be summed up to properties of functional groups (like for the polarizabilities described in section 2.3) and to properties of the entire molecule. These are exact within the precision of basin integration and of course, like any molecular property calculation, limited by the precision of the computational method itself.

In order to study interactions and bonds between two atoms, properties at the BCP between them can be investigated. First of all, this is the electron density $\rho(\mathbf{r} = BCP)$ itself as well as the Laplacian (second derivative of $\rho(\mathbf{r})$) $\nabla^2(\mathbf{r} = BCP)$. The latter is positive when the electron density is locally depleted and negative when it is locally concentrated. Thus, at generic positions it informs about the local Lewis acidity while at the BCP it can indicate a local charge concentration and therefore is diagnostic of a rather covalent bonding. When atoms with many electrons are examined (such as for the iodine-iodine bonds studied in

chapter 7), caution is required since the atomic shells also lead to local charge concentration and depletion and this behaviour might expand until the BCP.
The bond ellipticity ϵ describes the shape of the electron density in the plane containing the bond path. By definition its value is between 0 and 1. $\epsilon = 0$ designates a cylindrically symmetric bond and thus a single or triple bond while a higher ϵ indicates a double bond with π character.
The energy density at the BCP can as well give insight about the covalent versus ionic character of a bond and therefore was also employed in the chapter about bonding in polyiodides. Cremer and Kraka defined the electronic energy density $H(\mathbf{r})$ as the sum of the potential energy density $V(\mathbf{r})$ (always negative, stabilizing) and the gradient kinetic energy $G(\mathbf{r})$ (always positive, destabilizing). [29, 30] Thus, interactions where the energy density at the BCP $H_{r=BCP} = G_{r=BCP} + V_{r=BCP}$ is negative are dominated by the potential term designating a covalent interaction and vice versa.

Other properties that can be evaluated with QTAIM are localization and delocalization indices giving evidence about the sharing of electrons between two atomic basins. They provide information about whether an interaction is happening via a bonded pathway (high sharing) or rather through space (low sharing). The most used criterion is the delocalization index (DI) [31]. Such analysis, however, requires the full one-electron RDM as well as the second order density-matrix and thus cannot readily be done with DFT wavefunctions and experimental electron densities.

2.1.3 Software

QTAIM analysis can be done for both, molecular as well as periodic (crystal) wavefunctions and electron densities. Within the scope of this thesis, the software used for molecular ones is *AIMAll* (version 14.11.23 [32]) while the one for crystals is *TOPOND* as implemented in *CRYSTAL17* [23].

2.1.4 Interaction analysis with other methods

Besides the QTAIM method presented in the previous section, a plethora of techniques to study bonds and interactions exists. These are for example the non-covalent interaction (NCI) [33] method, electron localization function (ELF) [34], Roby-Gould index [35], orbital analysis with NBOs [36] or ELMOs [37] to mention only a few of them. Furthermore, bonding energies and their decomposition into different contributions can be analysed as described in the next section.

2.2 Interaction energies and their decomposition into classic and non-classic contributions

2.2.1 Overview

The bonding or interaction energy (subsequently, only the general term 'interaction energy' will be used) between two or more molecular fragments is defined as the difference between the sum of the fragment energies and the energy of the final composite structure. At least apart from an error introduced by the use of finite basis sets, which can be corrected with BSSE (basis set superposition error) counterpoise methods.
It is interesting to analyse the different contributions to the bonding energy. Two prominent decompositions methods are the interacting quantum atoms (IQA) [38] approach and the energy decomposition analysis (EDA) [39] used in chapter 7 for the analysis of iodine-iodine interactions in polyiodides, where it was probed how these interactions change when pressure is applied on the crystal. These two partitionings are further detailed here.

The philosophy behind the two partitionings is very different. The IQA partitioning is based on the electron density separated into atomic basins according to QTAIM and subsequent examination of atomic and interaction energies. The EDA is based on a partitioning of the Hamiltonian in the Schrödinger equations of the separated and interacting fragments.
IQA terms are always calculated for pairs of atoms. When the interaction energy of entire fragments is to be considered, the respective terms have to be summed. The method is implemented in *AIMAll* [32]. IQA analysis can be done for wavefunctions obtained from Hartree-Fock, B3LYP (using a specific correction in *AIMAll* for compensating the lack of a 2-electron density matrix), and complete active space calculations.
In the EDA approach, the interaction energies between fragments consisting of one or more atoms are evaluated. The method is for example implemented in *ADF* [40] but the terms can in principle be obtained from any quantum chemistry software. The partitioning based on the Hamilton operator works only for Hartree-Fock type calculations but an analogous partitioning for DFT functionals is implemented in *ADF*. In order to compare the two approaches, we will focus on Hartree-Fock calculations.

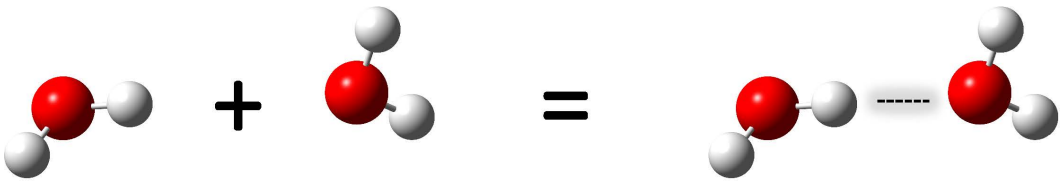


Figure 2.2 – Interaction between two water molecules. This example is used in section 2.2.4.

Both methods allow for the calculation of classic (electrostatic) and non-classic (covalent) terms. When looking at these terms and comparing their value in literature, they are very different. Indeed, the electrostatic term in EDA is always stabilizing whereas in IQA it is often destabilizing. Pendás [41] ascribes the large difference to the fact that with the IQA approach

all the electron density within a basin belongs to this atom. Electron density outside the interatomic surface belongs to a different atom. Therefore, it does not matter from which atom or fragment the electrons are initially coming. EDA instead assigns the electron density to the fragments by taking as reference the two fragments in vacuo, which is before the interaction. Thus, the electron densities are interpenetrating in EDA whereas in IQA they are concentrated around the nuclei. This leads to a lower electrostatic interaction of electrons to another nucleus in IQA. However, the electrostatic interaction with their own nucleus is instead stronger.
A more important point besides the difference in the assignment of the electron densities is that the IQA partitioning examines the interaction in the final molecule whereas EDA considers the interaction in steps. EDA starts from the two individual fragments which are the reference state and then lets them interact. It intrinsically takes into account the starting state of the fragments into the overall partitioning, which is not done in IQA where just the final interaction energy is considered. One can, however, also introduce an initial reference state in IQA and decompose in addition to the interaction energy also the deformation energy into classic and non-classic terms. Like that, the EDA and IQA results become directly comparable. This procedure was introduced by Racioppi *et al.* [42] and is detailed below.

2.2.2 IQA - Energy terms from partitioned electron density

The IQA terms can be calculated with the software *AIMAll*. The atomic energies are obtained by integrating over atomic basins. The atomic basins instead are obtained from a partitioning of the electron density according to QTAIM. In order to calculate the exchange-correlation terms, the two electron density matrix is needed. With Hartree-Fock calculations the exchange term is available (correlation of course intrinsically not) but DFT calculations are based on the electron density and do not provide a two-electron density matrix. Maxwell *et al.* [43] developed a method to circumvent this issue, which is described in section 2.2.2 'treatment of DFT wavefunctions'.

IQA theory

In the following paragraph the partitioning of the interaction energy between two fragments is presented using a consistent nomenclature; first in analogy to the original IQA derivation by Pendás *et al.* [38, 44], second in analogy to Cukrowski *et al.* [45, 46] and third according to Racioppi *et al.* [42] The numerous equations in this section are useful for somebody who wants to undertake the presented evaluation because there is no software automatically selecting or summing the correct terms.

Nomenclature
T: kinetic energy
V: potential energy
E: atomic/fragment/molecular energy

A: one-electron term
AA: interaction within the same atomic basin
AB: interaction between different basins
$\mathscr{G}$,$\mathscr{G}$: interaction of atoms in the same fragment
$\mathscr{G}$,$\mathscr{H}$: interaction of atoms on two different fragments

Partitioning of Pendás The total molecular energy of a molecule consists of one-body / monoatomic (Pendás) or self-energy (Cukrowski) terms and two-body or interaction terms.

$$E = \sum_A E_{self}^{AA} + \sum_{A>B} E_{int}^{AB} \tag{2.3}$$

The first sum goes over all atoms and the second sum goes over all pairs of atoms excluding double counting. The one-body or self-energy terms consist of the kinetic energy (always referring to the kinetic energy of the electrons, Born-Oppenheimer approximation), the electron-nucleus interaction within one atomic basin and the electron-electron interaction within one atomic basin:

$$E_{self}^{AA} = T^A + V_{en}^{AA} + V_{ee}^{AA} \tag{2.4}$$

The two-body or interaction contributions are the nucleus-nucleus interaction, the nucleus-electron interaction (between the nucleus of one basin and the electrons of another basin and vice versa) and the electron-electron interaction. These terms sum up to the interaction energy:

$$E_{int}^{AB} = V_{nn}^{AB} + V_{en}^{AB} + V_{ne}^{AB} + V_{ee}^{AB} \tag{2.5}$$

The sum over all above mentioned IQA atomic self and interaction energies in the molecule gives the total molecular energy where one needs to make sure that the interaction energies are not counted twice.

$$E = \sum_A E_{self}^{AA} + \sum_{A>B} E_{int}^{AB} = \sum_A \left(T^A + V_{en}^{AA} + V_{ee}^{AA}\right) + \sum_{A>B} \left(V_{nn}^{AB} + V_{en}^{AB} + V_{ne}^{AB} + V_{ee}^{AB}\right) \tag{2.6}$$

T^A is the atomic kinetic energy; the V_{en}, V_{ne}, V_{ee}, and V_{nn} terms describe the potential energies of the pair interactions between the electrons and nucleus of the same (AA) or different (AB) basins.

The electron-electron interaction term can be further split into the Coulomb interaction term (V_{Coul}^{AB}) and exchange-correlation term ($V_{xc}^{AB} = V_{ee}^{AB} - V_{Coul}^{AB}$) (blue) which is associated with a covalent interaction. Therefore, one can define the classic part (green) of the interaction energy as:

$$V_{cl}^{AB} = V_{nn}^{AB} + V_{en}^{AB} + V_{ne}^{AB} + V_{Coul}^{AB} \tag{2.7}$$

and the interaction energy itself as:

$$E_{int}^{AB} = V_{cl}^{AB} + V_{xc}^{AB} \tag{2.8}$$

The partitioning of the interaction energy E_{int}^{AB} in equation 2.8 gives rise to Pendás' partitioning

into classic and exchange-correlation contributions to the bonding energy.

By inserting these terms into equation 2.6 the total molecular energy can be expressed in terms of kinetic, classic and non-classic parts.

$$\begin{aligned} E &= \sum_A \left(T^A + V_{en}^{AA} + V_{ee,cl}^{AA} + V_{ee,xc}^{AA} \right) + \sum_{A>B} \left(V_{nn}^{AB} + V_{en}^{AB} + V_{ne}^{AB} + V_{ee,cl}^{AB} + V_{ee,xc}^{AB} \right) \\ &= \sum_A T^A + \sum_A E_{self,cl}^{AA} + \sum_A E_{self,xc}^{AA} + \sum_{A>B} E_{int,cl}^{AB} + \sum_{A>B} E_{int,xc}^{AB} \end{aligned} \tag{2.9}$$

The classic and non-classic terms obtained with this procedure are difficult to compare to terms obtained from an EDA partitioning where the total binding energy is partitioned and not only the interaction energy.

Pendás has indeed defined the binding energy as:

$$E_{bind} = E - \sum_A E_{vac}^A = \sum_A E_{def}^A + \sum_{A>B} E_{int}^{AB} \tag{2.10}$$

Where E_{vac}^A is the energy of the fragments in vacuo.
But he did not go on to partition the complete binding energy into classic and non-classic terms. In order to obtain the classic and non-classic term of the complete binding energy, one has to partition also the deformation energy (difference between final and initial/reference state).

Partitioning of Cukrowski Cukrowski [46] proposed the *fragment attributed molecular system energy change* (FAMSEC) scheme based on IQA going in the direction of partitioning also the deformation energy. He studied intramolecular interactions between two atoms A and B in the final state (*fin*). In order to quantify the contribution of the interaction between these two atoms to the binding energy he defined a reference state (*ref*) taking a different conformer where the two atoms are not interacting.

The fragment $\mathscr{G}$ consists of the interacting atoms A and B. The change in their self-energy is:

$$\Delta E_{self}^{\mathscr{G}} = \Delta E_{self}^A + \Delta E_{self}^B \tag{2.11}$$

The change in the interaction energy of these two atoms with all other atoms in fragment $\mathscr{H}$ is:

$$\Delta \sum_{X \in \mathscr{H}} E_{int}^{\mathscr{G}X} = {}^{fin}\sum_{X \in \mathscr{H}} E_{int}^{\mathscr{G}X} - {}^{ref}\sum_{X \in \mathscr{H}} E_{int}^{\mathscr{G}X} \tag{2.12}$$

The change of the interaction energy between atoms A and B is

$$\Delta E_{int}^{AB} = {}^{fin}E_{int}^{AB} - {}^{ref}E_{int}^{AB} \tag{2.13}$$

These terms sum up to:

$$E^{\mathscr{G}}_{attrb-mol} = \Delta E^{\mathscr{G}}_{self} + \Delta E^{AB}_{int} + \Delta \sum_{X \in \mathscr{H}} E^{\mathscr{G}X}_{int} \tag{2.14}$$

This is the energy contribution of the fragment $\mathscr{G}$ (atoms A and B or also more) to the molecular energy change when it goes from *ref* to *fin*. Since these terms do not include the change of the self-energy of all other atoms as well as the change of the interaction among all others, it is not equivalent to the binding energy $E^{fin} - E^{ref}$.

Partitioning of Racioppi Racioppi et al. [42] introduced a procedure which includes all terms and therefore partitions the entire binding energy into kinetic, classic (electrostatic) and non-classic (exchange-correlation) parts. Due to the many interaction energy terms, this calculation is extremely time consuming both in computation and human evaluation. Therefore, it is only adequate for small systems.
In order to calculate the binding energy between several fragments ($\mathscr{G}$, $\mathscr{H}$, etc.) one needs to calculate three terms:

1) The change of the self-energy. This is the difference between the self-energies E^{AA}_{self} of all atoms A in the final state and the self-energies of all atoms in the initial state $E^{AA}_{self,vac}$. The sum goes over all atoms in all fragments $\mathscr{H}$. The fragments $\mathscr{H}$ have the very same geometry in the the initial and final state.

$$E_{def,self} = \sum_{\mathscr{H}} \left(\sum_{A \in \mathscr{H}} E^{AA}_{self} - E^{AA}_{self,vac} \right) = \sum_{\mathscr{H}} \left(\sum_{A \in \mathscr{H}} \Delta E^{AA}_{self} \right) \tag{2.15}$$

2) The change of the interaction energy within the same fragment from before to after the interaction. The interaction energies between atoms of any fragment $\mathscr{H}$ change when going from the isolated fragment to the final compound. This gives rise to the deformation energy term $\Delta E^{\mathscr{H}}_{def,int}$ which is summed over all interaction within the same fragment for all the fragments.

$$E_{def,int} = \sum_{\mathscr{H}} \left(\sum_{\substack{A>B \\ A,B \in \mathscr{H}}} E^{AB}_{int} - E^{AB}_{int,vac} \right) = \sum_{\mathscr{H}} \left(\sum_{\substack{A>B \\ A,B \in \mathscr{H}}} \Delta E^{AB}_{fragment,int} \right) \tag{2.16}$$

3) The total interaction energies of all atoms from fragment $\mathscr{H}$ with all atoms of fragment $\mathscr{G}$. There is no Δ here because this term designates interaction energies between atoms of different fragments and they are not present in the initial state.

$$E_{fragment,int} = \sum_{\substack{A>B \\ A \in \mathscr{G}, B \in \mathscr{H}}} \left(\sum_{A \in \mathscr{H}} \sum_{B \in \mathscr{G}} E^{AB}_{int} \right) = \sum_{\substack{A>B \\ A \in \mathscr{G}, B \in \mathscr{H}}} E^{\mathscr{G},\mathscr{H}}_{int} \tag{2.17}$$

The sum of the deformation energies is positive (destabilizing) while the interaction energy is usually negative (stabilizing).

The total binding energy between all fragments is the sum of these three components.

$$E_{fragment,bind} = E_{def,self} + E_{def,int} + E_{fragment,int} \tag{2.18}$$

Only the last term is present in Pendás decomposition scheme.

Example for two interacting fragments:
Equation 2.18 written out for two fragments $\mathscr{G}$ and $\mathscr{H}$ becomes:

$$E^{\mathscr{G},\mathscr{H}}_{fragment,bind} = \sum_{A\in\mathscr{G}} \Delta E^{AA}_{def,self} + \sum_{A\in\mathscr{H}} \Delta E^{AA}_{def,self} + \sum_{\substack{A,B\in\mathscr{G}\\A>B}} \Delta E^{AB}_{def,int} + \sum_{\substack{A,B\in\mathscr{H}\\A>B}} \Delta E^{AB}_{def,int} + \sum_{\substack{A\in\mathscr{G}\\B\in\mathscr{H}}} E^{AB}_{int} \tag{2.19}$$

The self energy consists of:

$$E^{AA}_{self} = T^A + V^{AA}_{en} + V^{AA}_{ee} = T^A + V^{AA}_{en} + V^{AA}_{Coul} + V^{AA}_{xc} = T^A + V^{AA}_{cl} + V^{AA}_{xc} \tag{2.20}$$

and the interaction energies of

$$E^{AB}_{int} = V^{AB}_{nn} + V^{AB}_{en} + V^{AB}_{ne} + V^{AB}_{ee} = V^{AB}_{nn} + V^{AB}_{en} + V^{AB}_{ne} + V^{AB}_{Coul} + V^{AB}_{xc} = V^{AB}_{cl} + V^{AB}_{xc} \tag{2.21}$$

Inserting equations 2.20 and 2.21 into equation 2.19 leads to:

$$\begin{aligned}
E^{\mathscr{G},\mathscr{H}}_{fragment,bind} = &\sum_{A\in\mathscr{G}} \Delta T^{A,\mathscr{G}}_{def-self} + \Delta V^{AA,\mathscr{G}}_{def-self,en} + \Delta V^{AA,\mathscr{G}}_{def-self,Coul} + \Delta V^{AA,\mathscr{G}}_{def-self,xc} \\
&+ \sum_{A\in\mathscr{H}} \Delta T^{A,\mathscr{H}}_{def-self} + \Delta V^{AA,\mathscr{H}}_{def-self,en} + \Delta V^{AA,\mathscr{H}}_{def-self,Coul} + \Delta V^{AA,\mathscr{H}}_{def-self,xc} \\
&+ \sum_{\substack{A,B\in\mathscr{G}\\A>B}} \Delta V^{AB,\mathscr{G}}_{def-int,nn} + \Delta V^{AB,\mathscr{G}}_{def-int,en} + \Delta V^{AB,\mathscr{G}}_{def-int,ne} + \Delta V^{AB,\mathscr{G}}_{def-int,Coul} + \Delta V^{AB,\mathscr{G}}_{def-int,xc} \\
&+ \sum_{\substack{A,B\in\mathscr{H}\\A>B}} \Delta V^{AB,\mathscr{H}}_{def-int,nn} + \Delta V^{AB,\mathscr{H}}_{def-int,en} + \Delta V^{AB,\mathscr{H}}_{def-int,ne} + \Delta V^{AB,\mathscr{H}}_{def-int,Coul} + \Delta V^{AB,\mathscr{H}}_{def-int,xc} \\
&+ \sum_{\substack{A\in\mathscr{G}\\B\in\mathscr{H}}} V^{AB,\mathscr{G}\mathscr{H}}_{frag-int,nn} + V^{AB,\mathscr{G}\mathscr{H}}_{frag-int,en} + V^{AB,\mathscr{G}\mathscr{H}}_{frag-int,ne} + V^{AB,\mathscr{G}\mathscr{H}}_{frag-int,Coul} + V^{AB,\mathscr{G}\mathscr{H}}_{frag-int,xc}
\end{aligned} \tag{2.22}$$

The green terms are classic energies and the blue terms are exchange-correlation energies.

At this point, one can summarize all terms into kinetic, classic and exchange-correlation deformation parts as well as classic and exchange-correlation interaction parts and generalize to more than two fragments:

$$E^{\mathscr{G},\mathscr{H}}_{fragment,bind} = \Delta T^{\mathscr{G}}_{def} + \Delta V^{\mathscr{G}}_{def,classic} + \Delta V^{\mathscr{G}}_{def,xc} + V^{\mathscr{G},\mathscr{H}}_{classic} + V^{\mathscr{G},\mathscr{H}}_{xc} \tag{2.23}$$

Where $\mathscr{G}$ designates terms on the same fragment and $\mathscr{G},\mathscr{H}$ terms on different fragments.

Thus, a partitioning of the entire binding energy into kinetic, classic and exchange-correlation terms is obtained, which can be used for comparison to the EDA terms.

Treatment of DFT wavefunctions

In *AIMAll* version 14.04.17 a treatment for the popular functional, B3LYP, was implemented [43] allowing to calculate all the above listed terms. The total atomic exchange-correlation energy is available from the functional. The separation into inter- and intra-atomic contributions is done as follows: The Kohn-Sham orbitals are handled as if they were Hartree-Fock orbitals and the interatomic exchange is calculated with HF treatment. The intra-atomic term is simply the total atomic exchange-correlation minus this interatomic pseudo HF exchange term. This treatment allows to recover the total exchange-correction energy exactly (withing the limitation of the method) and it also allows a somewhat ambiguous separation into inter and intra-atomic terms. In order to activate this option, the keyword 'B3LYP' has to be inserted in the extended wavefunction (.wfx) file.

2.2.3 EDA - Energy terms from partitioned Hamiltonian

The binding energy of two or more fragments can be defined as

$$E_{bind} = E^{tot} - \sum_{fragments} E^{frag} \tag{2.24}$$

The Hamilton operator is composed of a one-electron part and a two electron part. The one electron part contains the kinetic energy of the electrons and the nucleus-electron interaction. The two electron part consists of the electron-electron interactions. With this decomposition in the Schrödinger equation it is easy to distinguish between the Coulomb part and the exchange part of the electron-electron interactions. (There is no correlation in the Hartree-Fock case.)

$$\begin{aligned} E_0 &= \langle\Psi_0|\hat{H}_{el}|\Psi_0\rangle \\ &= \sum_{i=1}^{n_{occ}} \langle\chi_i(1)|\hat{h}_i|\chi_i(1)\rangle + \frac{1}{2}\sum_{i=1}^{n_{occ}}\sum_{j=1}^{n_{occ}} \{\langle\chi_i(1)\chi_i(1)\left|\frac{1}{r_{ij}}\right|\chi_j(2)\chi_j(2)\rangle - \langle\chi_i(1)\chi_j(1)\left|\frac{1}{r_{ij}}\right|\chi_j(2)\chi_i(2)\rangle\} \\ &= \sum_{i=1}^{n_{occ}} \epsilon_i + \frac{1}{2}\sum_{i=1}^{n_{occ}}\sum_{j=1}^{n_{occ}} \{J_{ij} - K_{ij}\} \\ &= \sum_{i=1}^{n_{occ}} (T+V)_i + \frac{1}{2}\sum_{i=1}^{n_{occ}}\sum_{j=1}^{n_{occ}} \{J_{ij} - K_{ij}\} \end{aligned} \tag{2.25}$$

Applying this on the wavefunctions of the fragments and of the total assembly (final state) leads to their respective kinetic energy, electron-nucleus interaction energy, Coulombic electron-

electron interaction energy and exchange interaction energy.

$$\left[\sum_{i=1}^{n_{occ}}(T+V)_i+\frac{1}{2}\sum_{i=1}^{n_{occ}}\sum_{j=1}^{n_{occ}}\{J_{ij}-K_{ij}\}\right]^{frag}\Psi^{frag}=\left[T^{frag}+V_{ne}^{frag}+V_{coul}^{frag}+V_{xc}^{frag}\right]\Psi^{frag} \tag{2.26}$$

Subsequently, components belonging together can be grouped into classic and exchange terms and the corresponding terms are subtracted:

$$\Delta T_{=}T^{final}-\sum_{fragments}T^{frag} \tag{2.27}$$

$$\Delta V_{classic}=V_{classic}^{final}-\sum_{fragments}V_{classic}^{frag} \tag{2.28}$$

$$\Delta V_{xc}=V_{xc}^{final}-\sum_{fragments}V_{xc}^{frag} \tag{2.29}$$

Like this, the total binding energy can be partitioned. However, papers about EDA (for example by Morokuma [47], Bickelhaupt [14], or Franking [39]), usually focus on the partitioning of the binding energy into electrostatic interaction, Pauli repulsion, and orbital interaction term.

$$E^{bind}=E_{elstat}+E_{Pauli}+E_{orb} \tag{2.30}$$

Nevertheless, it is also possible to retrieve the classic and non-classic terms as a by-product of an EDA calculation, for example with the software *ADF* [40]. In that case a partitioning of the Pauli repulsion and orbital terms into kinetic, coulombic and exchange-correlation part is obtained such that one can extract the kinetic, classic and non-classic parts.

2.2.4 Exemplary calculation for water

To highlight the differences and similarities between IQA and EDA, an exemplary calculation of the interaction energies between two water molecules arranged as shown in Figure 2.2 was done. To allow for a direct comparison, the EDA terms were derived from integrals obtained with *Gamess* [48] because *ADF* uses Slater-type functions, whereas IQA analysis with *AIMAll* requires a wavefunction with Gaussian-type functions.

As previously written, the generally used terms in literature resulting from an EDA calculation are the electrostatic, Pauli, and orbitalic terms. Whereas with the IQA method, the classic V_{cl} and exchange-correlation terms V_{xc} of the final state are typically analysed.[41, 49] While for the H_2O dimer, the agreement between IQA classic and EDA electrostatic terms is not bad, this is definitely not generally the case. When these literature partitionings are for example applied to the polyiodides in chapter 7, the classic term in I_5^- (I_3^- and I_2 interacting) is positive (destabilizing) whereas the electrostatic term in EDA is negative (stabilizing).

What results from applying the IQA method as done in literature, are energy terms of only the final compound. The choice of considering just these values stands in contrast to how bonding is normally interpreted with regard to energetic considerations. In fact, bonding is usually explained by the stabilization (lowering of total energy) going along with the interaction of two units. Furthermore, this IQA approach considering only energies in the final compound leads to different numbers with respect to EDA. When subtracting, in contrast, the corresponding terms of the reference state (I_3^- and I_2) from the final state (I_5^-), the classic term becomes negative as it is the case in EDA.

Thus, it makes sense to introduce reference molecules also in an IQA analysis and to consider all the terms of equation 2.22. Upon additionally rearranging the EDA terms into classic and exchange as done in the bottom part of Table 2.1, the respective IQA and EDA values become very similar. Therefore, the two conceptually very different partitionings can be compared, and it is found that the agreement is excellent.

Table 2.1 – Enumeration and comparison of IQA and EDA terms for the water dimer. The values are given in kJ/mol. The first two IQA terms refer to numbers obtained from analysis of the dimer only. All other numbers are differences between a given term in the dimer minus the sum of the respective terms in the monomers.

	IQA	EDA
classic (V_{cl})	-19.09	
exchange-correlation (V_{xc})	-32.27	
electrostatic (E_{elstat})		-17.84
Pauli repulsion (E_{Pauli})		13.98
orbitalic (E_{orb})		-16.41
1-electron part (V_{pot})		-46548.22
1-electron part (V_{kin})	83.34	77.08
nucleus-nucleus (V_{NN})	46343.34	46346.36
Coulomb (V_{eeCoul})	23139.19	23296.51
exchange (V_{eexc})	-19.32	-15.22
kinetic	83.34	77.08
exchange	-19.32	-15.22
classic	-83.30	-82.99
Total (E_{tot})	-19.28	-21.14

2.3 Atomic polarizabilities

An electric field, which can be externally applied or provoked by surrounding molecules, can displace the charge distribution of an atom or molecule with respect to its unperturbed state. This invokes or changes the dipole moment of the molecule and is quantified by the polarizability stating how easily the electron density gets distorted. [50]

For small electric fields the induced dipole moment **p** is proportional to the electric field **E** and the proportionality constant is the polarizability α.

$$\mathbf{p} = \boldsymbol{\alpha}\mathbf{E} \tag{2.31}$$

The dipole moment and the electric field are vectors with three components and therefore $\boldsymbol{\alpha}$ is a 3x3 tensor.

$$\begin{pmatrix} p_x \\ p_y \\ p_z \end{pmatrix} = \begin{pmatrix} \alpha_{xx} & \alpha_{xy} & \alpha_{xz} \\ \alpha_{yx} & \alpha_{yy} & \alpha_{yz} \\ \alpha_{zx} & \alpha_{zy} & \alpha_{zz} \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix} \tag{2.32}$$

The isotropic polarizability is one third of the trace of matrix $\boldsymbol{\alpha}$.

In analogy to atoms or molecules, the polarization **P** informs about how much an applied electric field polarizes a dielectric material. For small electric fields and homogeneous dielectrics the proportionality constant is the electric susceptibility $\boldsymbol{\chi}$.

$$\mathbf{P} = \epsilon_0 \boldsymbol{\chi}\mathbf{E} \tag{2.33}$$

From the susceptibility $\boldsymbol{\chi}$ the dielectric constant ϵ_r and the refractive index **n** are easily obtained with equation 2.34 assuming that the relative permeability is 1.

$$n = \sqrt{\epsilon_r} = \sqrt{\boldsymbol{\chi} + 1} \tag{2.34}$$

The dielectric constant as well as the refractive index are important properties of a material. There is a high demand for low dielectric constant materials in the semiconductor industry [51] and high refractive index materials for optoelectronic devices [52].

The macroscopic polarization can be expressed as sum over all dipole moments of N molecules per volume:

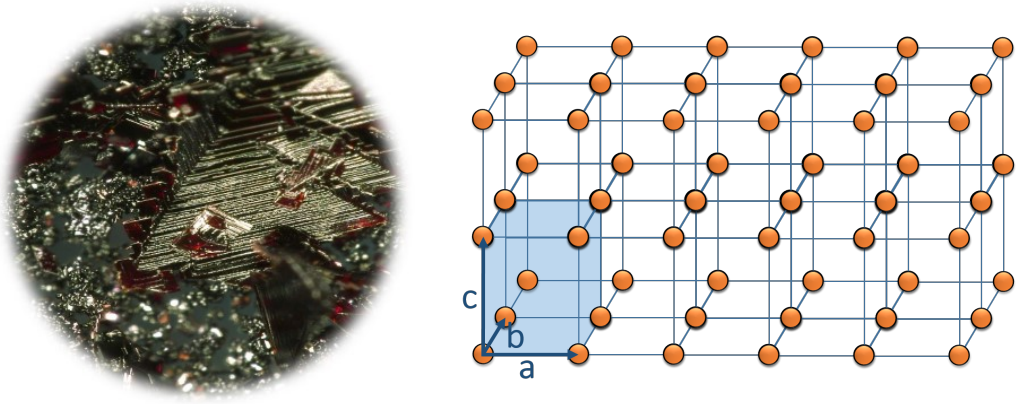
$$\mathbf{P} = \frac{N\mathbf{p}}{V} = \frac{N\,\boldsymbol{\alpha}\mathbf{E}_{local}}{V} \tag{2.35}$$

This equation relates macroscopic with microscopic quantities. However, the external electric field does not correspond to the one a molecule is exposed to. The molecule feels a local electric field, which depends on the external one and on on its chemical surrounding.
Different approaches to relate the molecular polarizability to the refractive index of metal organic frameworks (MOFs) are described in a paper resulting from my master thesis. [53]

3 Crystallography and X-ray Diffraction

3.1 Crystallography

Crystals are three-dimensional objects with periodic translational symmetry, where the repeating unit is called unit cell. In addition to the translational symmetry, other symmetry elements (mirrors, rotational and screw axes) are often present giving rise to 230 different space groups.



(a) Polyiodide crystals used in chapter 7. (b) Tetragonal crystal lattice and unit cell dimensions

Figure 3.1 – Macroscopic and microscopic features of crystals.

The periodicity of the crystal lattice enables both experimental structure determination and calculation of structure and properties. In section 1.4 it was described how one makes use of the periodicity in order to do *ab-initio* calculations of crystals. On the experimental side, scattering experiments make use of periodic lattices. Most commonly X-rays are used because they are easily produced and have wavelengths in the order of atomic bonds, but other particles such as electrons and neutrons can be used as well to study crystals. The irradiation of a crystalline sample leads to a diffraction pattern, which can be used to determine its structure. Both, single crystals and polycrystalline powders can be used for diffraction experiments. In contrast to other analytical methods like nuclear magnetic resonance (NMR), ultra-violet (UV), or mass spectroscopy (MS), X-ray diffraction experiments provide the structure of the

sample at atomic or even subatomic resolution making this method a very important tool in science. Besides the structural determination, which is the most wide-spread use of X-ray diffraction, one can also aim at estimating the electron charge density distribution and thus obtain information about the underlying bonding and interactions in the crystal.

3.2 The relationship between structure factors and electron density

Not so long after the discovery of X-rays by Röntgen, the laws of X-ray diffraction were established by van Laue [54], the Braggs [55] and others. X-rays are scattered by the electrons around atoms and if the atoms are periodically ordered like in crystals, the scattered radiation is subject to constructive and destructive interference leading to a diffraction pattern.
The positions of the diffraction spots are determined by the periodicity and symmetry of the crystal while their intensities depend on the scattering power of the present atoms and their positions relative to each other. Therefore, X-ray diffraction patterns were used to determine the atomic structures of crystals since William Lawrence Bragg proposed his equations [56] and nowadays the main application remains the structure determination of crystals consisting of atoms, molecules, polymers, or other substances.

The measured diffraction intensities I_{hkl} are proportional to the squares of the structure factors F_{hkl}, where h,k,l are the Miller indices and designate the reflection positions.

$$I_{hkl} \propto |F_{hkl}|^2 \tag{3.1}$$

The structure factors in turn are the sum of the atomic scattering factors [57].

$$F_{hkl} = \sum_{j}^{N\ atoms} f_j(\mathbf{k})\ e^{2\pi i(hx_j+ky_j+lz_j)} = \sum_{j}^{N\ atoms} f_j(\mathbf{k})\ e^{2\pi i(\mathbf{k}\cdot\mathbf{r})} \tag{3.2}$$

The sum goes over all atoms and $\mathbf{k}$ is the scattering vector in reciprocal space with $|k| = \frac{2\sin\theta}{\lambda}$, 2θ being the scattering angle, and λ the X-ray wavelength. Furthermore, it is $\mathbf{k}\cdot\mathbf{r} = (hx_j + ky_j + lz_j)$ with x_j, y_j, z_j being the fractional coordinates of atom j. Atoms in a crystal are always in movement and therefore the structure factor $f^{(0)}$, which an atom would have at rest, has to be convoluted with a temperature-dependent factor $T(\mathbf{k})$ describing the movement of the nuclei. [58] It is assumed that the electrons follow the movement of the nuclei instantaneously and unaltered.

$$f^T = f^{(0)} * T(\mathbf{k}) \tag{3.3}$$

By inserting equation 3.3 into equation 3.2, the structure factor becomes

$$F_{hkl} = \sum_{j}^{Natoms} f_j^{(0)}(\mathbf{k})\ T_j(\mathbf{k})\ e^{2\pi i(\mathbf{k}\cdot\mathbf{r})} \tag{3.4}$$

$T(\mathbf{k})$ is called the Debye-Waller factor. It depends on the displacement of the atoms from their mean positions, which is described by the vector $\mathbf{u}$. $\langle \mathbf{u}_\mathbf{k}^2 \rangle$ is the mean-square displacement projected on the direction of $\mathbf{k}$. At the lowest order approximation, the atoms undergo isotropic displacements from their mean positions and therefore $\langle \mathbf{u}_\mathbf{k}^2 \rangle$ is constant for all $\mathbf{k}$ and $T(\mathbf{k})$ is called the isotropic Debye-Waller factor. If the movement is described anisotropically, $T(\mathbf{k})$ depends on $\mathbf{k}$.

$$T(\mathbf{k}) = e^{-8\pi^2 \langle \mathbf{u}_\mathbf{k}^2 \rangle \sin^2\theta/\lambda^2} \tag{3.5}$$

When transformed into a cartesian basis, the expression becomes:

$$T(\mathbf{k}) = e^{-2\pi^2 \mathbf{k}^\mathbf{T} \mathbf{U} \mathbf{k}} \tag{3.6}$$

Where the matrix $\mathbf{U}$ contains the atomic displacement parameters (ADPs).

$$U = \begin{pmatrix} U_{11} & U_{12} & U_{13} \\ U_{21} & U_{22} & U_{23} \\ U_{31} & U_{32} & U_{33} \end{pmatrix} \tag{3.7}$$

The elements U_{ij} have the dimension of (length)^2.

The atomic scattering factors depend on the electron density distribution of the atom, the more electrons an atom has the more it scatters the X-rays. Therefore, hydrogen atoms with only one electron are very difficult to see, whereas heavy atoms are well visible. More precisely, the atomic scattering factors are defined by the atomic electron density distribution weighted with the correct phase factor

$$f_j(\mathbf{k}) = \int \rho(\mathbf{r}_j) e^{2\pi i(\mathbf{k}\cdot\mathbf{r})} d^3r \tag{3.8}$$

The structure factors for the entire thermally averaged electron density distribution $\langle \rho(\mathbf{r}) \rangle$ of the unit cell are its Fourier transforms: [59]

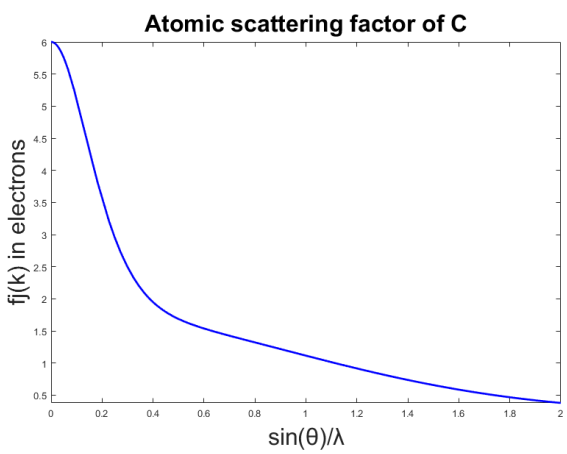
$$F(\mathbf{k}) = \int_V \langle \rho(\mathbf{r}) \rangle e^{2\pi i(\mathbf{k}\cdot\mathbf{r})} d^3r = FT \langle \rho(\mathbf{r}) \rangle \tag{3.9}$$

More diffuse electrons lead to low angle scattering (small value of $\sin\theta/\lambda$) whereas electrons closer to the nuclei lead to higher angle scattering.

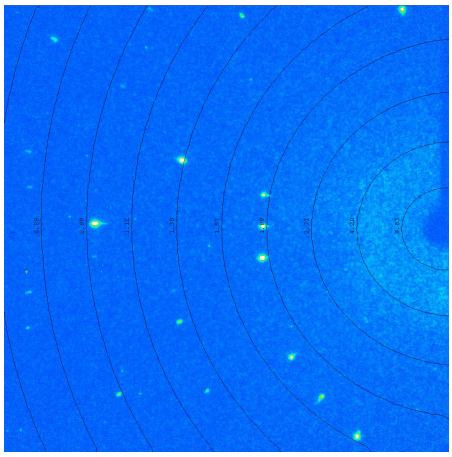
In principle the relationship between electron density and structure factors is mutual and the electron density is simply the Fourier transformation (FT) of the structure factors

$$\langle \rho(\mathbf{r}) \rangle = \frac{1}{V} \sum_{h,k,l} \mathbf{F}_{hkl} e^{-2\pi i(\mathbf{k}\cdot\mathbf{r})} \tag{3.10}$$

However, there are some issues prohibiting the direct transformation of the measured intensities into the electron density distribution of the crystal. [60] First of all, X-rays are not perfectly scattered in a crystal and do not pass without other interactions; in reality other effects than



(a) Structure factor for carbon atom.



(b) Diffraction pattern

Figure 3.2 – The scattering power of an atom decreases with higher angle. On the right side one frame of the diffraction pattern of NTA-B used in chapter 9 is shown.

pure Bragg diffraction are present. Among the most important ones is diffuse scattering which is observed due to a deviation of the crystal lattice from a perfectly regular lattice. Thermal diffuse scattering, in particular, is due the vibration of atoms which brings them out of perfect structural correlation. Besides the elastic and coherent scattering leading to Bragg diffraction, inelastic and incoherent scattering may be present, where the energy of the incident photon is transferred to an electron which gets expelled from the atom. This is called the Compton effect and it can be used to probe the momentum of the electrons. Within the kinematic approximation these effects are neglected assuming that the X-rays are are not modified by the crystal, meaning there is no (thermal) diffuse scattering, no Compton scattering, no multiple diffraction, no anomalous dispersion, no extinction, etc. In addition, X-ray photons can get absorbed by the material if the X-ray wavelength is close to the excitation energy of an atom. This is especially the case for heavier elements which have many electrons. In an isotropic sample, absorption is regular and can be modelled.
Second, the atoms in a crystal are always in movement due to various vibrations and one measures an averaged electron density which cannot simply be described by a static one.
Even more critical are two other issues. One is that the sum in equation 3.10 runs over an infinite amount of reflections hkl, whereas the measurement goes to a given resolution and thus obviously provides only a limited set of reflections. The other one is that the measured intensities $I_{h,k,l}$ are the square of the structure factors $F_{h,k,l}$ and therefore only their absolute magnitude can be obtained and not their phases (under the kinematic approximation).

3.3 Modelling the structure and the electron density

For the reasons introduced in the previous section, it is necessary to model the electron density. Various models have been developed to do that. Some are based on statistics (like the

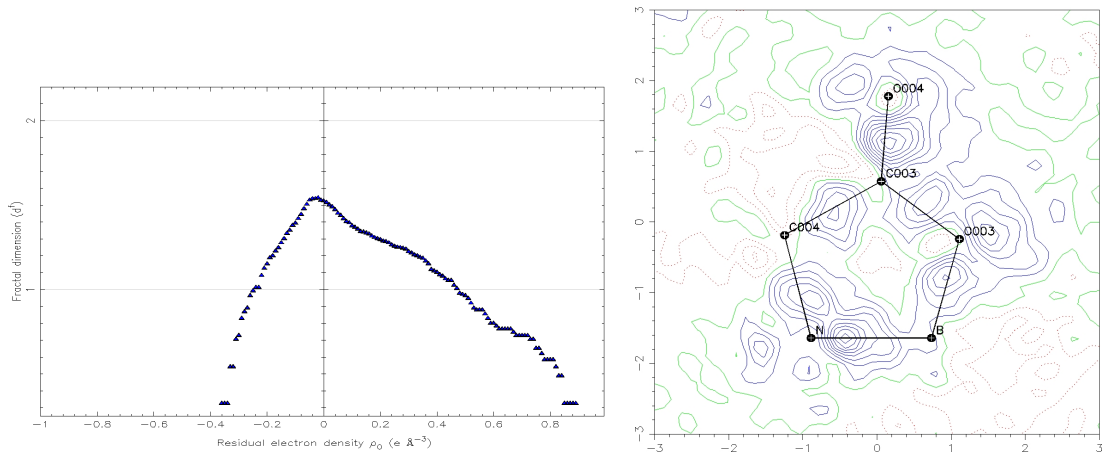
maximum entropy method [61] where the vibrationally averaged electron density is described not the static one), others are based on the least-squares refinement of model parameters to fit the experimental structure factors. The standard procedure to refine structures is based on the independent atom model (IAM). In order to describe the electron density the IAM uses spherically averaged ground-state electron densities for every atoms, which are furthermore non-interacting. This is obviously an approximation because in reality the density is not spherical. Nevertheless, the deformation from sphericity is small compared to the overall accumulation of electrons around the nuclei so that spherical atoms are a good approximation. In a structure refinement based on the IAM the difference between the calculated and the measured structure factors is minimized in a least-squares procedure. The phase of the structure factors stems always from the model. The refined parameters are the atomic positions, the atomic thermal parameters and the scale factor.

Despite the IAM being a good approximation, the fractal dimension plot and the residual density plot in Figure 3.3 prove that there is a systematic error in this model. A substantial amount of electron density is not explained and it is also not randomly distributed since in that case the fractal dimension plot would show a Gaussian distribution. The electron density difference maps reveals, furthermore, that the unexplained density is located in the bonding and lone pair regions. These are aspherical contributions to the electron density not explained with an IAM. The problem is especially pronounced for the hydrogen atom since it has no core electrons and its only electron is involved in bonding. Therefore, the position of hydrogen atoms appears shifted towards their bonding partners when obtained from an IAM of observed structure factors and compared to the real position.
The IAM is often good enough to determine the structure and in many cases, like for example when dealing with structural disorder in the crystal, twinning, limited amount of reflections or limited resolution, it is the only possibility. However, if the goal is a description of the electron density as well, an aspherical atom model is necessary.

In order to describe the electron density remaining after an IAM with spherical form factors, one can use the multipolar model (MM) [62]. The so-obtained information about the electron density allows the examination of bonds and interactions in the crystal. The models can also be used for the extraction of properties or for comparison with electron densities from theory or other experiments. Electron densities obtained with these models are considered "experimental" despite the use of theoretical information in the calculation of the atomic scattering factors. A different type of modelling introducing even more theory is the X-ray constrained wavefunction (XCW) calculation [63]. Due to their importance in this thesis, both, the multipolar model and the XCW approach are explained in chapter 4.

3.4 Structure factor simulation

While it is not possible to obtain an unbiased electron density or even a wavefunction from a limited set of structure factors, the inverse is possible. From a calculated wavefunction for a



(a) Fractal dimension plot

(b) Residual density F_{obs} - F_{IAM}. Positive contours are depicted as blue lines and negative ones as red lines. The contour level is 0.1 e/Å^3.

Figure 3.3 – Electron density not explained by an independent atom modelling of NTA-B.

given crystal one can simulate the structure factors up to any desired angular resolution. The purpose of this can be comparison to experiment. However, this procedure was used for many more reasons throughout the thesis.

One reason to simulate structure factors is to replace experimental ones in case they are not available or the quality is too poor. This might be the case when no crystal can be obtained or when the crystal or the measurement are not good enough for a charge density analysis. Furthermore, structure factors can be simulated for crystals under high pressure for pressure ranges, which cannot be obtained experimentally.
Additionally, simulated structure factors allow for including or excluding specific effects such as correlation [64] or the crystal field (see chapter 9).

In *CRYSTAL17* [23] an algorithm to simulate structure factors of periodic systems is implemented. In order to simulate structure factors of molecular wavefunctions, molecules were placed in a fictive unit cell, which is big enough for that there are no interactions between the molecules in different unit cells (10x10x10 Å^3 if not otherwise specified). The simulation was performed with the structure factor analysis module of XD [65].

Part II

Introduction: Theory of Quantum Crystallography

4 Quantum Crystallography

4.1 History of combining X-ray diffraction and quantum chemistry

About one century ago, scientists started to develop quantum mechanics and to understand the probabilistic laws that govern the world of molecules, atoms and smaller particles; laws which cannot be grasped by classical mechanics. Based on these discoveries the field of quantum chemistry was evolving (see chapter 1).
At a similar time, the foundations for the structure analysis of crystals with X-rays were laid, for example by Max von Laue who discovered the diffraction of X-rays by crystals [66] and William Henry Bragg and his son William Lawrence Bragg who formulated the laws for X-ray diffraction [67].
About half a century ago, first attempts to combine the equations of theoretical quantum mechanics with the results of various experiments were undertaken. Obviously experiments were always needed to ensure that the developed theories actually describe reality. After all, a central point of quantum mechanical calculations are expectation values which, by definition, are related to measurable quantities. Previously however, calculation and experiment had been two clearly distinct thing which had not directly been obtained at the same time. Yet, in that time researchers started to actually introduce experimental observations directly into their quantum mechanical calculations. Subsequently, the focus will mainly be on the combination of quantum mechanics and X-ray diffraction, thanks to which the field of quantum crystallography evolved.

Since X-rays are scattered by electrons, they contain information about the electron density distribution in a crystal. The electron density is highest in the core region of the atoms and this is why the independent atom model, which assumes spherical atoms, works well for solving structures. In fact, the additional density which is in the lone pair and bonding regions contributes much less to the total intensity of the reflections such that it was for a long time not possible to gain information about bonding electrons. Yet, it is to be mentioned that for example Bragg already observed the 222 reflection of diamond, which is forbidden

when assuming spherical carbon atoms, and he correctly attributed his observation to some "directional property" of the atoms [68]. When the quality of X-ray diffraction measurements increased, scientists realized that such information should be accessible and the striving for obtaining information of the actual electron density distribution in crystals arose.

Two differing strands developed. One strand is making use of density matrices and wavefunctions which are supplemented with or fitted to experimental data (in particular X-ray diffraction data). This method was initiated by Clinton and Massa [69] who worked with idempotent density matrices constraint to experimental data and it was further developed into the X-ray constrained wavefunction approach by Jayatilaka [63]. The other strand was pioneered by Stewart and Coppens and is concerned with the fitting of (mainly) atom-centred functions to experimentally measured diffraction data. Initially these approaches relied on the availability of predetermined coordinates and thermal parameters for example from neutron diffraction or an IAM and they were only used to obtain the electron density distribution. Later, the multipolar refinement (section 4.2) or the Hirshfeld atom refinement (HAR) (section 4.5) also allowed to refine positions.

Nowadays the goals of quantum crystallography are manifold. [70] Within a narrow definition, they can be summarized into four points:
I) Improving the structural information which can be obtained from crystallographic refinements.
II) Obtaining not only positions and displacement parameters but also the electron density $\rho(\mathbf{r})$ and related quantities in order to derive properties from it. This includes also derivatives of the electron density like the Laplacian $\nabla^2\rho(\mathbf{r})$ and topological properties of the $\rho(\mathbf{r})$ distribution as well as for example the electrostatic potential $V(\mathbf{r})$. Instead of a charge distribution also the spin density or momentum density distribution can be the goal.
III) Obtaining an experimental wavefunction or an otherwise improved wavefunction.
IV) Comparing theory and experiment in order to challenge, confirm, and improve each other.

Quantum crystallography can create a link between the macroscopic world of the crystal and the microscopic world describing the behaviour of the molecules. The availability of an electron density or a wavefunction allows to relate the building blocks a material consists of and the interactions among them to the macroscopic features, a very important point for crystal engineering, since understanding these mechanisms and thus connecting structure and function allows to design functional materials.

The results presented within the scope of this thesis, mostly use measurements with elastically scattered X-rays for the crystallographic part, however, this does not need to be so. Other scattering techniques like inelastic Compton scattering and diffraction with γ-rays, electrons, and (polarized) neutrons are widely used. These other techniques not only measure the charge density but also the momentum density or the spin density (see section 4.6). In addition, different experiments like spectroscopy, electrical or magnetic measurements can be used to study crystals and all of them can be combined with other experimental or theoretical

parts to probe a system. Crystallography can also refer to ab-initio calculations with periodic boundary conditions, a method often employed in this thesis and described in section 1.4. On the other hand, the quantum part does not necessarily need to be computational since many experiments directly probe the quantum behaviour of materials. In any case theory was always needed to build models from measured data for example to compensate for the missing information about the phase in X-ray diffraction experiments.
The field of quantum crystallography [71, 72] is broad, a lot of research is going on under the umbrella of this term and there are many links to other scientific branches. Different viewpoints on quantum crystallography as well as an overview about the current disciplines are given in a recent article [70].

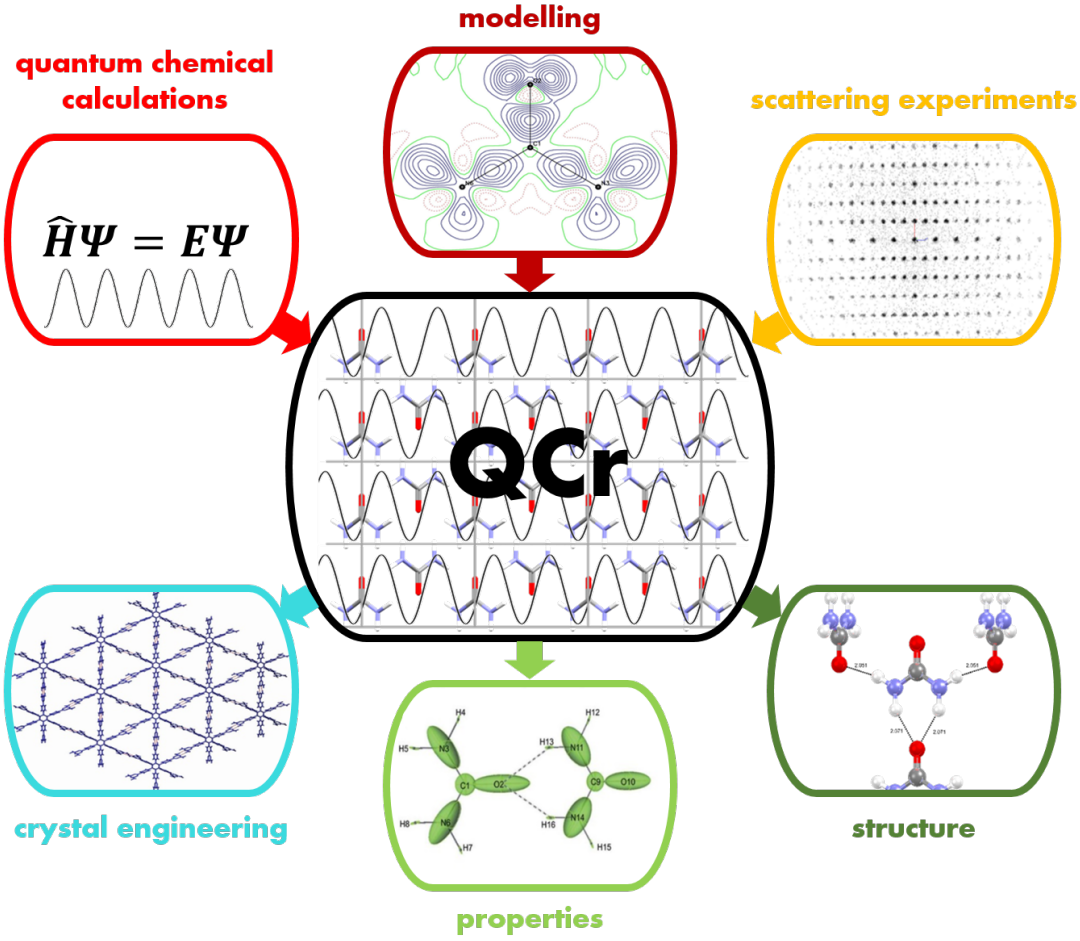


Figure 4.1 – Some inputs and outputs of quantum crystallography (QCr)

4.1.1 From the first quantum mechanical scattering factors to more elaborated ones based on atom-centred aspherical functions

The first atomic scattering factors in use were based on the assumption that electrons behave according to the classical laws of electrodynamics. But soon after Hartree presented his method for the calculation of wavefunctions in a self-consistent way, scattering factors were obtained based on quantum mechanical calculations; for example by James and Brindley [73] who used Hartree wavefunctions (and not Hartree-Fock ones, since Fock published his article about the calculation of correctly antisymmetrized wavefunctions only in 1930). Although they were aware that "*in any real crystal structure, the arrangement of the outer electrons of the atoms must be to some extent distorted from spherical symmetry*" they tabulated spherically averaged scattering factors. Spherical scattering factors remained the ones used until McWeeny published a series of four papers from 1951 to 1954.
In the first paper [74] he calculated atomic scattering factors based on wavefunctions for 1s, 2s, 2p and thereof hybridized atomic orbitals, where the proposed scattering factors for atoms with non-spherical electron configurations were indeed **aspherical**. In the second paper of the series [75], additional scattering factors for bonds were introduced. Together with the scattering factors of the atoms they gave rise to effective atomic scattering factors which were not spherical and explicitly took into account electrons in the bonds. He performed calculations for the hydrogen molecule where the effect of non-sphericity is largest since all electrons are involved in bond. Since these calculations were quite laborious he introduced in the third paper [76] Gaussian type orbitals for expressing the Slater type orbitals. In the fourth paper [77] he applied the formulas to the carbon atom in diamond and graphite. Since he calculated the structure factors for atoms at rest, he noted that at this point the most significant improvement would come from including thermal vibration.

Kleinman and Phillips [78] as well as Weiss [79] and later Dawson [80] also calculated aspherical scattering factors and used them for the calculation of the structure factors of diamond; specifically for the prediction of the 222 reflection which, with purely spherical scattering factors, would be absent. De Marco and Weiss [81] expanded this analysis also to silicon and germanium and obtained results for the 222 reflection of silicon which were in good agreement with experiment whereas no definite conclusions about germanium could be drawn since in that case the contribution of the valence electrons is too small compared to the core electrons.

Stewart, Davidson and Simpson [82] showed in 1965 that when the hydrogen scattering factors are taken from isolated hydrogen atoms, too low temperature parameters for bonded hydrogens as well as positions too much centred into the bond (compared to results from neutron diffraction or spectroscopy) are obtained. They calculated new scattering factors for a polarized hydrogen atom in H_2. The predicted form factors were used for **least-square fits** of positions and temperature factors.

In 1969, Stewart [83] presented a paper in which he calculated generalized scattering factors as Fourier transforms of atomic orbital products, where the two functions can be lying on the

same atom or on two different atoms. The structure factors obtained from these scattering factors were multiplied with a population parameter and compared to experimental structure factors. With the least-squares method the population parameters were refined. Orbital products among 2s and 2p AOs on the same atom were used, therefore leading to atomic monopole, dipole and quadrupole terms. In 1973 Stewart [84] expanded the method to 3d atomic orbital products extending the deformation terms to fourth order spherical harmonics. In the conclusions of this paper he also mentioned the possibility of using instead of atomic orbital products a **multipole expansion** in spherical harmonics where the radial part r^k is expanded with a multiplicative function $\exp(-\alpha_k r)$ where α_k is a variable parameter in the least squares equations. In that case, the electron density in the chemical bonds is accounted for by atom-centred terms and not by one or two-center atomic orbital products. In another paper in 1972, Stewart [85] calculated physical properties based on this model with refined population parameters.

In a series of papers in 1971, Coppens *et al.* [86, 87, 88] further developed the AO product based description of structure factors in order to extract information about bonding in molecules in a quantitative manner. They stated that the success of the spherical atom model is partially also due to the fact that thermal factors and positions erroneously include effects of electron density which is in bonds. Two different models were tested, one with only one-center terms and one including also products of atomic orbitals on two centres. They obtained expressions for atoms (one center) and for bonds (two center). The one and two-center atomic orbital products were multiplied with a population parameter and a temperature factor giving rise to structure factors for both the atoms and the bonds. The population parameters were refined with least-square fits to the diffraction data. The correlation between the population parameters and the conventional least-squares parameters might be large, wherefore they took the positional and thermal parameters from neutron diffraction instead of refining them and the total number of electrons was constrained, leaving only the scale factor and the electron population parameters to be refined. Only a very limited amount of parameters was used (symmetry was exploited) but they had nevertheless problems with correlation of parameters. For example, the contribution of the $2s^2$ term and the sum of $2p_x^2$, $2p_y^2$, $2p_z^2$, which is also spherical, can not be distinguished and thus the $2s^2$ parameter was arbitrarily set to some value.

It turned out that the approach with atomic orbital products on both, the same and different centres might be worse compared to **atom-centred expansions of the electron density**. An atom-centred expansion of the electron density with spherical harmonics was already proposed in 1968 by Kurki-Suonio [89], though without refining any population parameters. In 1971 Hirshfeld [90] further developed this approach. He defined a deformation function $\rho = \rho_s + \delta_\rho$, where ρ_s is the sum of spherical free-atom densities and δ_ρ the deformation density which is expanded with functions also centred on atoms. Stewart introduced what he called the pseudoatom model. A pseudoatom consists of all atomic expansion terms of one atom. With these multipolar terms the scattering factors can be obtained and then selected coefficients of the multipole model can be refined by a least squares fit to the experimental

structure factors. In two articles in 1975 and 1976 Stewart summarized his efforts about the pseudoatom model. [91, 92]

In 1978 Hansen and Coppens [62] introduced the multipolar model, which became the most widespread model for describing aspherical atom densities. It is explained in section 4.2.

4.1.2 Density matrix and wavefunction refinement

As written in the previous section McWeeny published on one hand a series of papers (1951-1954) [74, 75, 76, 77] where he proposed for the first time aspherical atomic scattering factors to model the electron density from X-ray diffraction measurements and his ideas were the starting point for Stewart, Hirshfeld, Coppens and others to develop the formulas that lead to the Hansen and Coppens model. On the other hand, McWeeny also published a paper in 1960 [3] about the purification of a non-idempotent density matrix which was the initialization for Clinton and Massa's developments of idempotent density matrices that fit the X-ray structure factors. The idea behind their approach is based on earlier attempts to include experimental information into the calculation of WFNs and density matrices.

Since WFNs are not physical observables they were and are typically obtained from pure quantum mechanical calculations using the variational principle and thus via minimization of the total energy. With an infinite amount of determinants and expansion functions one could in principle obtain the correct WFN and energy of a given system, see Figure 1.2. In practice, this is obviously not possible and one has to deal with different approximations. The best approximate WFN in terms of energy must not necessarily be the best one for the electron density distribution and thereof resulting properties. Therefore, the idea of introducing additional constraints for the calculation of the WFN arose. Besides the Hamiltonian, additional operators are applied on the WFN and their expectation values are requested to be consistent with experimental values or values from complementary theories. In principle the expectation values could be forced to be exactly identical to the constraint but in practice the expectation value is usually constrained such that the difference to the experimental value is lower than a particular value. Thus, during the variational procedure not only the energy is minimized but also the difference to the experimental constraints.

Mukherji and Karplus (who later won the Nobel Prize for the *development of multiscale models for complex chemical systems*) are among the first researchers calculating experimental WFNs. [93] They used the dipole moment and the deuteron quadrupole coupling constant as external constraints for the calculation of a WFN for the hydrogen-fluoride molecule. The deuteron quadrupole coupling constant is a constraint for the electron density close to the nucleus whereas the dipole moment constrains rather the outer region of the charge distribution. As a starting guess they used the best WFN from an unconstrained calculation with a minimal basis set expressed as a linear combination of atomic orbitals with associated expansion coefficients c_{ip}. Subsequently, they optimized the coefficients to obtain a new WFN fitting including the constraint. Some of these coefficients were reserved to ensure the correct normalization of

the WFN and some to ensure the orthogonalization of the orbitals. Only one coefficient was used to minimize the difference of the calculated dipole moment and deuteron quadrupole coupling constant to the experimental one while all other coefficients were only indirectly affected by the constraint. The new coefficients were determined by several minimization cycles.
With this approach, the difference of the expectation value of the additional operators to the experimental constraint was considerably reduced while the energy changed only little. Furthermore, the constrained WFN resulted in improved agreement with other experimental properties not used as constraints.

Rasiel and Whitman (1965) [94] developed a more generalized procedure to introduce *empiricism* (as they call it) into WFNs. They expanded the constrained WFN from the solution of the unconstrained (or free-variational as they call it) WFN. The constrained part was weighted with a Lagrange multiplier and this modified equation was solved. Specifically, they performed calculations for lithium hydride and used the experimental electric dipole moment as constraint. The constrained WFN removed a 12% error in the dipole moment while the energy sacrifice was only 0.06%. Other properties obtained from the constrained WFN were also improved with respect to the unconstrained WFN.

In contrast to the initial approaches of Mukherji and Karplus, Rasiel and Whitman, and Byers Brown [95] who worked with the wavefunctions, Clinton, Nakhleh, and Wunderlich [96] were determining the one-electron reduced density matrix $\rho_1\left(\vec{r'};\vec{r};\vec{R}\right)$. In analogy to the wavefunction-based approaches, they used empirical data to constrain the first-order RDM. However, in the first out of a series of five papers, they did not use experimental data as constraints for the determination of the density matrix but physical theorems. These were the electrostatic theorem, the virial theorem and the "density-matrix" theorem (meaning that the total amount of electrons must be correct). $\rho_1\left(\vec{r'};\vec{r};\vec{R}\right)$ was inserted into the integral equations of these theorems. In order to be able to solve the equations, different forms for the density matrix can be assumed and they chose atomic orbital functions to expand it. The so-obtained semi-empirical densities were not N-representable, meaning they could not be represented by a proper WFN and analysis of natural orbitals does not result in occupancies always between 0 and 2 as it should be. They mentioned the possibility to circumvent this problem by directly calculating the WFNs, however, they stated that this would lead to unnecessary numerical problems and they favoured the first-order reduced density matrix.

In the second paper of the series Clinton, Galli, and Massa [97] proposed a method based on McWeeney density matrix purification method to make the density matrix idempotent. In an iterative procedure, a matrix was purified:

$$P_{n+1} = 3P_n^2 - 2P_n^3 \tag{4.1}$$

Herein the constraint equations were introduced with a Lagrangian multiplier λ:

$$P_{n+1} = 3P_n^2 - 2P_n^3 \sum_k \lambda_k^{(n)} O_k \tag{4.2}$$

The solution of these equations was not an energy minimization procedure, nevertheless, for very simple cases with only two variables, they anyhow reproduced the Hartree-Fock density matrix when they took as constraints Hartree-Fock calculated values.
In paper three [98] only theoretical constraints were used. Specifically, a combination of the electrostatic and the virial theorem into a single differential equation. Infinitely many densities would fulfil this equation. Therefore a known density was taken and scaled and only the scale factor was optimized. They demonstrated with diatomic molecules that the so obtained scale factors are in qualitative agreement with the variational ones. In paper four [99] they used a specific theorem of Parr as a constraint on their idempotency equation and in paper five [100] the equations were applied to the calculation of local energies of H_2^+ and He. In a later paper Clinton and Massa [101] used the cusp condition as constraint on the electron density matrix.

Henderson and Zimmerman [102] later also made use of the Clinton formula. They demonstrated that these equations can be fulfilled by several idempotent solutions for the density matrix and thus require an additional criterion to be solved uniquely. For this they introduced an energy minimization criterion. However, they did not exert any variational minimization but basically error and trial. They proposed to take a set of operators and calculated initial guesses for the expectation values of these operators (so no experimental or theoretical constraints were introduced). To solve the Clinton equations they took these expectation values as constraints and calculated the energy from the resulting density matrix. With this density matrix a new guess was obtained, the procedure was repeated and in the end the solution with the smallest energy was taken.

In 1972 Clinton and Massa [69] used for the first time coherent X-ray diffraction data to obtain the electron density of the unit cell and thus of the molecules via a density matrix fitting. Like in the previous papers, they imposed the quantum-mechanical constraint of N-representability on the density matrix. The structure of the crystal was supposed to be known. The elements of the density matrix were determined by least-squares fit to the scattering factors using the Clinton equations. Therefore, the density matrix has a form automatically ensuring N-representability and it fits the diffraction data:

$$P_{n+1} = 3P_n^2 - 2P_n^3 + \sum_{\vec{G}} \lambda^n(\vec{G}) f(\vec{G}) \tag{4.3}$$

They also already thought about lattice vibration effects which they introduced by multiplying the scattering-factor data with a thermal factor.

In 1978 Frishberg and Massa [103] worked out some advantages of this approach. Due to the N-representability condition they faced less parameter correlation problems than Stewart, Coppens and others with their approach of atom centred basis functions to expand the

electron density. In addition, they dealt with less independent parameters. The solution of the iterative equations is a purified and thus N-representable density matrix additionally constraint to diffraction data and thus in accordance with structure factor data by the least squares criterion. Although the scattering factors depend only on the electron density (diagonal elements of the one-electron reduced density matrix), also the off diagonal elements are obtained because the idempotency criterion also includes the off-diagonal elements. They did exemplary calculations using structure factors of the H_2 molecule, once with an idempotent and once with a non-idempotent density matrix. Although the latter led to a lower R-factor, it was associated with problems (correlation of parameters, non-physical meaning of exponents, negative density, higher number of parameters) which were not there for the idempotent density matrix.

In 1995, Massa, Huang and Karle [104] published a paper in which they used for the first time the term "*quantum crystallography*". They specified that this is a method for "*combining X-ray diffraction data for crystals and quantum mechanics with the objective of obtaining accurate wave functions that are consistent with the X-ray data*" and thus "*extracting quantum mechanically valid properties from X-ray diffraction experiments*". They did this again in terms of an N-representable density matrix expressed with atomic orbitals. The Fourier transform of the density matrix was least-squares fitted to structure factors. Thus, this first definition highlights the opportunity of improving a wavefunction with crystallographic information.
In a later paper [105] the same authors expand the definition of quantum crystallography, specifying that "*conversely, quantum mechanics, with the aid of crystallography, can greatly enhance the information available from a crystallographic experiment.*" Thus, also pointing out the possibility to improve the crystallographic description of a system with the use of quantum mechanics.

The one-electron RDM which is the central component of the Clinton equation and Massa's early quantum crystallographic ideas can, however, not only be refined against X-ray diffraction data. Compton scattering also offers itself for that purpose since it contains information about the off-diagonal elements of the 1-electron RDM (in contrast to X-ray diffraction which measures the electron density, that is the diagonal part of the RDM). The off-diagonal elements are related to the momentum of the electrons and thus, the 1-electron RDM is common to position and momentum space (see section 4.6).

Eventually, a different method was born developed by Jayatilaka in 1998 [106]. In that paper he described for the first time the refinement of an entire wavefunction and not reduced density matrices against X-ray data. He termed this the X-ray constrained wavefunction (XCW) refinement. More details of this method are explained in section 4.4.

4.2 Hansen-Coppens multipolar model

The multipolar model published in 1973 by Hansen and Coppens [62] is probably the most widely used method to obtain the electron density from an X-ray diffraction experiment.

The description of the electron density with a multipolar model is done with three different terms: A spherical core density ρ_{core}, a spherical valence density $\rho_{valence}$ and real spherical harmonics y_{lm} to describe the asphericity of the valence electron density.

$$\rho_{atomic}(\mathbf{r}) = P_c\rho_{core} + P_v\kappa^3\rho_{valence}(\mathbf{r}) + \sum_{l=0}^{l_{max}} \kappa'^3 R_l(\kappa'_l\mathbf{r}) \sum_{m=-l}^{l} P_{lm}y_{lm}(\mathbf{r}/r) \tag{4.4}$$

P_c, P_v, and P_{lm} are population coefficients. The sum of P_c, P_v, and P_{00} is the number of electrons of one atom because the multipolar P_{lm} terms with $l \neq 0$ integrate to zero. ρ_{core} and $\rho_{valence}$ are Hartree-Fock densities of the free atoms. The spherical valence function can be expanded and contracted by adjusting the κ parameter and the radial part of the deformation valence density with κ'.
The radial function is $R_l(r) = \frac{\zeta^{n_l+3}}{(n_l+2)!} r^{n_l} e^{-\zeta_l r}$ and the angular functions y_{lm} are real spherical harmonics like also used for atomic orbitals. They are normalized.

As written in section 3.2, the structure factors are the Fourier transforms of the thermally smeared electron density $\langle\rho(\mathbf{r})\rangle$. In analogy to the Born-Oppenheimer approximation, it is assumed that the electrons move instantaneously with the nuclei and are not affected by the movement of the nuclei. The nuclei, in turn, are displaced from their equilibrium positions due to lattice vibrations. This nuclear displacement can be described with a harmonic or anharmonic tensor $T_k(u)$, where u is the displacement from the equilibrium position and T a probability function. The structure factors are then calculated as the product of the two functions in reciprocal space (see equation 3.4), which are in this case the electron density expanded in multipoles and the thermal displacement function.

Hansen and Coppens wrote a program called *MOLLY* allowing to perform the least-square refinement of the adjustable parameters in the model which were: atomic positions, harmonic-anisotropic vibration, extinction parameters as well as population coefficients P_v, P_{lm} and radial expansion parameters κ and κ'. The spherical functions in the original Hansen-Coppens paper go up to hexadecapolar terms (l=4). Hansen and Coppens performed different tests to ensure that the improved agreement indices of the refined multipolar model are due to a better description and not just because of the increased number of parameters involving a larger flexibility of the model. They found that in their examples the fit is physically meaningful, a point that should be checked for every multipolar model.

In the standard procedure for a multipolar model the expansion still goes to hexadecapolar terms and ADPs can in principle be refined up to fourths order. The total charge is constrained to zero such that the crystal remains neutral. The number of parameters can be reduced by imposing crystal and chemical symmetry. Various software such as XD [65], VALRAY [107], MollyNX (joint charge and spin density refinement) [108], MoPro [109], and JANA2006 [110] was developed over the years allowing for slightly different refinement procedures.

Thus, in a multipolar model the electron density is expanded in atom terms (atom-centred multipolar functions) and no bond functions are used; or to put it differently, only one-center

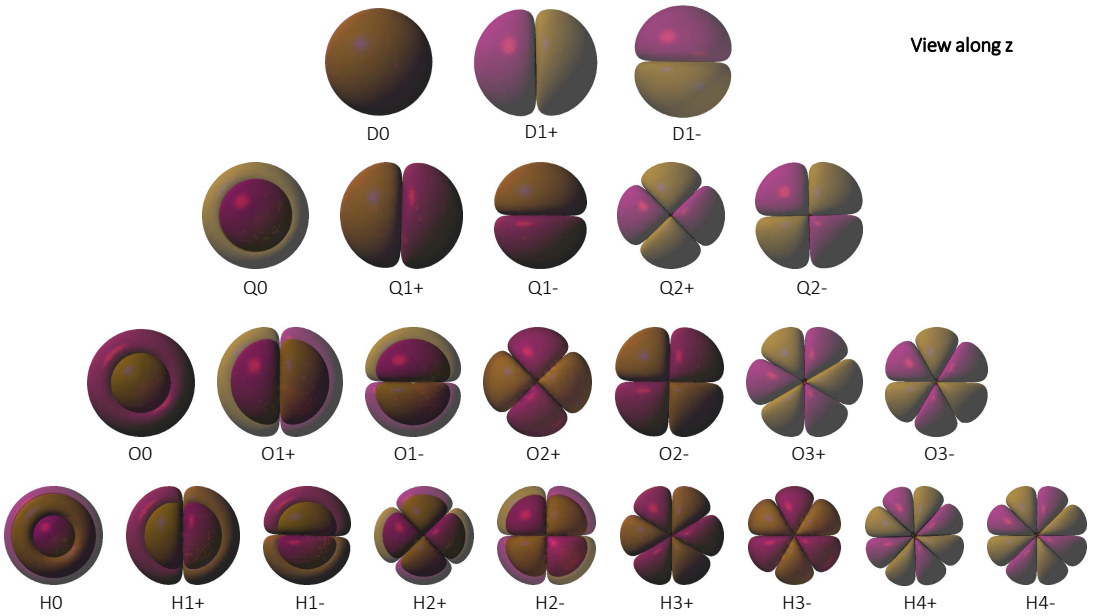


Figure 4.2 – Dipolar, quadrupolar, octopolar, and hexadecapolar functions used in a conventional multipolar refinement

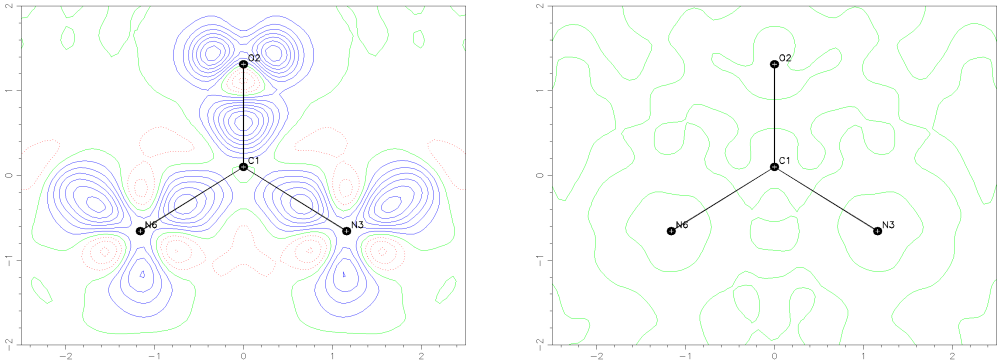
terms and no two-center terms are used. Due to the flexibility of the model it is nonetheless possible to describe the electron density in the bonds. However, because of the increased number of parameters with respect to the independent atom model, the data quality and the angular resolution (until which $\sin\theta \setminus \lambda$ reflections are measured) have to be higher than in a standard measurement.
If the data quality is excellent and the number of measured reflections very high, it is possible to add more parameters to the model which can be chosen depending on the interest. One can for example refine the core population [111] or add independent radial functions or κ's to every spherical harmonic. However, the reflection to parameter ratio should never be lower than 10, [59] otherwise parameter correlation is too large. In contrast to the X-ray constrained wavefunction calculations (see section 4.4) it is clear what is experimental data and where it enters the model. In addition, the influence of individual reflections on the model can be calculated. [112]

Since all functions are centred on one atom, atomic properties are easily obtained with the Hansen-Coppens multipolar model and chemical interpretation is straightforward, in contrast to the MO model where orbitals are spread over the entire molecule. The multipole parameters are defined with respect to a local coordinate system of every atom making it easier to account for local symmetry. This also allows multipoles to be relatively easily transferred between atoms in similar environments which led to the development of pseudoatom databases described in the next section.

Hydrogen positions can be constrained to average distances obtained from neutron measure-

ments. Hydrogen ADPs can be obtained with the SHADE web server. [113, 114] This is done by combining external and internal vibrations. The external contribution is obtained by a TLS (translation-libration-screw) analysis of the non-hydrogen-atom framework leading to a rigid body model that can be imposed on the hydrogen atoms. [115] The internal mean square displacements of the hydrogen atom corresponds to the sum of the contributions from the internal vibrational modes. These are obtained from a database derived from the analysis of neutron diffraction studies of organic compounds. Therefore, the hydrogen ADPs are calculated as: $U^{ij} = U^{ij}_{rigid} + U^{ij}_{internal}$

The multipolar model allows to obtain the electron density as it is in the crystal, that is including the effect of interactions with surrounding molecules, the effect of the crystal field, of temperature [116] or pressure [117]. Due to its flexibility it can recover electron density in the bonds, lone pair regions, interaction regions, etc. Its effect can be visually evaluated by plotting deformation density maps which show the difference between the electron density of the multipolar model and the IAM (sum of spherical atoms).



(a) Deformation density (F_{MM} - F_{IAM}) for the urea molecule (theoretically simulated structure factors).

(b) Residual density after a multipolar refinement (F_{obs} - F_{MM})

Figure 4.3 – Positive contours are depicted as blue lines and negative ones as red lines. The contour level is 0.1 $e/Å^3$.

4.3 Databases of multipolar pseudoatoms

The advantage of the multipolar model is that aspherical atoms can be refined and thus effects of bonding are taken into account. However, this advantage also goes along with a cost, which is the considerably increased number of parameters compared to the IAM. In order to have a reasonable data to parameter ratio, the number of measured reflections must be high and thus the measurement needs to be performed up to very high resolution. By far not all crystals diffract good enough to allow for that. To circumvent the problem, pseudoatom databases were introduced. Instead of freely refining all parameters in the model, one relies on tabulated values for the multipole parameters which vary depending on the atom type. Thereby, it is possible to obtain aspherical atom densities without increasing the parameter number.

Moreover, the refinement of very large molecules such as proteins becomes possible.
Based on the charge density derived from the multipoles one can calculate the aspherical scattering factors. Often, a normal multipolar refinement of positions, ADPs, and multipole parameters is performed after importing the database parameters. In order to set up the database, multipoles for given atom types are to be refined in different systems and the average is the representative of this atom type. An atom type is usually not just a particular element but rather an atom in a specific chemical environment, for example with defined first neighbours. In order to make the database parameters transferable from one system to another one, a local coordinate system needs to be defined, which is independent of the crystal coordinate system. For minimizing the number of different atom types, as few model compounds as possible should enter the database while at the same time one needs to ensure that there are as many as needed to keep a high accuracy of the result. These guiding concept was also used for the construction of a polarizability database described in chapter 6.

4.3.1 ELMAM

Pichon-Pesme *et al.* [118] set up the first database with multipole parameters. They refined multipole parameters for three small amino acid derivatives and used them for the refinement of a tripeptide molecule, which resulted in better agreement statistics than from a spherical atom refinement.

These initial attempts led to the development of the ELMAM database where the acronym stands for 'experimental library multipolar atom model' [119]. For all atom types occurring in the proteinogenic amino acids a database entry was obtained based on multipolar refinements of high-resolution X-ray diffraction data from small peptides and amino acid crystals. Afterwards, the multipolar parameters can be transferred to other peptides measured at lower resolution and used for a refinement. Thereby the crystallographic statistics is improved.

Later, the ELMAM database was generalized into the ELMAM2 database which contains not only parameters for atoms in peptides but for many common functional groups of organic chemistry. [120] Furthermore, the assignment of the local coordinate system was improved. [121] A good local coordination system is important because it allows to have less multipolar parameters by exploiting the local symmetry. The origin of the coordinate system is always located on the atom for which the multipolar parameters are stored and oriented with respect to the neighbours of this atom. They describe seven different ways of defining the level coordinate system depending on the atom type.
When setting up the polarizability database for functional groups (see chapter 6) we also had to choose a way to define local coordinate systems because polarizabilities are tensorial properties and therefore the coordinate system is important to express the magnitudes in every direction. To make the polarizabilities transferable from one molecule to another, they need to be fully independent of the coordinate system of the molecule and only expressed in a local coordinate system of the considered functional group itself. For functional groups

consisting of only two atoms, the choice of the local coordinate system was inspired by the method of Domagała *et al.* [121] using the direction of one bond (**u**) as first axis (I_1), then the cross product of **u** and the direction **v** of a second bond to obtain $I_3 = I_1 \times \mathbf{v}$. And the second axis is $I_2 = I_3 \times I_1$. For functional groups with more than two atoms, we chose to express the local coordinate system based on the intertial tensor as described in chapter 6.

4.3.2 UBDB

While the parameters in the ELMAM databases are based on fitting to experimental charge densities, the parameters in the University at Buffalo Database [122] are based on fittings to calculated charge densities. The authors did single-point calculations on experimental structures (with bond distances to hydrogen atoms normalized to neutron distances) for small organic molecules and then fitted them with a Hansen-Coppens multipolar model to set up the database. Atom types were defined based on the first and second neighbours. The database entry for one specific atom type was obtained by averaging over all calculated values for this atom type. It was checked whether the parameters within one type are homogeneous enough and if this was not the case a new type was defined. The procedure of transferring parameters of a multipole model describing charge densities of aspherical atoms was termed TAAM (transferable aspherical atom model).
The database was expanded in 2011 [123] to include more atom types present in biologically relevant molecules.

4.3.3 Invariom

The generalized invariom database [124] contains entries for different invarioms. Invarioms are unique atom types defined by their first and second neighbours according to specific rules. Therefore, molecules can be described by a string containing invarioms which have been assigned according to the structure and bond lengths. For every invariom a unique compound is defined by saturating all unused neighbours with hydrogen atoms. This compound is geometry optimized and then a multipolar model is fitted to the *ab-initio* density. The database is made for organic molecules mainly based on H, C, N, O but including also some other atoms. The local coordinate system is also defined based on the neighbouring atoms.

4.3.4 Comparison

The common goal of all databases is to use transferable aspherical scattering factors and improve the structure refinement. They are all designed for organic molecules. ELMAM relies on experimental charge densities whereas UBDB and invariom use calculated ones. However, UBDB uses experimental structures whereas invariom is based on geometry-optimized ones. Invariom has strict rules to define atom types and every atom type is unique such that the multipolar parameters do not need to be averaged. In ELMAM2 two atom types are identical

when chemical type, type of local axes, number and type of first neighbours, and planarity are equal. UBDB defines atom types according to neighbours and evaluates how homogeneous the parameters are; when they get too different a new type is defined. Bąk *et al.* compared structures and properties obtained with the three different databases. They did not find any profound differences in the structures and ADPs, while differences in the dipole moment and electrostatic interaction energies were larger. [125]

Table 4.1 – Comparison of pseudoatom databases

	ELMAM2	UBDB	generalized invariom database
structure	high resolution structures from selected compounds, X-H distances set to average value from neutron diffraction	from CSD, X-H distances set to average value from neutron diffraction	geometry optimized
multipoles of Hansen-Coppens model fitted to:	experimental structure factors	electron density from DFT single point calculations on experimental geometry	electron density from geometry optimization
definition of atom types	atom types based on chemical type, local axes, number and type of direct neighbours, planarity. In certain cases also second- and third-nearest neighbours are considered. Averaging over same atom types	based on chemical environment and statistical similarity criterion	unique invarioms: an atom type is defined based on its neighbours and in certain cases next-nearest neighbours, which are then saturated with H.

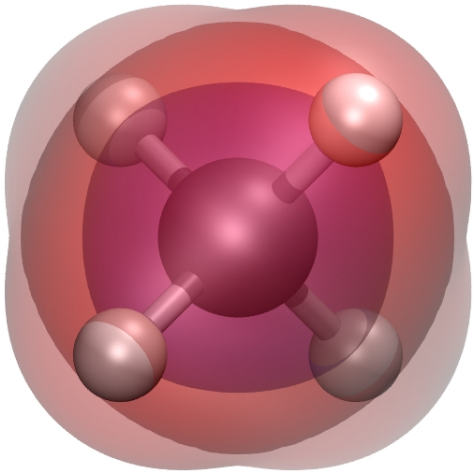
4.4 X-ray constrained wavefunctions

4.4.1 Alternative minimization criteria

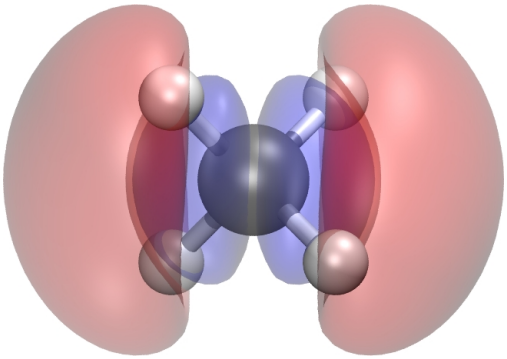
The variational principle states that the best wavefunction for a given system is the one leading to the lowest energy. For a perfect system calculated with an infinite basis set and an infinite number of Slater determinants this is true and properties calculated from such a wavefunction would be correct. Yet, in practice we always deal with approximate systems with truncated basis sets and low number of determinants and in these cases the wavefunction leading to the lowest energy is not necessarily the best one. Properties obtained from such a wavefunction are not completely accurate and it is possible to improve them within the same level of theory. A HF calculation, for example, leads to the best orbitals within the one-determinant approximation and with the given basis set. However, it is known that these orbitals are too diffuse

as shown in Figure 4.3. When electron correlation is introduced, the orbitals as well as the electron density contract and get closer to their real optimum form. Indeed, the orbitals from a DFT calculation (which results also in just one Slater determinant) are actually more contracted than the HF ones at the expense of increasing their energies. The effect of correlation keeps the electrons away from each other making them see the nucleus less screened and thus the orbitals less diffuse. In a Hartree-Fock calculation the electron-nucleus attraction is not negative enough while the kinetic energy and the electron-electron repulsion is also artificially minimized and thus the orbitals more diffuse. [14]

Thus, despite the HF solution being the one with the lowest possible energy in this framework, the HF-orbitals are not necessarily the best ones. Therefore, the lowest-energy criterion employed in the variational principle compensates for errors (finite basis set, number of determinants) by introducing other errors just to artificially lower the total HF energy. As a side remark it is noted, that the total DFT energy only appears lower than the total HF one because it is calculated in a different way. If the energy of a DFT wavefunction was calculated in the HF way it must necessarily be higher than the HF one (see Table 4.2).



(a) Difference between HF and DFT density of methane. Red signifies more density in the HF case and blue more in the DFT case. Isocontour level = 0.000465 $e/\text{Å}^3$



(b) Difference between HF highest occupied molecular orbital (HOMO) and DFT HOMO of methane. Red signifies more density in the HF case and blue more in the DFT case. Isocontour level = 0.00004 $e/\text{Å}^3$

Figure 4.4 – Differences between DFT (B3LYP) and HF calculation. Both were single point calculations with the 6-31G basis set on a geometry optimized at HF level.

Indeed there are other options such as DFT or X-ray constrained wavefunction (XCW) calculations. Within the framework of the latter, not only the energy is minimized but also an additional criterion forcing the electron density to be as close as possible to the one of a real crystal. Therefore, an X-ray constrained wavefunction necessarily has a higher energy than the respective unconstrained wavefunction but this does not mean it is worse. Apart from the fact that it describes the molecule in a crystal and not in vacuo or polarized medium as HF does, it also allows to indirectly include other effects such as for example correlation at HF

Table 4.2 – Some energies (in Hartree) of the methane molecule optimized at HF/6-31G level of theory. The Hartree-Fock energies of the wavefunctions were calculated with *AIMAll* [32] via integration over all atoms leading to small numerical imprecisions.

HF output energy*	-40.18055
HF energy of this WFN	-40.18055
HF sum or orbital energies	-13.78792
DFT output energy**	-40.51033
HF energy of this WFN	-40.17886
DFT sum or orbital energies	-12.02672
XCW output energy***	-40.18048
HF energy of this wavefunction	-40.18048
XCW sum or orbital energies	-13.76745

* HF 6-31G
** B3LYP/6-31G
*** Structure factors for the DFT wavefunction of methane in a hypothetical 10x10x10 $Å^3$ unit cell. Resolution: 2 $Å^{-1}$. XCW level of theory: HF/6-31G, $\lambda = 50$

level of theory.
For a fair comparison between HF and XCW one should not only improve the WFN but also the energy as done for DFT calculations where the energy is a functional of the density and not calculated in the HF way. Also within XCW a simple comparison of energies between constrained and unconstrained wavefunctions is biased since the constrained one is necessarily higher but this wavefunction instead contains other effects. In fact, we have tried to include a measure of energy also for the X-ray constrained part and the results are presented in section 8.1.

4.4.2 Advantages and disadvantages of XCW

While traditional methods for the refinement of X-ray diffraction data, like for example the multipolar refinement, provide only information about the diagonal elements of the one-electron reduced density matrix (the electron density), the XCW method also enables the calculation of the density matrices and the WFN itself. Since wavefunctions contain all the information of a quantum mechanical system in a specific state, this approach is not only providing information about one-electron properties but also about two and more electron properties. Thus, although the fitting is done to structure factors from Bragg scattering, depending only on the electron density, the other elements of the density matrices are obtained as a result of this particular method and thereby also properties which were not used to obtain the WFN can be calculated. The availability of a WFN goes along with further advantages. It allows to express the information about a system in a well known and compact form. A

lot of software exists to deal with WFNs and it is relatively easy to extract various properties. Parameter correlation is much less of a problem than in multipolar refinement due to the energy minimization criterion.

Disadvantages are that it is not so clear what part of the results comes from theory and what from experiment. The physical meaning of XCW derived quantities is questionable since they come only indirectly from the X-ray experiment. The information level between theory and experiment might be different like for example structure factors include correlation and calculation does not. Additionally, there are problems going along with experimental data such as extinction, absorption, or limited resolution.
The weighting of the theory and experiment part has to be user-defined and it is still unclear what criteria to use for that.
Another bias is that in order to calculate a molecular WFN one needs to truncate the crystal. This works well for molecular crystal and is more problematic for polymeric crystal.

Apart from all the issues, a non-periodic WFN built from atomic orbitals representing a crystal is quite a remarkable construction. The correct linear combination of all the X-ray constrained WFNs over all molecules in the crystal gives in principle the correct crystal WFN. Thus, XCWs represent a way to obtain a molecular wavefunction for a periodic system such that, in principle, the properties are those of the crystal.

4.4.3 Procedure for X-ray constrained wavefunction calculations

Nuclear positions and ADPs need to be known from previous refinements (e.g. HAR (section 4.5) or multipolar model (section 4.2)) before the XCW calculation. XCWs are obtained by simultaneously minimizing the electronic energy and an agreement statistics obtained by comparing the calculated structure factors to the measured structure factors (least-squares fitting). [63] The calculations can be done with the software *Tonto*. [126]
The energy is the expectation value of the Hamilton operator applied on the WFN $\Psi(\mathbf{r};\mathbf{p})$ which is determined by the parameters $\mathbf{p}$ and coordinates $\mathbf{r}$.

$$E(\mathbf{p}) = \left\langle \Psi \mid \hat{H} \mid \Psi \right\rangle \tag{4.5}$$

The XCW method has been expanded to Kohn-Sham orbitals, unrestricted calculations and relativistic calculations. [127]

χ^2 is an agreement statistics used to analyse how well the calculated structure factors F_k^{calc} agree with the measured ones F_k^{exp}. The squared difference of these structure factors is weighted with the uncertainty σ_k and summed over every reflection k.

$$\chi^2(\mathbf{p},\mathbf{q};\mathbf{F}^{exp}) = \frac{1}{N_r - N_p} \sum_k^{N_r} \frac{(F_k(\mathbf{p},\mathbf{q}) - F_k^{exp})^2}{\sigma_k^2} \tag{4.6}$$

N_r is the number of experimental data, N_p is the number of parameters, and $\mathbf{q}$ are additional

parameters of the X-ray diffraction experiment which are needed to represent the measured structure factors of a real crystal such as thermal motion (ADPs), extinction and an overall scale factor.

The energy minimization and the least-squares fit of the structure factors are combined in one global functional $L(\mathbf{p},\mathbf{q})$ to be minimized and one part is weighted with a factor λ. By convention, λ is usually put in front of the structure factor part. It is not a dimensionless factor since the unit of the first part is an energy and the second part are squared structure factors so that the unit is squared electrons. In chapter 8.1 it is described how to make the second part also have dimensions of energy.

$$L(\mathbf{p},\mathbf{q}) = E(\mathbf{p}) + \lambda\chi^2(\mathbf{p},\mathbf{q};\mathbf{F}^{exp}) \tag{4.7}$$

In order to decrease χ^2, one has to increase the weight of the experimental part and thus increase λ. This will, however, lead to an increase of the energy because likewise to the transformation from HF to DFT, the electrons contract and get on average closer to the nuclei, thus, increasing their kinetic energy which is always destabilizing (and this dominates over the effect of lowering the stabilizing potential energy). Vice versa, decreasing the the energy, that is lowering λ (up to the point where $\lambda = 0$ and thus an unconstrained HF calculation) leads to a worse fit of χ^2. [128]

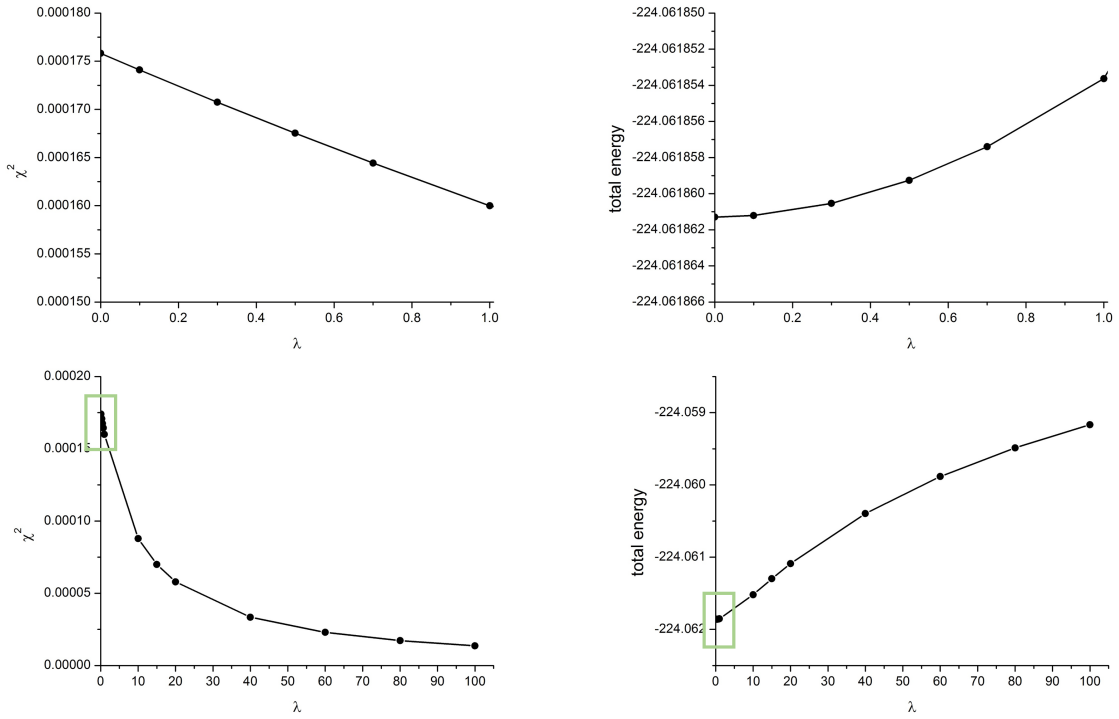


Figure 4.5 – XCW fitting to simulated structure factors of urea. The figures on the top are magnifications of the regions in the green rectangles.

There are two ways to introduce thermal effects. Either one partitions the electron density into atomic contributions (for example with a Hirshfeld atom partitioning) and then convolutes these with the nuclear-position probability distribution function (ADPs) or one calculates an approximate two-nucleus distribution from one-nucleus thermal parameters according to the procedure by Stewart, Coppens, or Tanaka as done in *Tonto*. Eventually, the structure factors are calculated by summation over all atoms in the molecule and by applying the correct crystal symmetry.

An additional parameter Δ can be added to the overall functional, which represents the desired agreement. The optimum value and targeted agreement is $\Delta = 1$ which means that on average the modelled structure factors are within one standard deviation of experimental structure factors.

$$L(\mathbf{p},\mathbf{q}) = E(\mathbf{p}) + \lambda\chi^2(\mathbf{p},\mathbf{q};\mathbf{F}^{exp} - \Delta) \tag{4.8}$$

In case the desired agreement is not reached, λ should be increased and the calculation repeated until χ^2 has reached the desired value. However, as visible in Figure 4.5 at a certain point χ^2 does not decrease much more whereas the energy increases a lot. Therefore, it is not sensible to lower χ^2 by all means. Furthermore, convergence problems arise when λ is too high. Other proposed termination criteria are to increase λ until χ^2 reaches the same weighted residual $R_w(F)$ as in a multipolar refinement of the same system [129], to increase it until the change in χ^2 or E is below a given level, or to increase until $\frac{\Delta\chi^2}{\Delta\lambda} > -0.5$ or $\frac{|E_{el} - E^{el}_{\lambda=0}|}{|E^{el}_{\lambda=0}|} > 5 * 10^{-4}$. [130]

4.5 Hirshfeld atom refinement and X-ray wavefunction refinement

Scattering factors in the IAM approach are based on spherical, atomic electron densities. However, the electron density of an atom depends obviously on its chemical environment and thus, an improved method consists in replacing the spherical atom densities by aspherical ones. Another way is to replace them by tailor-made electron densities for each atom and use these for the structural refinement. This idea led to the development of the HAR. [131] In contrast to the multipolar model, only coordinates and ADPs are refined whereas the electron density comes from theory and is not refined from experimental data. Since no electron density parameters are refined correlation between parameters is smaller.
With X-ray diffraction data one can determine cell parameters, atomic coordinates, and atomic displacement parameters but while the coordinates and ADPs of elements with several electrons can usually be determined quite accurately, hydrogen atoms are an issue. This is due the fact that they only have one electron and therefore are almost invisible to X-rays and furthermore, the density is shifted towards their bonding partners. Yet, in many fields it is important to know their positions. To circumvent the problem, neutron diffraction can be used where the H-positions can be accurately determined. Neutron diffraction is by far not as widespread as X-ray diffraction is because it requires access to large-scale facilities.

Therefore, H-positions are often fixed to average bond lengths from neutron diffraction data [132] and the hydrogen ADPs are determined with SHADE, combining a rigid-body analysis of the molecular backbone and internal vibrations from a database. Pseudoatom databases (see section 4.3) led to improved description of hydrogen-atom positions and thus bond lengths. [119] Due to the use of theoretical electron densities, HAR can further improve the situation and it is possible to obtain good H-positions and reasonable ADPs. [131]

The steps of a Hirshfeld atom refinement according Capelli *et al.* [131] are:
- An initial set of atomic coordinates is required from a previous refinement, for example an IAM or MM refinement.
- A wavefunction is calculated for a molecule which is as large as the asymmetric unit or larger, usually at HF or DFT level of theory. From this wavefunction an initial electron density is obtained.
- The electron density is partitioned into atoms with Hirshfeld's stockholder partitioning scheme. [133], where the electron density of an atom A is obtained with its weighting function w_A.

$$\rho_A(\mathbf{r}) = w_A(\mathbf{r})\rho(\mathbf{r}) \tag{4.9}$$

The weighting function is obtained by dividing the spherically averaged atomic electron density by the sum of spherical electron densities of all atoms in the molecule.

$$w_A(\mathbf{r}) = \frac{\rho_A^0(\mathbf{r}-\mathbf{R}_A)}{\sum_B^{molecule}\rho_B^0(\mathbf{r}-\mathbf{R}_B)} \tag{4.10}$$

- Charges and dipole moments of the atoms are calculated and used to simulate a crystal field around the molecule by placing them at the positions of atoms of adjacent molecules.
- In this field a new electron density is calculated and this is called an ED cycle.
- ED cycles are repeated to convergence in the molecular energy which is then called an ED step.
- The Fourier transforms of the Hirshfeld atoms are used to calculate the aspherical scattering factors of every atoms. Subsequently, these are used for a least-squares refinement of coordinates and ADPs. This is called a structural refinement step.
- ED steps and structural refinement steps are repeated alternately until the shift of parameters falls below a given level.

In the initial method of Jayatilaka and Dittrich from 2008 [134] the initially calculated Hirshfeld atoms were kept fixed and not updated during the procedure whereas in the more recent procedure presented above cycles of electron density and non-spherical atomic scattering factor calculations followed by coordinate and ADP refinements are repeated to convergence.

In a study with L-phenylalaninium hydrogen maleate, Woinska *et al.* [135] compared hydrogen bond lengths obtained with IAM, MM, HAR (anisotropic ADP, isotropic ADP and ADP from SHADE), transferable aspherical atom modeling TAAM (database) and neutron diffraction. It was shown that the hydrogen bond lengths obtained with the anisotropic Hirshfeld atom

refinement are in very good agreement with the ones from neutron diffraction. In contrast, bond lengths to hydrogen atoms have to be fixed to neutron distances in the multipole models and TAAM to derive sensible properties. ADPs of hydrogen can be refined with HAR but the resulting shapes are not in such a good agreement with neutron data as the positions are.

The XCW method described in section 4.4 is a way to obtain a wavefunction from experimental structure factors. In its original setting it is not foreseen that it is used for a refinement of atomic coordinates and thermal vibration parameters. However, in combination with HAR it can be used to also refine coordinates and ADPs and this method is named X-ray wavefunction refinement (XWR). It works very similar to the HAR just that the calculation of the electron density is done with XCWs so that this step also includes experimental info. Thus, it is an iterative process including an X-ray WFN refinement adjusting the electronic info agaist experimental data followed by a least-squares refinement. The method was first introduced by Grabowsky *et al.* [136]. They studied the bond in SO_2 and an SO_2-containing molecule and used the resulting wavefunction to analyse its character. They showed that ionic bonding rather than multipole covalent bonds are present in these cases. Chęcińska *et al.* [137] compared the structures of three urea derivates obtained from MM, HAR and XWR. XWR gave the best results.

4.6 Other scattering techniques than X-ray diffraction

4.6.1 Compton scattering

In section 1.2 the one-electron reduced density matrix was introduced. It is:

$$\gamma(\mathbf{x};\mathbf{x}') = N \int \Psi_0(\mathbf{x},\mathbf{x}_2,...,\mathbf{x}_N) \times \left(\Psi_0(\mathbf{x}',\mathbf{x}_2,...,\mathbf{x}_N)\right)^* dx_2,...dx_N \tag{4.11}$$

The 1e-RDM describes the position and spin distribution of the electrons. By integrating over both spin components the position 1e-RDM $P(\mathbf{r};\mathbf{r}')$ is obtained. Fourier transformation over all six coordinates leads to the momentum 1e-RDM $\bar{P}(\mathbf{p};\mathbf{p}')$ [138]. The diagonal elements of the 1e-RDM in position space correspond to the electron density while in momentum space the diagonal corresponds to the electron momentum density. Both of them are observables; the electron density from X-ray diffraction and the momentum density from Compton scattering. X-ray diffraction is the elastic scattering of X-rays and Compton scattering is the inelastic scattering of photons on the electrons. As a result of the inelastic scattering, the energy of the incident photon changes and the change depends on the momentum of the electrons along the scattering direction p_z as well as on the scattering angle. Plots showing the scattering with respect to the electron momentum distribution are called Compton profiles $J(p_z)$. [139] Compton scattering also has a very small spin-dependent contribution which can be studied by polarizing the incident photons such that a magnetic Compton profile is obtained. [139]

4.6.2 Neutron diffraction

In contrast to X-rays which are scattered by electrons, neutrons are mainly scattered by the core of the atoms. This goes along with a number of important implications: First of all, the scattering power of the atoms is not anymore directly dependent on the number of electrons. With X-rays light elements scatter less than heavy elements which is the reason for why the position of hydrogen atoms cannot be well determined. The neutron scattering power instead varies arbitrarily among the periodic system (at least from a chemical perspective, there are physical reasons for it) and therefore hydrogen atoms are well visible and their positions and ADPs can be accurately determined. Second, neutron scattering depends on the isotope meaning that one can distinguish between different isotopes in a sample (e.g. determine the hydrogen to deuterium ratio).
It is possible to do refinements using X-rays and neutrons at the same time (X-N method) and thus to refine accurate hydrogen positions as well as the electron density. [140]
Neutrons penetrate samples much better than X-rays and if single crystal samples are used, their size must be in the order of 1 mm^3 [141] which is much larger than for X-ray diffraction. Furthermore, neutron sources are considerably less common than X-rays sources which can be produced in any lab with home-source diffractometers. In order to measure with neutrons one has to go to large scale facilities such as spallation sources or nuclear reactors.

4.6.3 Polarized neutron diffraction

Neutrons possess a spin, which is employed in polarized neutron diffraction. Their magnetic moment interacts with the magnetic moment of the electrons and can be used to characterize magnetic crystalline materials. The scattering depends on the angle between the scattering vector and the spatial distribution of the magnetization and, moreover, the neutrons interact with both the nuclei and the electrons making this technique quite complicated. In practice, the polarization of neutrons is flipped and the respective diffracted intensities are measured giving rise to so-called flipping ratios.

With the measured data one can refine the spin density using magnetic form factors analogously to the electron density. From these results one can learn about the electronic structure of open shell systems and magnetic interaction mechanisms. Polarized neutron diffraction results can be coupled with the ones from X-ray diffraction. Such a joint refinement gives rise to much better spin-resolved electron densities than when just polarized neutron diffraction is used. [108]

4.7 Quantum crystallography from other measurements than diffraction

Crystallographic information can not only be obtained from scattering experiments but from any other measurement involving crystals. Thus, quantum crystallography is not limited to

the combination of scattering techniques with theoretical quantum chemistry.
Indeed, the conductivity measurements we performed on the polyiodide crystal (see chapter 7) provide very useful information about the bonding of the different polyiodide units, which cannot be obtained by diffraction. The interaction between them cannot only be understood in terms of atom-atom distances because structurally the I_2 fragments continuously approach I_3^- while the electrical transport properties change rather suddenly. Even the electron density and its properties at different interaction distances do not provide this information; they indeed corroborate the measured increase of the conductivity by predicting a more and more covalent bond but they do not forecast any sudden change as well. [142]
Deplano *et al.* [143] demonstrate in their article that Raman spectroscopy can provide information about the bonding in polyiodides, which is inaccessible from structural information. X-ray diffraction measures the averaged (in time and space) electron density whereas Raman spectroscopy provides information about the movement of two atoms with respect to each other and the correlation of this movement. In general, many spectroscopic techniques are used to study crystals and they can provide important information inaccessible from diffraction.

Different information can be obtained from the measurement of the dielectric constant of MOFs. MOFs are crystals with pores, which can be filled with solvent. In a study of Scatena *et al.* [144] the uptake of water molecules in a MOF was studied where the molecules can be physisorbed, chemisorbed, or are freely moveable in the pores. Depending on the content of water in the MOF and the way the water molecules are connected to the framework, the dielectric constant changes. Since many water molecules are only filling the voids without being structurally visible, this information cannot be obtained from diffraction measurements. Thus, quantum crystallography is by far not limited to diffraction experiments.

Results Part III

5 Outline and Goals

In chapter 2, theoretical chemistry tools to partition various chemical quantities were presented. One is, for example, the Quantum Theory of Atoms in Molecules (QTAIM), with which an unambiguous definition of atoms in molecules is enabled. Since we are interested in crystalline systems, our studies are particularly guided by the identification of atoms, functional groups, and molecules in crystals. This is not merely an issue of nomenclature but serves to rationalize the properties of materials in terms of their building blocks and interactions among them. Furthermore, these building blocks often have remarkably invariant properties. [145, 146] The four chapters in the results part of this dissertation make use of this concept and apply it to different problems. Chapters 6 and 7 are published articles [147, 142] while chapter 9 is an article submitted for publication. These three chapters are reported in the dissertation without any modification, but supplemented with additional results and comments at the end. My contribution to these articles is explained in the last paragraph of this outline. Chapter 8 summarizes different results obtained in the field of X-ray constrained wavefunctions.

In chemistry, nuclei are always well separated from each other, while this is not the case for the electron density. Due to the interactions between atoms, the electron density is shared between them and spread over the molecule; thus, it cannot be easily attributed to atoms. The QTAIM approach, presented in section 2.1, is a way of uniquely ascribing all the electron density in the molecular space to atoms. The identification of atoms is useful because one can thereby assign properties to them. This, in turn, offers the opportunity to understand the behaviour of given compounds in terms of their building blocks and, therefore, also to rationally design new materials.
Chapter 6 demonstrates how to obtain polarizabilities of atoms and functional groups based on QTAIM. The polarizability is an important characteristic that defines material properties such as the dielectric constant or the refractive index. Furthermore, it is not only used to obtain macroscopic optical properties but, by partitioning these properties into contributions of building blocks, it additionally serves to analyse the interactions between the building blocks themselves. Polarizabilities of functional groups in various molecules were calculated, compared and classified into groups. For every group the mean polarizability tensor was

evaluated and served as representative database entry for that particular group. Subsequently, this average polarizability was transferred to other functional groups, which therefore did not need to be calculated *ab-initio*. The availability of a database allows for a quick estimation of functional group polarizabilities, which otherwise would have to be laboriously computed or could not be computed at all. In our article we demonstrated that the polarizabilities obtained from the database are good estimates, such that the refractive index derived thereof is not further away from the *ab-initio* calculated value than the precision of a measurement is.

Another possible partitioning is that of the interaction energy presented in section 2.2. Several techniques exist to partition the energy of two interacting atoms or fragments into different contributions such as ionic or covalent. Although there is not a clearly best way of doing that and the results vary depending on the chosen method, such partitionings are more than just nomenclature. They allow us to analyse the characteristics of a bond and to compare them with those of other bonds. Absolute numbers of partitioned bond energies are to be taken with care, but relative numbers compared for different situations can bear a lot of information. This was done in the article about polyiodides (chapter 7), where we initially examined one particular iodine-iodine contact in a crystal subjected to increasing external pressures. The studied crystal consists of well separated I_2 and I_3^- molecules at ambient pressure. Upon compression these units approach each other and form bonds, until the crystal eventually polymerizes. This process appears continuous and is made possible by the nature of iodine atoms able to form hypervalent structures. The goal was to find out whether the I_2 and I_3^- units can be considered separated molecules in the crystal and, if yes, at what compression level this is not any longer the case. In order to answer this question, several different partitioning methods of the bond energy were employed. Indeed, we found that when I_2 and I_3^- approach each other due to the external pressure, the covalent character of the interaction increases. This change is not just a theoretical feature but has a tremendous effect on the macroscopic properties. We measured the conductivity of the crystal and found a dramatical increase with pressure and thus with the transformation of electrostatic bonds to covalent ones. A second example showed that also the number of possible polyiodide pathways present in the crystal has an important effect on the conductivity.

The study presented in chapter 9 started with the compound boron-nitrilotriacetate (NTA-B), which has three carbonyl groups. In the crystalline state they interact with the carbonyl groups of neighbouring molecules. The arrangement in the crystal brought up the idea that at high pressure, the carbonyl groups could form bonds and the crystal would polymerize. Thus, again this study is guided by the question of how molecules are identified in crystals and when their identification becomes blurry. We used X-ray constrained wavefunction (XCW) calculations to study NTA-B and other compounds. On one hand, we did conventional XCW calculations explained in chapter 4.4 and on the other hand XCW calculations with extremely localized molecular orbitals (XC-ELMO). We compared the electron density and the orbitals including and excluding the effect of the crystal and rationalized the observed differences in terms of interactions.

Chapter 8 deals with various different aspects of X-ray constrained wavefunctions. First, the XCW equation was changed in a way that both terms, the quantum chemical minimization of the energy and the fitting to the crystal structure factors, have units of energy and are therefore on the same scale. Second, the electronic energy calculated from structure factors was examined. Third, the effect of an electric field on the structure factors and thereof derived electron densities was studied.

To sum up, the common point of all chapters is the identification of atoms in molecules and molecules in crystals in order to study their interactions and contributions to the behaviour of the total system. Various different kinds of quantum chemical calculations were performed, either for molecules *in vacuo* or crystals. Through the use of various quantum chemical and crystallographic methods and the combination of them to gain additional insight, this work can rightly be put in the context of quantum crystallography, and the topics studied represent some questions and answers in the broad spectrum of that research field.

Specification of personal contributions to the published articles
The results presented in the published book chapter constituting chapter 6 of this dissertation are based on my calculations. Furthermore, I wrote the code for the genetic algorithm clustering, as well as the code for the definition of a local coordinate system in specific functional groups and the transformation of polarizabilities into this coordinate system.
For the article about the polyiodides, Tomasz Poręba performed the experimental measurements while I did the calculations.
Within the scope of the XC-ELMO project I stayed one month at the University of Lorraine in Metz to work with Alessandro Genoni, who introduced me to the XC-ELMO calculations. The calculations in the article were carried out by me while the high-pressure experiments of NTA-B were done by Fabio Montisci.

6 Towards a Generalized Database of Atomic Polarizabilities

Apart from the addendum, this chapter originates from the previously published book chapter "Towards a Generalized Database of Atomic Polarizabilities" in *Understanding Intermolecular Interactions in the Solid State*, **2018**. [147]
(https://pubs.rsc.org/en/content/chapter/bk9781788010795-00211/978-1-78801-079-5)
Reproduced by permission of The Royal Society of Chemistry

Abstract: *This chapter deals with the transferability of functional group polarizabilities. The motivation of this study is the need for a rapid evaluation of dielectric properties of materials consisting of aggregation of molecules (nanocrystals, fibres, films and crystals), with the purpose of a rational design and an efficient screening. Criteria based on the quantum theory of atoms in molecules define functional groups and provide a good transferability of their properties. This is true not only for properties depending solely on the ground state electron density, but also for those defined by mixtures of states, like the polarizability, which is the main subject of this study. In order to allow for the transferability, we use a generalized definition of a local coordinate system for each functional group, necessary to express the polarizability tensors in a fully self-contained way without biases. Transferability tests show that the best local coordinate system is based on the nuclear charge moment tensor of the functional groups. The a priori definition of the functional groups was further refined by clustering their polarizabilities, using genetic algorithm according to some similarity criteria, which leads to an even better estimation of properties for molecules that are not used to build the original database. This procedure allows to rapidly estimate polarizabilities of large molecules or, together with the dipole moment, to approximate the crystal field. Applications of this approach concern, among others, the calculations of linear optical properties of crystals.*

6.1 Introduction

Present-day quantum chemical methods allow for precise calculations of molecular wave-functions and, consequently, electronic properties of molecules. The same holds true for

crystals, for which periodic wavefunctions, atomic structures as well as electronic structures and properties can be quite accurately calculated. In this context, it could be expected that the interest in sub-molecular units, like functional groups, is small because the most obvious break-down of a material may be molecules (for molecular materials), monomers (for polymers) or atoms (for extended covalent, metallic and single-atom ionic solids). However, when a materials scientist or a chemist aims at designing a new functionalized material, the basic constituents of molecules and crystals are still considered. As a matter of fact, the concept of functional group is so deeply entangled in modern chemistry that it cannot easily be avoided. The characteristic behaviour of two or more functional groups justifies for instance the use of the arrow notation in synthetic chemistry, it allows to explain reaction mechanisms and design new synthetic routes. Functional groups are also important for the characterization of a material. For example, typical vibrations of atoms in a group give rise to absorption bands that allow their identification in an IR spectrum. Finally, the optical, electric, magnetic or mechanical properties of molecular materials can be analysed in terms of the contributions of the functional groups they consist of; ultimately, just one or a few functional groups might be enough to explain a molecular or crystal property, an important aspect for retro-designing new materials (reverse crystal engineering). The fact that functional groups approximately retain their identity and properties in different molecules goes along with the concept of *transferability* [148, 149, 150] according to which the electron density and thereof derived properties depend, to a first approximation, on the local nuclear configuration and the bonding situation. Nevertheless, the properties of a functional group depend obviously also on the chemical environment, which perturbs its "isolated" electron density. These perturbations are of two kinds: through-bond and through-space. The first one is due to conjugative effects that may enable outer atoms to modify the electron distribution of a given functional group, even if they are quite distant. Of course, this perturbation decreases with the number of bonds that separate the perturbing atom and the functional group under analysis. The through-space perturbation, instead, is due to the polarization induced by the electric field of the perturbing group and it rapidly decreases with the distance. Thus, despite an overall similarity, if we compare a given property of a specific functional group in different molecules, a relatively large variance may be observed. The variance is smaller for functional groups of the same molecule embedded in different crystalline environments (for example in co-crystals or polymorphs). In fact, the electric field due to the surrounding molecules in a crystal, despite being of the order of magnitude of GV/m, is much smaller than an intra-molecular field. Moreover, through-bond effects are almost absent in intermolecular interactions, apart from very strong hydrogen bonds. Therefore, changing some substituents in a molecule is much more perturbative than changing the surrounding of a molecule.

For these reasons, to use functional groups in chemistry and materials science, the concept of *similarities* and *dissimilarities* must be properly quantified and it is quite useful to establish a classification that allows to sort them into groups (clusters) within which some *descriptors*, for example physical properties, are similar. This means that elements of a given cluster can be well represented by the average values of their descriptors, whereas functional groups

belonging to different clusters are clearly distinguished. After this delicate procedure, one can build up a database of functional groups and their associated properties, to easily evaluate the role of a group of atoms in a molecular crystal. The cluster averages not only represent the molecules used to build up the database, but they are also representative (within a given confidence criterion) of the same functional groups in other molecules not employed for constructing the database. If this purpose is reached, the database entries would allow to rapidly estimate properties of large molecules or their aggregation in crystals, without the need of extensive and time-consuming calculations. Furthermore, the role of individual atoms or functional groups to the build-up of a molecular or crystal property can be evaluated, an important step towards the rational design of molecular-based materials.

Even though our approach is general, we will focus on the first-order polarizability tensor in this work. The electric polarizability is responsible for the optoelectronic properties of a material like refractive index or dielectric constant. Thus, a good comprehension of the key factors that affect the polarizability allows one to tune the optoelectronic properties and design new materials with targeted applications. For example, high refractive index (HRI) materials are desirable for lenses, optical waveguides [151], or a high light extraction efficiency in LED components [152]. Anti-reflective coatings used in solar cells require a tuneable refractive index [153]. The ongoing downscaling of semiconducting devices demands for materials with continuously lower dielectric constants for the insulation of wires [154]. Moreover, fast market changes and the objective to produce more sophisticated opto-electronical devices stimulate the development of more environment-friendly HRI materials. So far, aromatic rings, heteroatoms like sulphur, or halogen atoms were used to enhance refractive indices. [155, 156, 157, 158] However, due to environmental concerns, compounds with halogen substituents are not suitable. Moreover, a big challenge is to increase the refractive index of a material while keeping its low viscosity at ambient temperature. At present, a wide group of halogen-free acrylate derivatives is used for fabricating HRI optical devices but so far no organic compound with low birefringence, small optical dispersion and absence of visible spectra absorption exhibited a refractive index above 1.7. [159, 160, 161] Inorganic compounds may exhibit higher refractive indices but they carry a number of disadvantages such as high density, low flexibility and highly intense visible absorption bands. [162] Therefore, there is still a demand for new and improved organic compounds with excellent optical properties.

In this context, the purpose of our research is to simplify the efforts necessary to select new promising materials (either organic, inorganic or metal-organic), before undergoing expensive tests and more sophisticated calculations, and before embarking new and perhaps expensive fabrication procedures. For this purpose, a critical issue is being able to rapidly estimate an approximate value of the refractive indices of a molecular-based crystalline material in order to select the most promising candidates.

With modern computational resources, the calculation of static as well as frequency dependent polarizabilities of small to medium sized molecules is possible using quantum-mechanical methods at many levels of theory thus allowing for acceptable precision. However, given the

relatively high level of electronic correlation usually necessary for very accurate estimations of the optoelectronic properties, ab-initio calculations for large molecules and condensed matter are often prohibitive. [157] Noteworthy, a high correlation level is necessary to well represent local and non-local effects perturbing a functional group. This suggests that a partition of the material in functional groups may be preferable to compute the overall polarizability of the system. In fact, a database of functional groups may simultaneously guarantee the requested accuracy (depending on the level of theory used to pre-calculate the functional groups) and a very high rapidity (because new *ab initio* calculation is not necessary and the polarizability of the whole system is simply the sum of the polarizabilities of their constituent parts). The only concern is a proper spatial orientation of the database polarizability tensors, which can be solved with a proper treatment of local tensor axes, as deeply discussed in this chapter.

Once a database of functional group polarizabilities is compiled and at hand, optical properties of molecular materials can be rapidly predicted. If the database also contains dipole moments of the functional groups, the inner crystal field can be approximated and therefore the dielectric properties of a crystalline material as well, by evaluating also the mutual polarization of the molecules.

The road-map of our research is summarized in Figure 6.1, focusing on the build-up of a polarizability database, which is the subject of this chapter.

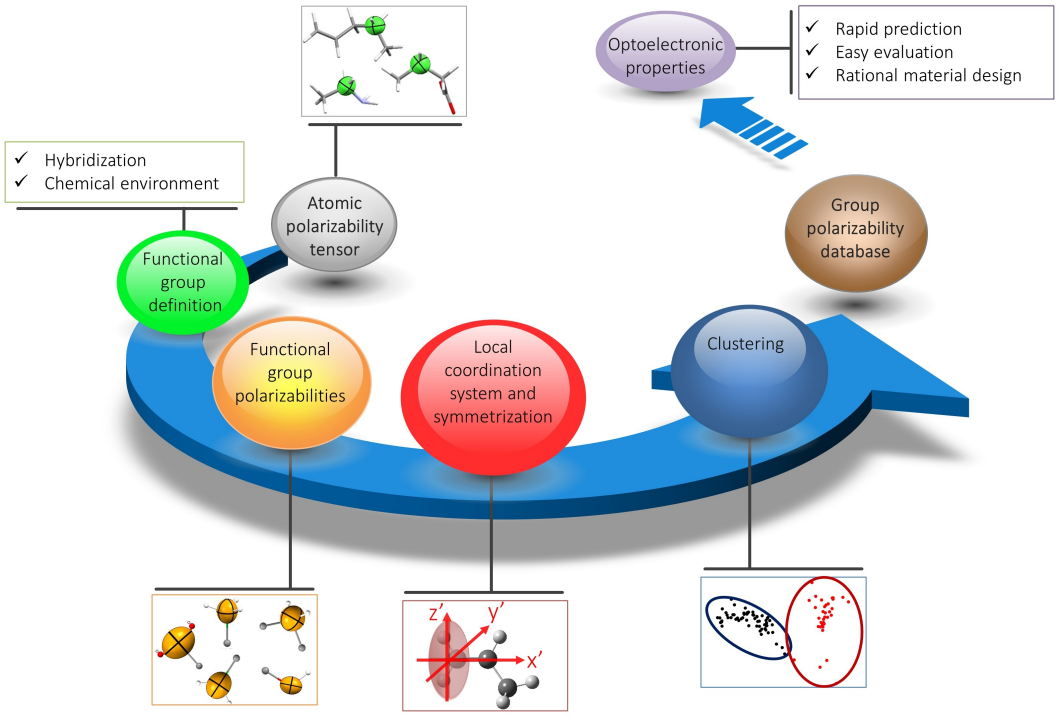


Figure 6.1 – A step-by-step pathway for creating a generalized database of functional group polarizabilities.

6.2 Theoretical Background

6.2.1 Earlier Atomic Polarizability Databases and the Need for a New One

A molecule exposed to an electric field will respond to this perturbation by a polarization of its charge density (affecting both the electron and nuclear distributions). The strength and type of this response is described by the electric (hyper)polarizabilities, which are symmetric tensors relating the three-dimensional electric field vector $\boldsymbol{E}$ to the change in the dipole moment $\boldsymbol{\mu}$. If only the dipolar first order polarizability is considered, the molecular dipole moment will change linearly with the applied field and the induced dipole moment change is easily derived: [163]

$$\Delta\boldsymbol{\mu} = \boldsymbol{\alpha} \cdot \boldsymbol{E} \tag{6.1}$$

The first order polarizability $\boldsymbol{\alpha}$ is a 3 x 3 tensor whose individual components can hardly be determined experimentally. Alternatively, it is common to refer to the isotropic polarizability (the arithmetic average of the main diagonal tensor components), easily obtained from refractive measurements in solution or polycrystalline materials. However, knowing the individual tensor components would be essential as they inform on the orientation of the polarizability ellipsoid, which may be highly anisotropic. This is of relevance for rationally designing crystalline materials, because the anisotropy of the molecular polarizability may explain the most important intra- and inter-molecular contacts that cause the anisotropy of the optoelectronic properties.

Vogel [164] was one of the first to evaluate and produce a database of isotropic polarizabilities for different atom and functional group types using additivity schemes. In this kind of approach, the group constituents of a chemical system are considered to not interact with each other, hence the "unperturbed" polarizabilities are simply summed together to produce the whole property. Based on Vogel's isotropic values, Applequist [165] set up a more advanced way to evaluate polarizabilities of molecules via a non-additive model taking into account the interaction of atomic polarizabilities via induced dipole moments. Therefore, the molecular polarizabilities, despite based on isotropic atomic values, were transformed into anisotropic ones. Miller [166] used the so-called atomic hybrid components (AHC), which were defined by the atomic number and the hybridization state of the atom, to construct isotropic molecular polarizabilities. Every AHC had an assigned polarizability which was obtained by fitting to experimental data. Kassimi et al. [167] calculated isotropic polarizabilities of amino acids as the sum of smaller fragments they consist of. The fragments had to be saturated with hydrogen atoms and this contribution was afterwards removed from the summation:

$$\boldsymbol{\alpha}(AB) = \boldsymbol{\alpha}(AH) + \boldsymbol{\alpha}(BH) - 2\boldsymbol{\alpha}(H) \tag{6.2}$$

The reference polarizability for hydrogen was obtained from the polarizabilities of methane

and ethane in the following way:

$$\boldsymbol{\alpha}(H) = \boldsymbol{\alpha}(CH_4) + \boldsymbol{\alpha} - \frac{1}{2}\boldsymbol{\alpha}(C_2H_6) \tag{6.3}$$

The polarizabilities of the small fragments AH and BH can be calculated with much less effort or they can also come from a previously built database. Therefore, the authors pointed out the need for a database of such fragment polarizabilities.

Dykstra et al. [168] developed a database for different atom types including also anisotropic polarizabilities obtained by fitting their tensor components to the corresponding calculated molecular components. Unfortunately, no definition was given for the local coordinate system used to import the anisotropic tensors from the database, an issue which is not trivial to solve (as we will discuss in section 6.3.2).

Alternatively, Stone [169] calculated distributed polarizabilities from the Hilbert-space based distributed multipole analysis previously developed by himself and co-workers.[170, 171, 172] After analysing many partition schemes, Stone advocated that the most transferable polarizabilities would be obtained from a real-space based partitioning, like the ones proposed by Bader or Hirshfeld. Ivanov-Smolenskii et al. [173] obtained polarizabilities of ions based on X-ray diffraction data. They examined crystals of the NaCl type where the Wigner-Seitz cell contains only one atom and the separation into what they defined as pseudoatoms is very similar to Bader's scheme described in section 6.2.2. From the number of electrons and their mean radius they estimated atomic polarizabilities. Whitten et al. [129], instead, pointed out that the one-electron density, as for example obtained from a multipolar refinement, is not sufficient for determining polarizabilities. They proposed the X-ray constrained wavefunction approach to evaluate polarizabilities from X-ray measurements. Suponitsky et al. [174] calculated the intermolecular interaction energy of a molecule in a crystal as the energy difference of this molecule embedded in the crystal and in isolation. The induced dipole moment, necessary to calculate the electrostatic part of the energy, was derived from the difference of the electron density of the molecule in the crystal and in isolation. The electron density was described with a Hansen-Coppens multipolar model fitted on one hand to X-ray data and on the other hand to Hartree-Fock calculations. They concluded that the uncertainty of the electron density obtained from a measurement, in their example of a non-centrosymmetric crystal of urea, is too large to allow this calculation while it worked well for electron densities obtained from Hartree-Fock calculations.

Bader et al. [149] demonstrated the additivity of group polarizabilities for acyclic hydrocarbons and rationalized it based on the partitioning of the electron density with the Quantum Theory of Atoms in Molecules (QTAIM). From the polarizabilities of alkanes, Bader derived average polarizabilities for the methyl and methylene group. The QTAIM approach was further developed by Keith who mainly contributed to the solution of the origin dependence problem of dipole moment derivatives of non-neutral subunits of a molecule (like in fact atoms). This scheme

was improved and implemented in the software *PolaBer*, which allows the calculation of full polarizability tensors for atoms and functional groups in isolated molecules or crystals. [175] Using this method, Dos Santos et al. [176] calculated average polarizabilities of functional groups present in naturally occurring amino acids and evaluated the changes with respect to their aggregates, thus explaining the role of hydrogen bonds on perturbing isolated-molecule polarizabilities. On one hand, this work paved the way for the database discussed here. On the other hand, that work also provided accurate analysis of the effect of intra- and intermolecular interactions on the transferable functional groups, thus showing how to use the database for accurate determination of polarizabilities of large molecular aggregations, and eventually also crystals.

In all above studies, functional groups were defined *a priori* based on chemical intuition. Except for the QTAIM based studies, the polarizabilities of these atom types, bonds, or groups were always obtained by fitting them to a set of experimental or calculated molecular polarizabilities. As anticipated, our goal is to create a first principle and generally transferable database for functional group polarizabilities, obtained by a partitioning of the molecular electron density according to QTAIM (see section 6.2.2). This method is self-consistent and does not require any assumption. Moreover, also the recognition of functional groups should emerge from the statistical analysis of calculated polarizabilities on a number of representative molecules without assuming any *a priori* knowledge and adopting only unsupervised clustering algorithms. [177] Using a multivariate cluster analysis, one may in fact fulfil this condition. In addition, since the number of clusters may exceed that of functional groups, also subtypes of functional groups may be distinguished. Within one cluster, the polarizabilities are sufficiently similar and the average values of a particular cluster is the best representation of its constituent polarizabilities. Ideally, the cluster coincides with a functional group or a sub-type functional group.

Eventually, a database can be established where each entry contains information about the functional group type and a 3x3 (symmetric) polarizability tensor.

The following approximations are adopted: a) the nuclear relaxations under the applied field and the vibrational contributions to the polarizability are ignored; b) only static (*i.e.* zero-frequency) polarizabilities are considered. By applying a), the electric (hyper)polarizabilities are only determined by the electron density. The use of only the first order polarizability $\boldsymbol{\alpha}$ (a 3 x 3 symmetric tensor) is justified by the *quasi*-linear response of the dipole moments to electric fields of the magnitudes of those of most interacting radiations. Of course, this implies that non-linear effects (like for example second-harmonic generation) cannot be evaluated with the database. Even in the strong inner fields of molecular crystals the range of the response is in general sufficiently close to linearity, therefore the field induced dipole moments and inner fields are also accurately estimated. The use of static fields is justified by the relative similarity between high frequency electric polarizabilities and static polarizabilities. In fact, at very high frequency, all the motions of the molecules and specific strong absorptions are quenched. A simple scaling could be in case adopted to estimate the refractive index in the

desired wavelength spectrum region. [178]

As described above, many scientists have proposed methods to estimate atomic $\boldsymbol{\alpha}$ tensors, and constructed similar databases, either empirically or semi-empirically, but lacking exportable local coordinate systems for the tensors. A major problem was also the definition of polarizabilities that avoid issues related to the origin-dependence of the framework. The novelty of our approach is to use the transferable Bader's zero-flux-bounded functional groups, as previously anticipated [175, 176] to create a database that allows for easy transferability of properties, hence contributing to the rapid estimation of optoelectronic properties of rather complex materials such as proteins, polymers or crystals.

This kind of analysis is not restricted to the polarizability, but could be extended to other properties, like for example the atomic multipole moments. Good estimations of electric moments (in particular the dipole moment) are important to evaluate the electrostatic potential 'produced' by a molecule. As a matter of fact, in the field of charge density analysis, many databases of atomic electric multipoles have been developed in the past. [124, 121, 179] Together with the polarizability, they allow to approximate the crystal field, which polarizes molecules in a crystal.

6.2.2 Distributed Atomic Polarizabilities

In order to derive atomic polarizabilities, it is necessary to define an atomic polarizability density. In the approach proposed here, we use the quantum theory of atoms in molecules, that provides a topological partition of the molecular electron density into atomic basins Ω, based on the gradient field. This enables defining atomic properties, like the atomic dipole moments $\boldsymbol{\mu}(\Omega)$, consisting of two components, namely the atomic polarization $\boldsymbol{\mu}_p(\Omega)$ (i.e., the charge shift within the atomic basin) and the charge translation $\boldsymbol{\mu}_C(\Omega)$ (i.e. the shift from one basin into another).

$$\boldsymbol{\mu}(\Omega) = \boldsymbol{\mu}_p(\Omega) + \boldsymbol{\mu}_C(\Omega) = -\int_\Omega (\boldsymbol{r} - \boldsymbol{R}_\Omega)\rho(\boldsymbol{r})d\boldsymbol{r} + (\boldsymbol{R}_\Omega - \boldsymbol{R}_0)q(\Omega) \tag{6.4}$$

In equation (4), $\boldsymbol{R}_\Omega$ is the position vector of the nucleus Ω; $\boldsymbol{R}_0$ is the origin of the molecular coordinate system (arbitrarily chosen). If the atomic basin is not neutral, $\boldsymbol{\mu}_C(\Omega)$ does not vanish and is origin dependent. This is a serious issue for transferability. However, the procedure designed by Keith enables overcoming this dependency by re-writing the charge-translation component into a basin-to-basin summation:

$$\boldsymbol{\mu}(\Omega) = \boldsymbol{\mu}_p(\Omega) + \boldsymbol{\mu}_C(\Omega) = -\int_\Omega (\boldsymbol{r} - \boldsymbol{R}_\Omega)\rho(\boldsymbol{r})d\boldsymbol{r} + \sum (\boldsymbol{R}_\Omega - \boldsymbol{R}_{BCP}(\Omega|\Omega'))Q(\Omega|\Omega') \tag{6.5}$$

$Q(\Omega|\Omega')$ is the charge transferred to the atomic basin Ω from the bonded atom Ω' and $\mathbf{R}_{BCP}$ is the positional vector of the bond critical point between atoms Ω and Ω'. The scheme

proposed by Keith has a general validity beyond QTAIM itself. A different partitioning scheme may not only produce a different intrinsic polarization $\boldsymbol{\mu}_C(\Omega)$, but also a different connectivity between atomic basins, modifying quite substantially $\boldsymbol{\mu}_C(\Omega)$. Nevertheless, the sum of all atomic dipoles $\boldsymbol{\mu}_C(\Omega)$ and $\boldsymbol{\mu}_P(\Omega)$ must return the molecular dipole. A special case is that of non-neutral molecules. In this case, the molecular dipole is origin dependent. However, for any given partitioning, the procedure by Keith provides an origin independent solution (this basically transform the origin dependence of the $\boldsymbol{\mu}_C(\Omega)$ terms into a "partition" dependence). This aspect is very important when considering transferability of functional group properties, given that a functional group in a molecule is not supposed to be neutral.

The polarizabilities are eventually obtained as the field derivatives of the total atomic dipoles ($\boldsymbol{\mu}_C(\Omega)$+ $\boldsymbol{\mu}_p(\Omega)$), which could be computed analytically or numerically. Thus, atomic polarizabilities can be derived from the unperturbed and perturbed atomic dipole moments by numerical differentiation, provided that the electric field is sufficiently small to cause only a linear response:

$$\alpha(\Omega)_{ij} = \lim_{E \to 0} \frac{\mu_i^{E_j}(\Omega) - \mu_i^0(\Omega)}{E_j} \tag{6.6}$$

The resulting atomic polarizability is a 3x3 tensor, which is by its own definition not symmetrical, but for practical purposes it is symmetrized according to the classical scheme by Nye.[180] Otero et al.[181] have criticized the use of equation (6) arguing that because the unperturbed, field-free electron density and the densities perturbed by the electric fields are independently calculated and partitioned, the atomic basins are not constrained to be constant in the process. According to the authors, this would be the ultimate reason as to why the polarizability tensors result not symmetrical. However, this criticism is not correct, as the numerical approach of equation (6) gives the exact definition of a dipolar density derivative, which corresponds to the exact analytical definition in the limit of an infinitely small field. The reason why the tensors are not symmetrical, lies in the atomic basins themselves, which are in fact by definition not symmetrical, which does not enable equation (6) to be invariant under the exchange of *i* and *J*. This affects both $\boldsymbol{\mu}_p(\Omega)$ and $\boldsymbol{\mu}_C(\Omega)$) terms.

As QTAIM is an exact partitioning scheme, in the sense that all points in molecular or crystal space are assigned to one and only one atomic basin (except for the boundaries), the sum of all atomic polarizabilities in a functional group gives the polarizability of the functional group itself, and likewise for the molecule. The atomic, functional group and molecular polarizability tensors can be represented in position space as ellipsoids, centred at the centre of the group/molecule or at the atomic nuclei, with units of a volume.[175, 176, 182, 27] It is also important to emphasize that, while by definition the total molecular polarizability must be a positive definite symmetric tensor, the individual atomic polarizability must not. However, this is very rarely observed and only for polarizabilities of hydrogen atoms that, in specific intermolecular contexts, could be negatively defined. This issue does not affect any of

the molecules we used in the calculations reported here.

6.3 Constructing the Database

6.3.1 Computational Details

A sufficiently extended sampling is necessary to obtain a good estimation of the polarizability of a specific functional group, given that many different environments could be considered. However, because the amount of possible molecules that contain a particular functional group is extremely large, even under the condition that the molecule contains a limited number of atoms, the chosen molecules for a given group were constructed using a preconceived approach. In particular, molecules containing only C, H, N, and O atoms were used. The 16 common functional groups or atoms shown in Table 6.1 were attached as first neighbours and as second neighbours to the functional group under analysis.

Table 6.1 – Functional groups or atoms used in this work.

Added group	
-H	
-CH_3	Methyl
-OH	Hydroxyl
-OCH_3	Methoxy
-COOH	Carboxyl
-COH	Formyl
-$COCH_3$	Acetyl
-$COOCH_3$	Acetate ester
-NH_2	Primary amine
-$N(CH_3)_2$	Dimethylamine
-$CONH_2$	Primary amide
-$CHCOOHNH_2$	Amino acid group
-$CHCH_2$	Vinyl
-$C(CH_3)CH_2$	Isopropenyl
-CCH	Ethynyl
-C_6H_5	Phenyl

Only one of these functional groups was attached at a time and all other positions were saturated with hydrogen or methyl groups. The cooperative effect of various functional groups is therefore not taken into account. The geometry of every molecule, in its more commonly observed conformation, was optimized with GAUSSIAN09[183] at the B3LYP level of theory using the cc-pVDZ basis set. An extensive conformational analysis, in search of more stable conformers, was not performed. Only for 1,3-propandiol a full conformational analysis was performed with the *Open Babel confab generator*.[184] The results confirm that using more conformations for the same molecule would not increase the information, given the relatively

lower variance. In fact, the standard deviation of the isotropic polarizability of the nine calculated conformers of 1,3-propandiol is 0.0026 a.u. compared to a standard deviation of 0.0056 a.u. for the entire dataset of the hydroxyl group.

After the geometry optimization of the selected conformation, calculations with and without a static electric field were carried out at the CAM-B3LYP/aug-cc-pVDZ level of theory to obtain the wavefunctions and molecular electron densities. Our previous work demonstrated that this level of theory gives quite accurate functional group polarizabilities for many amino acids.[176] The chosen electric field was 0.001 a.u. and it was applied in +x, -x, +y, -y, +z, -z directions, to allow the application of equation (6). *AIMAll*[32] was used to partition the seven obtained electron densities into atomic contributions.

6.3.2 The Local Coordinate System

The 3x3 tensors representing the atomic and functional group polarizabilities are expressed with respect to a coordinate system. For each molecule used to generate the database, this coordinate system is originally oriented with respect to the molecular inertia tensor. However, all the polarizabilities of atoms within a given functional group must be compared based on a common reference frame, independent from that of the molecules they belong to. Therefore, a local functional group coordinate system is necessary to define the functional group polarizabilities that become therefore transferable to any molecule containing that functional group. The definition of a local coordinate system is not unique and it is of course a rather crucial step to properly use the information stored in the database. Several definitions of local coordinate systems are for example found in scattering factor databases used for charge density studies, like the invariom[124] database, ELMAM2[121], or UBDB[179]. In the invariom database, the axes are defined along bonds and the ranking of the bond lengths assigns the axes. ELMAM2 also uses bonds for the assignment; and additionally, seven different axes systems are distinguished depending on the different bonding situations. At variance from our approach, these definitions hold for individual atoms, where the centre of the coordinate system must obviously coincide with the atomic nucleus and the axes orientation depends only on other atoms. In order to implement similar concepts for coordinate systems of functional groups, additional criteria are necessary, because the group has its own inertia system and intra-group bonds. While individual atoms (if assumed just as material points) are inherently spherically symmetric, functional groups are less symmetric and often non-symmetric. Nevertheless, they often have a topological symmetry, which could be larger than the actual symmetry of the molecule. In order to refer the local properties to a molecule-independent system, we defined a local coordinate system for functional groups. For those containing more than two atoms, an inertia-like tensor **I** was calculated using the position of the atoms in the functional group and, instead of the atomic masses, the nuclear charges Z. This is sometimes called the nuclear charge moment tensor [185]. The elements of the symmetric **I** tensor were obtained as follows:

$$
\begin{aligned}
I_{xx} &= \sum_{i=1,N} (y_i^2 + z_i^2) Z_i \\
I_{yy} &= \sum_{i=1,N} (x_i^2 + z_i^2) Z_i \\
I_{zz} &= \sum_{i=1,N} (x_i^2 + y_i^2) Z_i \\
I_{xy} = I_{yx} &= -\sum_{i=1,N} x_i y_i Z_i \\
I_{yz} = I_{zy} &= -\sum_{i=1,N} y_i z_i Z_i \\
I_{xz} = I_{zx} &= -\sum_{i=1,N} x_i z_i Z_i
\end{aligned}
\tag{6.7}
$$

where N is the number of atoms in the functional group, x, y, z the atomic coordinates, and Z_i the atomic number. The main axes of the tensor are taken as the local coordinate system with the origin located at the centre of nuclear charges. Mathematically, the eigenvectors from the diagonalization of the nuclear charge moment tensor correspond to the components of the rotation matrix that rotates the global coordinate system to the local coordinate system, as shown in Figure 6.2a. The rotation matrix must be reordered according to the magnitude of its eigenvalues (while keeping the handedness of the system) in order to obtain consistent results for the same functional groups in different molecules. Furthermore, the same orientation of a given functional group in different molecules must be assured by choosing axes along particular directions specified by given atoms in the functional group. The advantage of this "nuclear charge moment tensor" definition is that it is less dependent on small changes of the molecular geometry compared to other definitions. For functional groups consisting of only two atoms, the local axes system was defined along the bond direction using a procedure very similar to the one developed for ELMAM2[121]. The first axis z' was defined along the bond connecting the two atoms, the second and third axes, x' and y', were simply defined as perpendicular to the first one. For an unambiguous definition, an additional criterion was obviously needed; therefore, a supporting vector s', shown in Figure 6.2b, was defined along the bond to a neighbouring atom and y' was chosen to be perpendicular to the plane spanned by z' and s'. In both cases the vectors were normalized, giving rise to a locally defined right-handed coordinate system.

The rotation matrix Q corresponds to the matrix rotating the global coordinate system into the local one. This matrix must be applied on the originally calculated polarizability tensor as well as on the molecular coordinates to express them with respect to the local coordinate system.

$$
\boldsymbol{Q}^T \boldsymbol{\alpha} \boldsymbol{Q} = \boldsymbol{\alpha}^{FG} \tag{6.8}
$$

After the polarizability tensors were rotated to the local coordinate system, they were symmetrized with respect to the topological symmetry of the functional group given in Table 6.2.

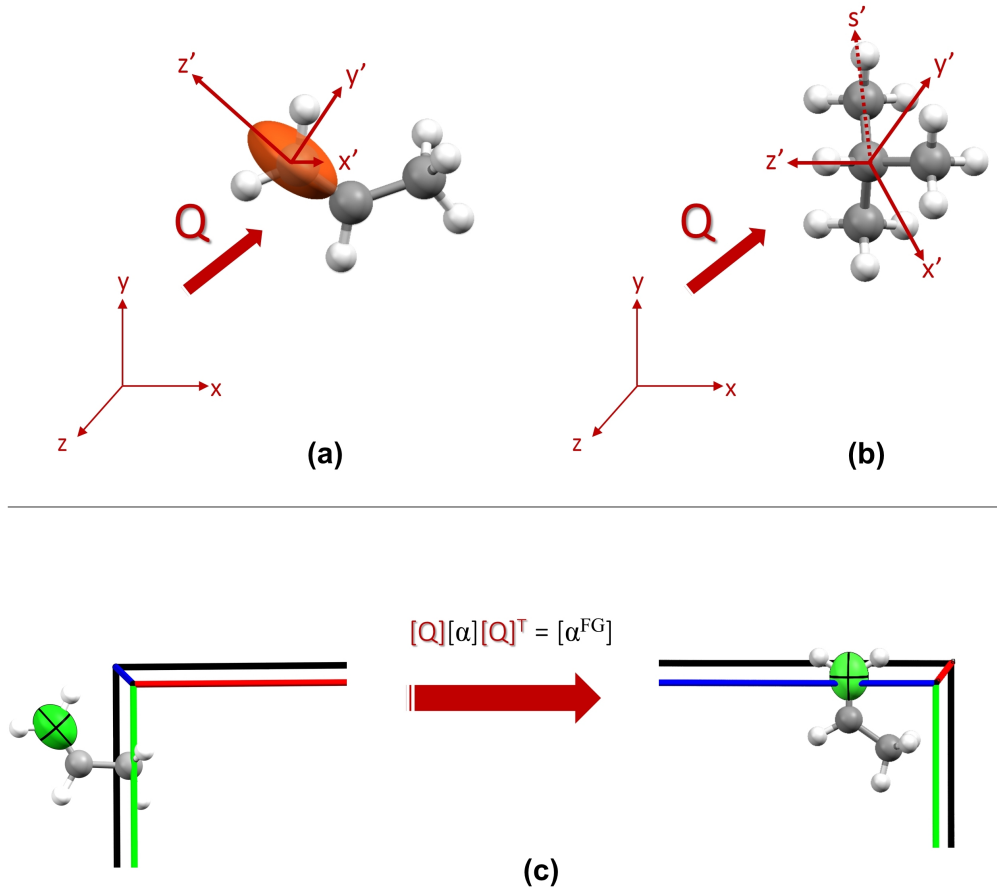
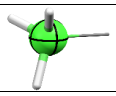
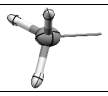
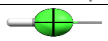
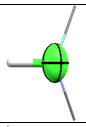
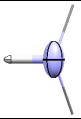
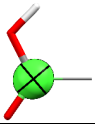
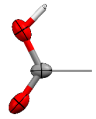
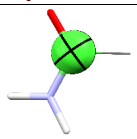
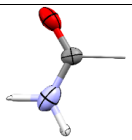
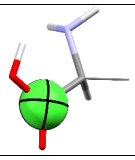
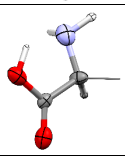


Figure 6.2 – (a) Definition of the local coordinate system based on the nuclear charge moment tensor of the CH_2 functional group. (b) Definition of the local coordinate system along bonds for a functional group with just two atoms (C-H). s' is a supporting vector to a neighbouring atom needed for a unique definition. (c) The matrix **Q** rotates properties from the external to the local coordinate system.

Due to the symmetrization of the polarizability tensor of the functional group, a small fraction of the information about the orientation of the tensor is lost. However, the isotropic polarizability is not influenced by this procedure.

In principle, the database can not only contain polarizability tensors for functional groups but also tensors for each atom in a given group. The atomic polarizabilities can enter the database without further symmetrization besides that required when applying equation (6). In order to use them, they would be averaged over all entries in the database and eventually symmetrized.

Table 6.2 - Molecules created for the CH_2 functional group.

Functional group	Hybridization	Functional group polarizability	Atomic polarizabilities	Topological symmetry of fragment
CH_3	C sp^3			C_{3v}
CH_2	C sp^2			C_{2v}
	C sp^3			C_{2v}
CH	C sp^3			C_{∞}
	C sp^2			C_{2v}
	C sp			$C_{\infty v}$
OH				C_s
NH_2				C_s
NH	N sp^3			C_s
	N sp^2			C_s
COOH				C_s
$CONH_2$				C_s
$CHCOOHNH_2$				C_1

6.3.3 Multivariate Data Analysis and Clustering

Once a representative amount of data is available, more insight comes from statistical analysis of the variables that describe the objects. The tools we adopted for the analysis of this multivariate data are the principal component analysis and the unsupervised cluster analysis.[177]

For a given dataset of calculated polarizabilities the mean $\langle \alpha_{ij} \rangle$ and the standard deviation σ_{ij} were calculated for every element of the polarizability tensor (α_{xx}, α_{yy}, α_{zz}, α_{xy}, α_{xz}, α_{yz}) individually. Afterwards, the data can be standardized according to the equation 6.9. This procedure corresponds to setting the mean to 0 and the standard deviation to 1 for all variables, thus bringing them on a comparable scale. In the case of polarizabilities, this is relevant given the large discrepancy between diagonal (large) and out of diagonal components (smaller, often close or equal to zero). In this way, all the elements of the polarizability tensor have the same weight.[186]

$$z_{ij} = \frac{\alpha_{ij} - \langle \alpha_{ij} \rangle}{\sigma_{ij}} \tag{6.9}$$

The next step is the clustering, which means recognizing that in some molecules a given functional group assumes values that are more similar than in others, implying a sub-classification of groups. For this purpose, one can adopt a cluster analysis. The average of all polarizabilities within the same cluster will be eventually the database entry for a functional group in a specific context. It is expected that this context coincides with a specific chemical environment. For example, the functional group CH_2 may behave differently according to the hybridization of the carbon atom and the nature of its bond(s). Therefore, one may expect to obtain a cluster for C(sp^2), one for C(sp^3) and several sub-clusters depending on the groups attached to the C atom. However, the multivariate analysis is unbiased therefore no *a priori* information is given on expected different CH_2 groups.

For a good representability of the average values and a more appropriate multivariate analysis, the ensemble of tested molecules should be sufficiently large. The purpose of the database is dual: on one hand, it shall represent as many bonding situations as possible for each functional group in order to represent well the “chemical variance”; on the other hand, it must synthesize the information in a not too large number of clearly distinguishable cases. A database would be useless if it is insufficiently representative (too few test molecules) or if it is too much diversified (it contains too many clusters). It is important to note that different clusters can contain different numbers of molecules, meaning that the clustering is only based on the variables and there are no external restraints or interventions.

In order to perform a cluster analysis, it is necessary to evaluate the proximity of two polariz-

ability tensors. For this purpose, we use the Euclidean distance:

$$d_{AB} = \sqrt{\sum_{K=1} (P_K(A) - P_K(B))^2} \tag{6.10}$$

where K is the dimension of the property P. Thus, for polarizabilities the summation runs from 1 to 6. A and B are different data points, thus representing a functional group in two different molecules.

Clustering was performed using a genetic algorithm, a procedure that is often adopted for finding the best set of parameters fitting a very large number of observations. Genetic algorithm use methods that mimic evolutionary processes in nature. A given state can be encoded in a string of numbers (or also matrices, etc.), so called chromosomes. Initially, a set of chromosomes representing different starting states is created giving rise to a set of parent chromosomes. Each of them is attributed a fitness according to the chromosome's ability to minimize or maximize a given function. Subsequently, two chromosomes are chosen out of the parental set and combined to give a new one. The probability of picking chromosomes depends on their respective fitness; the higher the fitness, the higher the probability to use them in the recombination step. Pairing of two chromosomes out of the parental set is repeated until a set of offspring chromosomes is generated with the same size as the parental set. Mutation of offspring can happen, which means that a chromosome is changed at an arbitrary number of random positions. After recombination and mutation, a new set of chromosomes is obtained with potentially higher fitness. While cycles of selection, recombination, and mutation are repeated, the general fitness should increase. The overall fittest chromosome corresponds to the best solution. When the genetic algorithm is applied to our case of grouping polarizabilities, the problem lies in finding an optimum clustering that minimize a global function F defining the average distance between the points in a cluster to their cluster average:[187]

$$F = \sum_{i}^{cluster} \sum_{\boldsymbol{\alpha}_j \in C_i} \| \langle \boldsymbol{\alpha} \rangle_{C_i} - \boldsymbol{\alpha}_j \| \tag{6.11}$$

where $\langle \boldsymbol{\alpha} \rangle_{C_i}$ is the average polarizability of cluster i and the sum runs over all polarizabilities in all clusters. The optimization is done by expressing the cluster averages as chromosomes that are subject to the evolutionary processes. After each cycle of fitness calculation, selection, recombination and mutation all polarizabilities are assigned to the closest cluster average. Thereafter, a new cluster average is calculated based on all the points in this cluster. Then the next generation is produced and this cycle is repeated many times.

The symmetrized functional group polarizability tensors can be grouped into clusters based on either standardized or unstandardized data. The only bias in the procedure is that the number of clusters is not a "result", but it is chosen *a priori* based on certain criteria. There is some degree of arbitrariness in selecting these indicators. A criterion similar to that of Popelier

was adopted here, namely the *intercluster ratio* (I.R.).[188] A Gaussian distribution has a given mean and standard deviation. Using the standard deviation, a confidence interval around the mean, e.g. 2σ, can be specified within which a data point is located with a given probability. The I.R. is a similar criterion to judge the separability of two clusters. An I.R. with a specific number signifies, in analogy to a single distribution, that the points in the two clusters are correctly classified with a given probability. It is calculated for each diagonal component of the polarizability tensor individually. For two clusters A and B the I.R. is defined as the ratio of the difference of their means to the sum of their respective standard deviations:

$$IR = \frac{\langle \alpha_A \rangle - \langle \alpha_B \rangle}{\sigma_A + \sigma_B} \tag{6.12}$$

The number of distinct clusters was chosen such that each pair of clusters has at least one component with an I.R. higher than a given number, which we chose to be 2. This means that for both clusters the chance that a data point is correctly classified is at least 95.4%.

6.3.4 Recognizing a functional group

The clustering into subcategories that we have presented is based on the polarizabilities of functional groups. Of course this information is not available when only the molecular geometry is known and an assignment is needed in order to use the database. Therefore, additional criteria are needed which are based solely on structural features. Calculations for CH_2 show that a clustering based on the inertia tensor of the functional group results in a very similar clustering as that based on the functional group polarizabilities. Selecting two clusters, for example, leads in both cases to a separation of sp^3 and sp^2 hybridized carbon atoms.

This correspondence is essential in order to classify the functional groups based solely on their geometry and therefore to make use of the more detailed information stored in the database. This enables one to make the functional group recognition and polarizability calculation quite automatic, following this simple sequence: a) define the functional groups within a molecule (based on the molecular topology); b) calculate the inertial tensor of each functional group; c) compare the inertial tensors with those calculated for the functional groups stored in the database and assign to the functional group the mean polarizability of the selected cluster; e) calculate the molecular polarizability as sum of the functional group polarizabilities.

6.4 Results

6.4.1 Clustering the CH_2 Polarizabilities

As introduced above, the polarizabilities of a given functional group depend on the type and connectivity of the atoms they consist of, perturbed by the chemical environment (mainly due to *through-bond* conjugative effects along a given connectivity path or through-space

interactions). The clustering procedure described in the previous section is a way to identify and classify these perturbations.

We illustrate here the functional group clustering using the CH_2 group as a working example. Obviously, the polarizability of this group depends strongly on the hybridization state of the carbon atom. An sp^3 hybridized carbon atom will form two additional bonds besides the ones to the hydrogen atoms, while an sp^2 hybridized carbon atom will form only one additional (double) bond. Therefore, we first expect to find a clear clustering of functional group polarizabilities based on the carbon hybridization. The effect of groups bonded to the carbon atom is less easy to predict, although one may anticipate different effects depending on the electronegativity of the bonded atoms, the presence of electron resonance forms and the proximity of other non-bonded groups (that may bear strong charges and therefore produce a strongly polarizing field). Due to the tensor symmetrization in C_{2v}, the principal axes of the CH_2 polarizability tensors coincide with the axes of the local coordinate system. Therefore, the off-diagonal elements are zero and the polarizabilities can be plotted with the values of the diagonal only (Figure 6.3a). In case of functional groups with lower symmetry, the off-diagonal elements are not constrained to zero. In this case, representation of the six independent variables is difficult to visualize, and the correlation is not easily understood. For this reason, a principal component analysis is necessary. This allows to identify the combinations of variables (principal components, PC) that enable explaining the largest variance in the dataset.

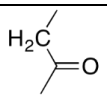
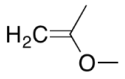
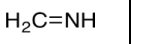
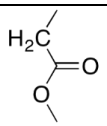
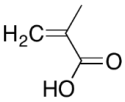
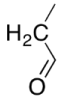
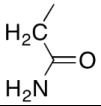
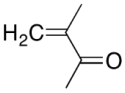
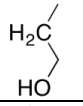
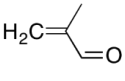
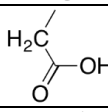
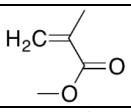
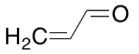
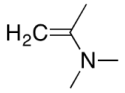
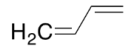
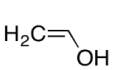
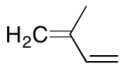
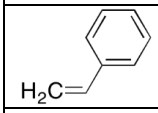
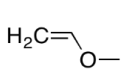
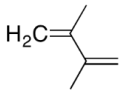
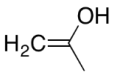
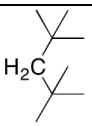
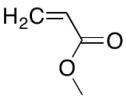
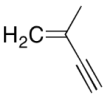
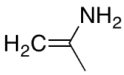
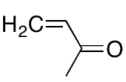
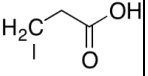
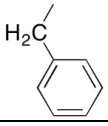
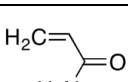
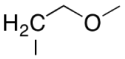
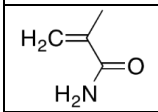
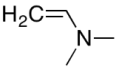
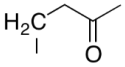
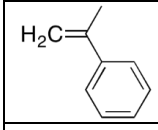
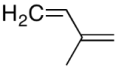
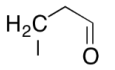
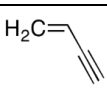
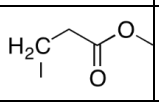
For CH_2, the scores of PC1 and PC2 and the local coordinate system are reported in Figure 6.3b. PC1 is composed of -0.56α_{xx}, 0.57α_{yy}, 0.60α_{zz} and PC2 is composed of 0.74α_{xx}, 0.67α_{yy}, 0.06α_{zz}. Noteworthy is that PC1, which expresses the negative correlation between polarizability in the direction x and y, explains more than 80% of the variance.

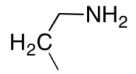
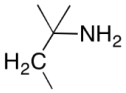
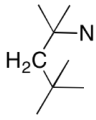
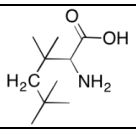
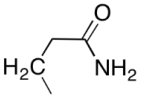
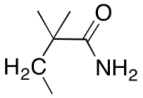
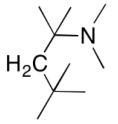
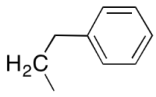
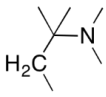
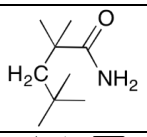
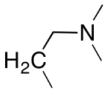
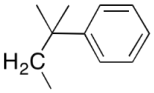
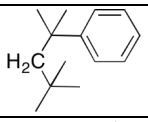
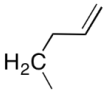
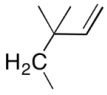
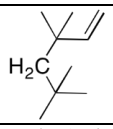
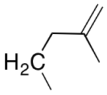
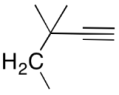
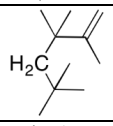
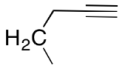
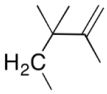
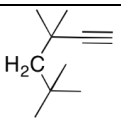
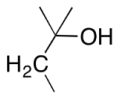
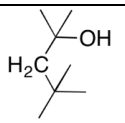
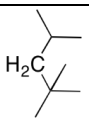
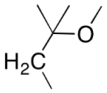
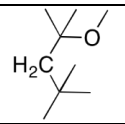
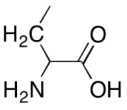
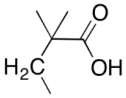
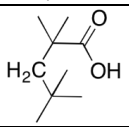
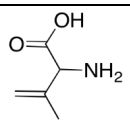
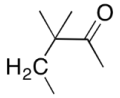
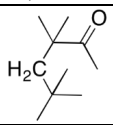
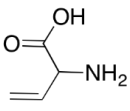
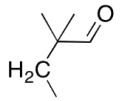
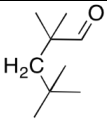
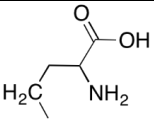
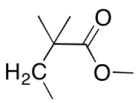
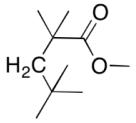
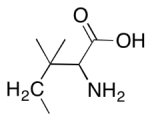
The molecules used for the statistical analysis are listed in Table 6.3.

The colours in Figure 6.3 indicate the hybridization of the carbon atom (red-orange for sp^2; black-blue for sp^3) as well as bonds to heteroatoms (blue or orange). As expected, the polarizability is affected by the hybridization (PC1) more than by the direct neighbour (PC2). The clustering depicted in Figure 6.3b is simply driven by chemical intuition. By applying the genetic algorithm clustering, we can check whether this intuition reflects the data.

In Figure 6.4, results are reported for calculation with 2, 4, 6 and 8 clusters. In keeping with the principal component analysis, the effect of the carbon hybridization overwhelms the presence of a heteroatom. In fact, the result of the clustering with two clusters corresponds almost entirely to the separation of sp^2 and sp^3 hybridized carbons. In the case where four clusters were chosen, the groups with different hybridizations remain separated, but very similar polarizabilities are distributed on three clusters (blue); a result which does not represent a chemically intuitive separation because it is not strictly related to the presence of heteroatoms. When going to eight clusters, the cases where the carbon atom is bound to a heteroatom are in a distinct cluster.

Table 6.3 - Molecules created for the CH_2 functional group.

molecule	label	molecule	label	molecule	label	molecule	label
	1		14f		43		57
	2a1		14g		44		58
	5		14h		45		59
	6		16b		46		60
	7a		16c		47		61
	7e		16d		48		62
	7i		16e		49		63
	7j		14b		50		64
	8a		14c		51		65
	9a		21b		52		66
	10a		28		53		67
	10g		40		54		68
	10h		41		55		69
	14a		42		56		70

	**71**		**84**		**97**		**110**
	**72**		**85**		**98**		
	**73**		**86**		**99**		
	**74**		**87**		**100**		
	**75**		**88**		**101**		
	**76**		**89**		**102**		
	**77**		**90**		**103**		
	**78**		**91**		**104**		
	**79**		**92**		**105**		
	**80**		**93**		**106**		
	**81**		**94**		**107**		
	**82**		**95**		**108**		
	**83**		**96**		**109**		

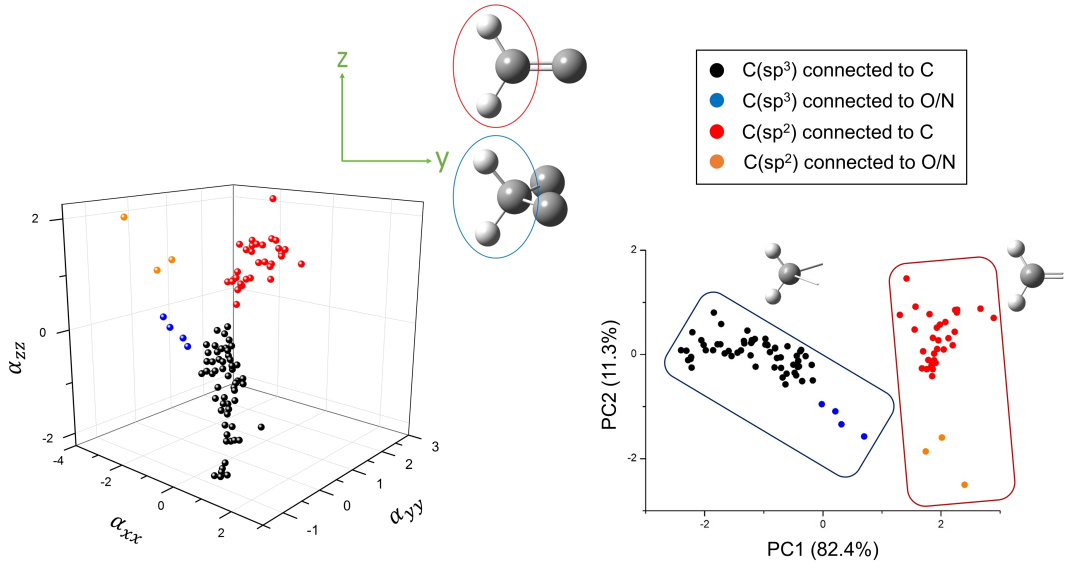


Figure 6.3 – (a) Representation of the standardized CH_2 polarizabilties in three-dimensional space and (b) projected on the plance of the two principal components. The percentage gives the amount of variance explained by each principal component. The blue rectangle highlights functional groups with an sp^3-hybridized carbon atom and the red rectangle highlights functional groups with an sp^2-hybridized carbon atom. furthermore, the orientation of the axes is shown with the green vectors y and z.

6.4.2 Clustering all Functional Groups

Figure 6.5 shows the 456 standardized polarizabilities of all the calculated functional groups plotted in principal component space. Figure 6.5a shows the different types of functional groups, highlighted with colours. The hybridization is indicated with the size of the squares. From this picture, one can see that the functional groups characterized by the highest polarizabilites, namely amino acids, amides and carboxylic acids, present the highest variability as well. Figure 6.5b shows the outcome of a clustering including the polarizabilities of all different functional groups

Figure 6.6 shows the 13 clusters originating from the clustering. The size of the symbol is proportional to the number of entries from a given functional group in this cluster. Since the polarizabilities of the amino acid functional group are rather high, the variability is high as well and they go into different clusters. On the contrary, the NH(sp^2) and CH(sp^2) polarizabilities go into the same cluster.

6.4.3 Using the Database to Compute Polarizabilities

The polarizabilities of some amino acids were calculated and results are summarized in Figures 6.7-6.10. The molecular tensors are computed in four different ways: 1) analytical double

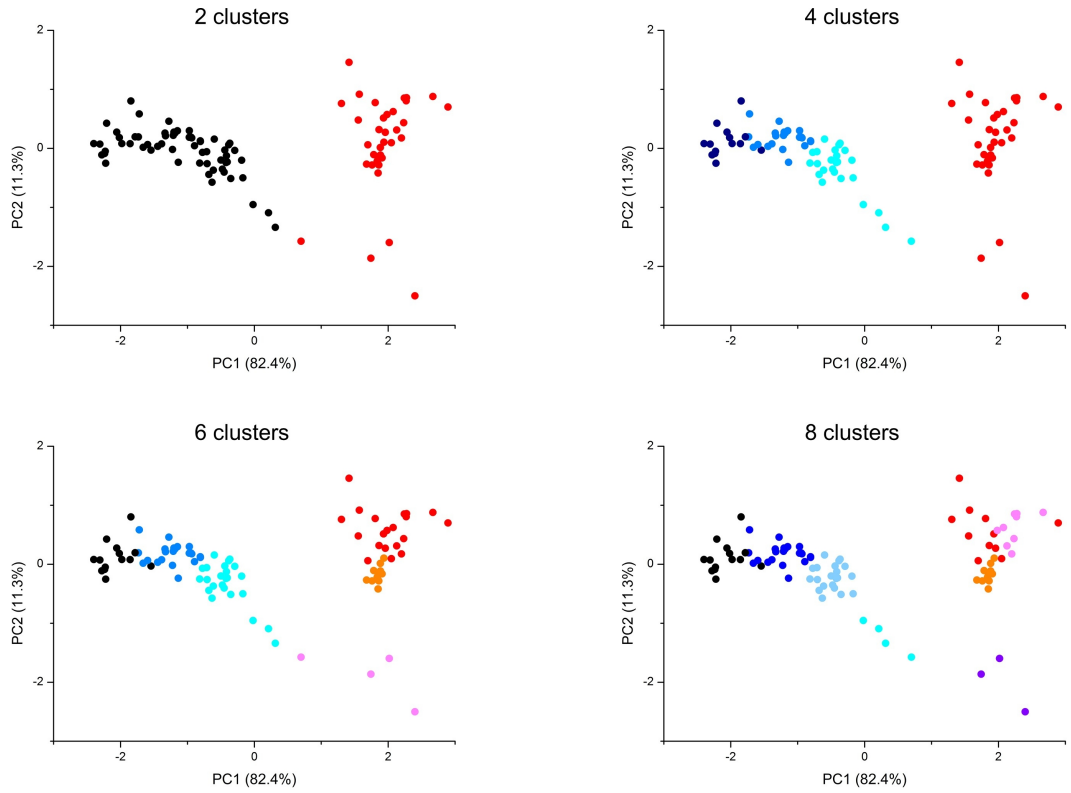
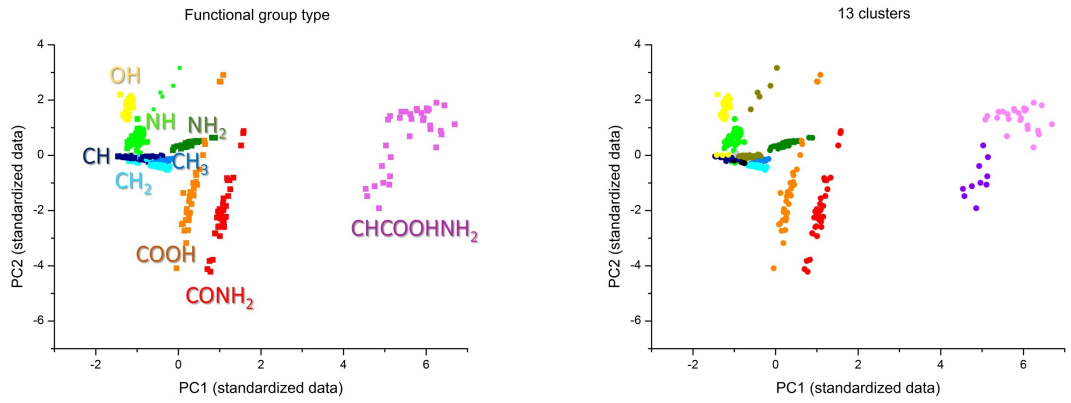


Figure 6.4 – Outcome of the genetic algorithm clustering with 2, 4, 6, and 8 clusters indicated with different colours.



(a) (b)

Figure 6.5 – (a) Polarizabilities of eight different functional groups. For NH, CH2 and CH the size of the squares corresponds to the hybridization of the carbon atom. (The smallest is sp, the middle one sp^2 and the largest squares are sp^3.) This makes the total number of functional groups equal 13. (b) Outcome of the genetic algorithm clustering with 13 clusters shown with different colours.

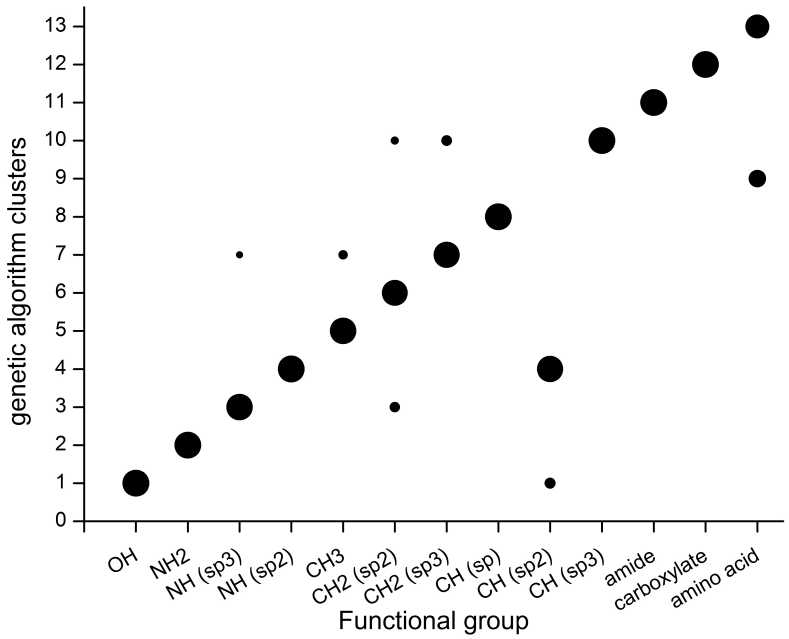


Figure 6.6 – 13 different types of functional groups for which polarizabilities were calculated are shown on the x-axis and the 13 clusters of the genetic algorithm clustering on the y-axis. The size of the circles is proportional to the percentage of the functional group type given on the x-axis in a particular cluster.

derivation of the molecular electronic energy with respect to the field; 2) sum of the QTAIM partitioned atomic polarizabilities obtained from the calculated molecular electron density and finite field approximation; 3) sum of average functional group polarizabilities, from our database entries without clustering; 4) sum of functional group polarizabilities, from our database entries after the genetic algorithm clustering procedure.

Both methods 1 and 2 use couple-perturbed Kohn-Sham theory to calculate the molecular wave function under perturbation of a field. From our previous studies,[182] we already know that the QTAIM distributed polarizabilities (method 2) reconstruct rather accurately the molecular polarizability calculated with method 1, in keeping with the electron density partitioning being an "exact" one. Slight differences are due to the numerical derivatives of the atomic dipoles and the numerical integrations of the density in the atomic basins. Because the purpose of the present work is testing the clustering and database procedures, we use as benchmark the QTAIM distributed atomic polarizabilities (method 2) and do not further discuss in terms of the exact polarizabilities. Moreover, the benchmark polarizabilities are calculated at the same level of theory as the database entries, in order to avoid issues due to the choice of functionals and basis sets (albeit of course important). In previous work, we have shown that the level of these calculations was sufficient to reproduce the experimentally measured molecular polarizabilities of the natural α-amino acids.

The number of clusters per functional group was determined by the intercluster ratio, chosen

in such a way that no clusters have an intercluster ratio lower than 2.

The purpose of the comparison is to assess whether the database is able to predict with sufficient precision the molecular polarizability, compared to a fully *ab initio* calculation on the same molecule, given the enormous difference in computational costs. Moreover, it is interesting to check if using a single entry per functional group is significantly worse than using clustered and thus more specific entries.

Figure 6.7 shows the percentage deviation of the database polarizabilities from the benchmark ones for the functional groups of glutamic acid for α_{xx}, α_{yy}, α_{zz}. Additionally, the functional group polarizabilities were calculated using Miller's[166] isotropic atomic polarizabilities. The deviation of the so obtained polarizabilities to the benchmark ones is also given. As all calculations refer to gas phase molecular geometries, the configuration which is considered is the neutral, instead of the zwitterionic (stable in the solid state) one. We already calculated a set of polarizabilities for zwitterionic forms, by fixing the molecular geometries to neutron diffraction values to mimic the actual behaviour of amino acids in solid state.[176] Figure 6.7 shows that the group most difficult to model with a database is in fact CH_2, although it plays a smaller role for the molecular polarizability because it carries less electrons. Noteworthy is the improvement with the clustering procedure. In fact, the simple database averages produce largest errors up to 25% for the CH_2 group, which are drastically reduced with the clustering (apart for α_{yy} of one CH_2 group and α_{xx} of the COOH group); all other components are estimated with 10% discrepancy from the benchmark. The use of Miller's isotropic polarizability leads in general to a worse approximation of the benchmark polarizability when compared to the database with clustered entries.

The sum of the various functional group polarizabilities tends to compensate some errors so that the molecular polarizabilities are estimated within 11% error on individual components without clustering and within 6% with clustering.

The error of the molecular isotropic polarizability is further reduced (because of error compensation) to 3.66% with the general database polarizabilities and to 1.24% with the clustered values. Given that we are interested in predicting the optical properties of crystals consisting of these molecules, it is important to calculate how this error would propagate, for example on the refractive index. If approximated with the gas phase molar volume, the error on the polarizability corresponds to an error on the refractive index of 0.011, when calculated without clustering, or 0.004 with clustering. In both cases, the deviation is well within the uncertainty of a typical refractive index measurement. This means that calculating an *ab initio* molecular polarizability for the evaluation of a crystal refractive index is not more accurate than using simply the database entries (especially after clustering). On the other hand, the estimation of the anisotropy and therefore the birefringence is worse.
As already discussed in previous work,[176] an accurate evaluation of the crystal refractive index should of course account also for the mutual polarization of molecules. Noteworthy, the functional group approach has an intrinsic advantage with respect to the molecular approach,

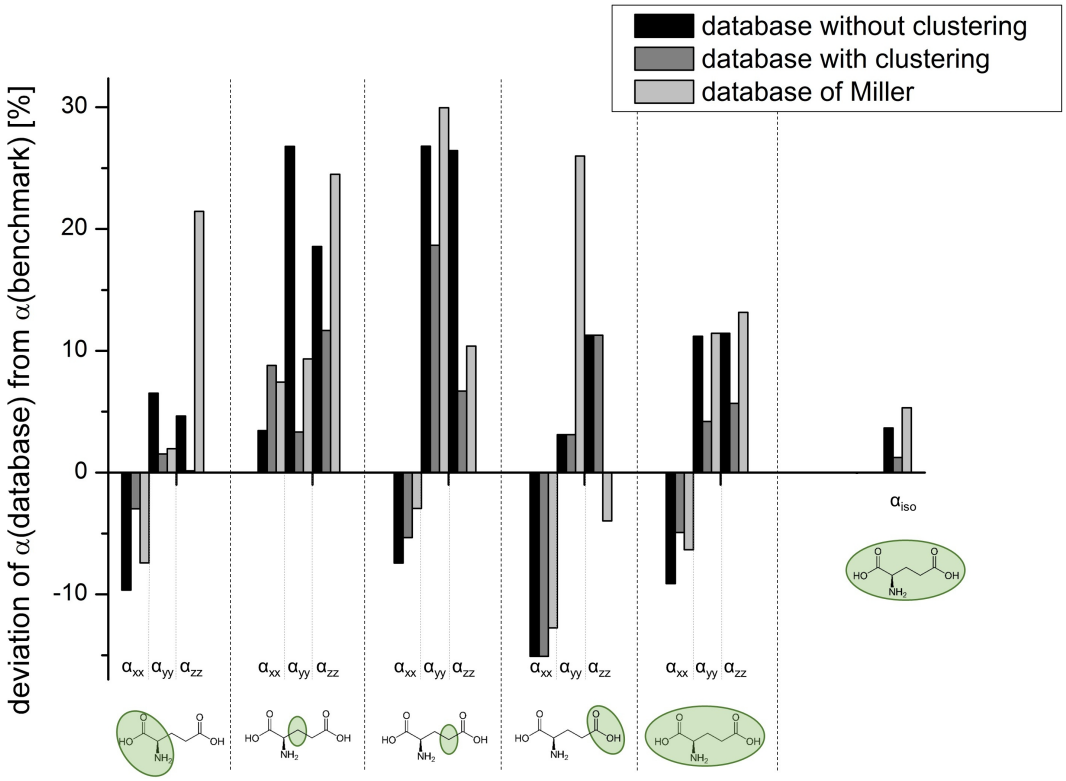


Figure 6.7 – Percentage deviation of the functional group polarizabilities of glutamic acid calculated with different databases from the benchmark values. The diagonal components are shown.

because it allows for a more detailed evaluation of the mutual polarizations.[176] Therefore, the database functional group approach may guarantee a more accurate evaluation of mutual polarizations with respect to the exact, but global, molecular polarizabilities (unless of course they are partitioned into distributed atomic or functional group polarizabilities).

Figure 6.8 illustrates the difference between the benchmark QTAIM polarizability tensors and the same tensors from the database with clustering for glutamic acid drawn with the software *peanut*[189].

A comparison between the benchmark isotropic polarizability calculated with *PolaBer* and the isotropic polarizability calculated from database entries was made for ten amino acids that were not used to construct the database. Figure 6.9 shows the percentage deviation of the molecular isotropic polarizabilities obtained with database entries from the benchmark values. The deviations between the reference values and the values obtained without clustering are shown in black whereas the deviations between the reference values and those obtained after clustering are in red. The average error of the isotropic polarizability among all examined amino acids calculated without clustering is 2.32% while it is 1.43% when calculated with clustering. It is interesting to note that almost all the values are positive, meaning that the

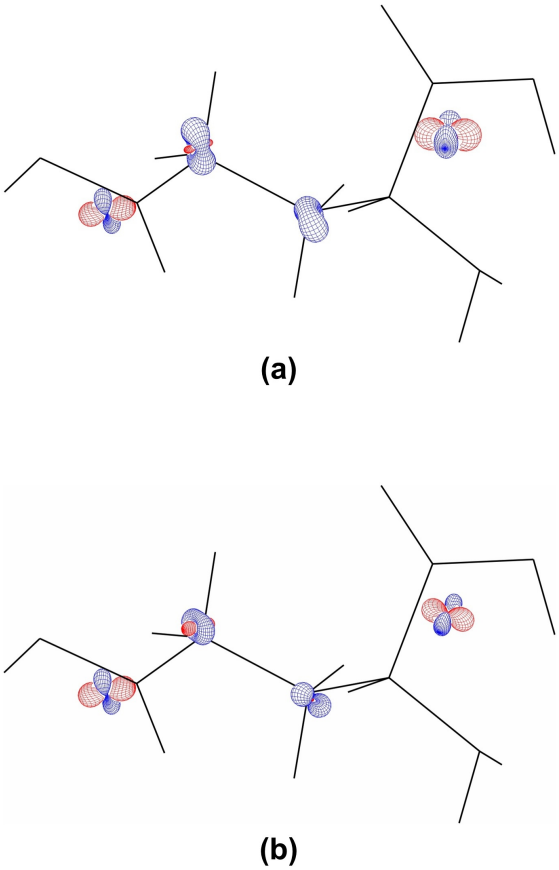


Figure 6.8 – Difference between the benchmark polarizability tensor and the polarizability tensor from the database for glutamic acid drawn at the center of charge. Red colour indicates that the benchmark polarizability is higher and blue that the database polarizability is higher. (a) The database without clustering and (b) the database with clustering.

database polarizability is always higher (Figure 6.10). This is in contrast to many models discussed in section 6.2.1 where the polarizability from the database is usually lower because the mutual induction of polarizability between different fragments is not explicitly taken into account.

The database was also tested for β- and γ-amino acids (Figure 6.11), which reveal similar features as the natural amino acids. The deviation of the polarizability obtained with the database from the benchmark polarizability is shown for the α_{xx}, α_{yy}, and α_{zz} component. The results from the database with clustering are better than when no clustering is performed. The deviation of the isotropic polarizability is less than 3% in both cases, while it is larger for the individual components.

These analyses enable us to conclude that the performances of the database are very promising, because they allow to estimate with accuracy of ca. 2% the isotropic polarizabilities for molecules that were not used to build the database. For anisotropic components the

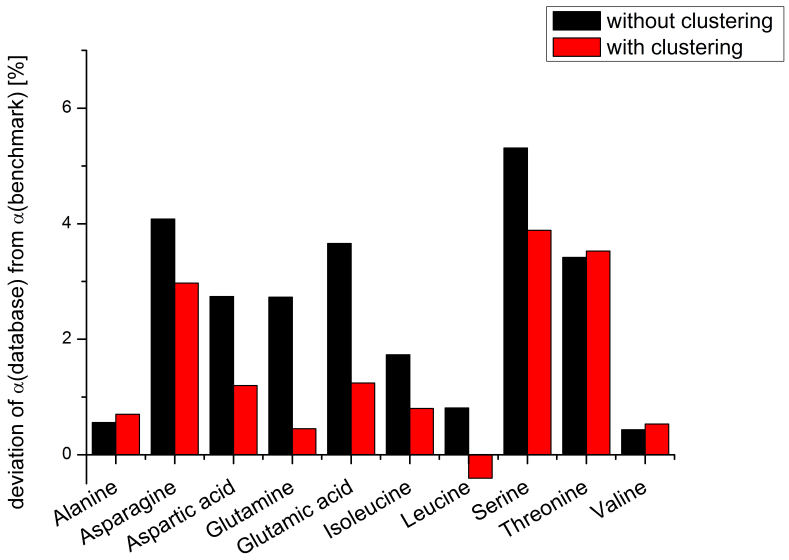


Figure 6.9 – Plot representing the percentage deviation of the database polarizabilities from the benchmark polarizabilities for ten amino acids. Black columns refer to a database without clustering and red ones to a database with clustering.

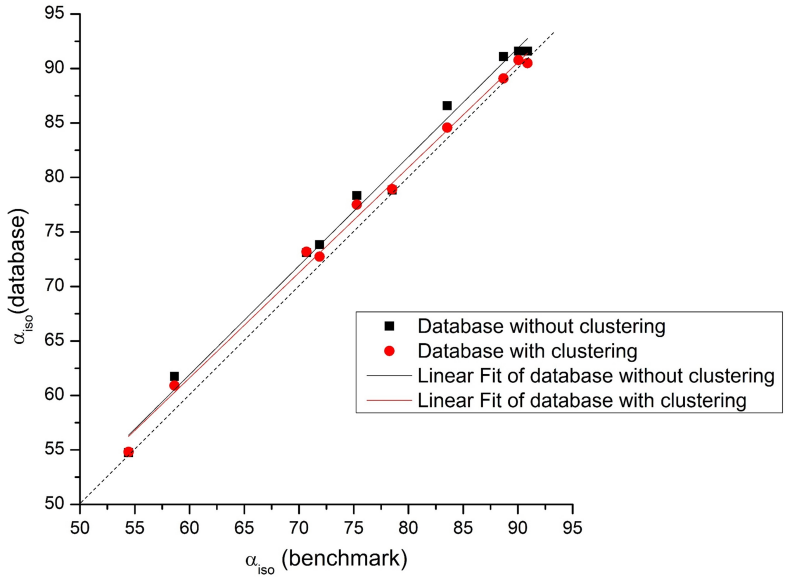


Figure 6.10 – Comparison of the isotropic benchmark polarizability to the database polarizabilities for the ten calculated amino acids. The dotted line represents perfect coincidence.

discrepancies are higher, but still acceptable. Therefore, we may conclude that the necessary condition of transferability is demonstrated. Of course, the success of the database depends also on the quality of the calculations on the set of test molecules. In this respect, there is room for improvement, for example using functionals and basis sets that are able to approach even better the experimental results. For this purpose, more tests are necessary as well as more experimental information. Noteworthy, the computational costs to produce a similar

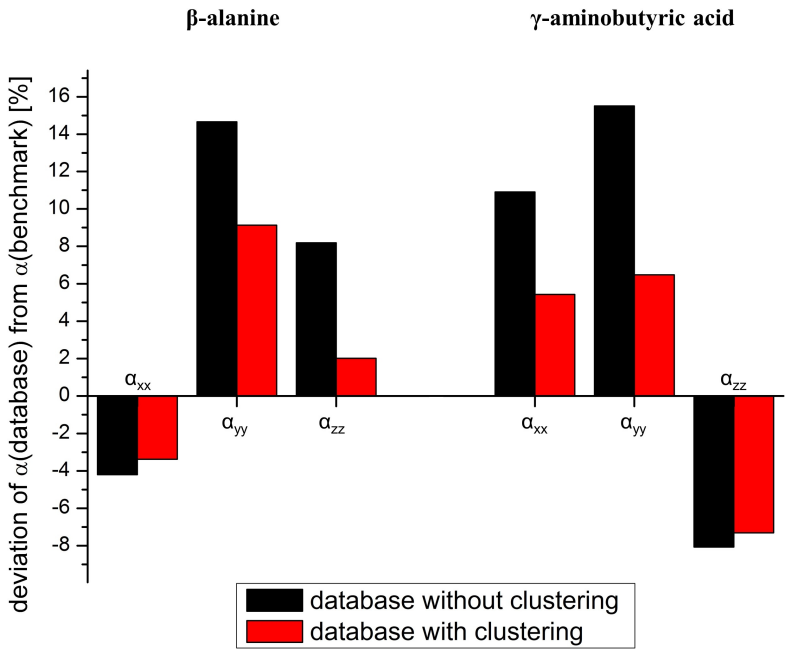


Figure 6.11 – Percentage deviation of the polarizabilities of β-alanine and γ-aminobutyric acid calculated with the database (with and without clustering) from the benchmark values. The diagonal components are shown.

database with a more sophisticated level of theory are not so high Nevertheless, the current database seems to be already suitable for the purposes mentioned in the introduction.

The database will be included in the forthcoming new version of the software *PolaBer*. For each functional group, the averages of atomic and group polarizability components are stored (in Bohr^3), together with the averages within each clusters. The database molecules used for calculations are listed for each cluster.

In the future, the program *PolaBer* will be modified in order to recognize a functional group based only on geometrical features and to automatically propose adequate polarizabilities (average group polarizabilities or specific cluster). The program will allow users to quickly estimate the polarizability of any given molecule for which they just need to input the geometry. Considering also the molecular volume, which may be given by the user or also stored in the database, dielectric properties can directly be calculated.

6.5 Conclusions

In this book chapter, we presented a methodology to derive a functional group database of fully exportable polarizabilities. We defined a local coordinate system that enables the transferability of functional group polarizabilities. These polarizabilities were symmetrized and clustered into subcategories using genetic algorithm. The database entries thereby obtained

were used to estimate the polarizabilities of different molecules.

The work has many purposes: 1) a multivariate analysis is not only a way to produce representative values for given quantities, but also a method to gain insight into the machinery of electron polarizability effects; 2) the database provides means to rapidly estimate polarizabilities in larger molecules or aggregates consisting of the building blocks present in the database; 3) the derived molecular polarizabilities can be used as such in force fields for the evaluation of polarization and dispersion interactions in molecular recognition processes; 4) the polarizability can be used to rapidly estimate dielectric constants in the solid state as well as derived quantities such as refractive indices.

One of the goals of our research is in fact connected with point 3): The availability of a polarizability database, especially if extended to all main group atoms and to transition metals as well, would enable a very rapid screening of molecular/polymeric materials for optoelectronic applications, for example HRI materials[151, 152, 153, 155, 156, 157, 158, 159, 160, 161, 162, 190] low dielectric constant porous materials[154], etc.

Moreover, our purpose is that of giving the guidelines for extending the database, so that interested researchers could contribute and expand it by adding functional groups interesting for their own specific purposes. Our future research, apart from further development of the database by adding more descriptors (e.g. dipole moments, hyperpolarizabilities, etc.) and improving the statistics with more test molecules, will be dedicated to improve the automatism of the calculations, using adequate routines that will enable recognition of the functional groups. Given the speed of database calculations of polarizabilities and the subsequent derivation of static dielectric constants and refractive indices, an automatic procedure of structure screening would be extremely rapid. These structures could be actually observed and already characterized structures (for example present in structural databases) as well as structures simulated during processes of crystal structure prediction. The implications for material design are then rather straightforward.

Acknowledgement

We thank the Swiss National Science Foundation for financial support (project 160157).

6.6 Addendum

6.6.1 Comments and additions to the published article

This study substantially contributes to the long-term goal of establishing a database of polarizabilities, which should enable a fast screening of optical properties. We showed how to automatically sort polarizabilities of various functional groups into clusters using genetic algorithms. The average of all functional group polarizabilities in one cluster is representative of this cluster and stored as database entry. Subsequently, we used this entry and imposed it on the respective functional group in other molecules. However, this back-transformation is ill-defined because it is unclear which database entry to take for a given functional group. This ambiguity arises due to the fact that the database entries were created based on an automatic clustering according to the similarity of their polarizability tensors and not based on well-defined geometrical criteria. We tried to establish links between geometry and polarizability such that the assignment of the correct database entry could be done based on the molecular structure. Indeed, the polarizability is closely linked to the structure, and strong correlations between them were found, but actually the purpose of the automatic clustering is exactly to *not* use geometrical criteria. As described above, this leads to the dilemma that the clustering is based on the polarizablity of a certain set of test molecules, while at the same time the polarizability is exactly what one does not know *a-priori* for a given target molecule because it is the quantity one wants to extract from the database. Therefore, ambiguity in the choice of the correct cluster is unavoidable. The only sensible way to circumvent this problem is to define criteria for the identification of functional groups based on the geometry, which take into account its chemical environment similar to how it is done in the multipolar pseudoatom databases. When well-defined geometrical criteria are introduced, the back-transformation is unambiguous.

However, genetic algorithm clustering is still useful also when clusters are defined based on geometrical criteria. Because in that case, the clusters defined according to geometrical criteria could be compared to the clusters obtained from the genetic algorithm clustering. With this comparison, one could learn whether the clusters defined according to geometrical criteria have a good size or not and thus, whether their entries have similar enough polarizabilities or if more or less clusters are needed. In case more clusters are needed because the polarizabilities within one cluster are too dissimilar, clusters should be split into several ones. On the contrary, if the polarizabilities of two or more clusters are very similar they could be merged into one. Subsequently, the well-defined geometrical criteria would be adjusted to that.

Indeed, Figures 6.5 and 6.6 point out that this approach is reasonable because the clusters from the genetic algorithm clustering resemble a lot the ones one would define by chemical intuition. The strongest deviation from that is represented by the CH(sp^2) and the amino acid groups, which seem to have quite varying polarizabilities and form different clusters.

It is also be pointed out again that the polarizabilities are symmetric tensors with six independent components. For graphical reasons they were reduced to two, using principal component analysis, but the remaining four components also carry information. For this reason different

clusters marked with different colors seem to overlap in Figures 6.4 and 6.5. But points which are close in the 2D plots, do not necessarily need to be as close as that in the 6D space of the polarizability tensor.

Besides this more fundamental issue, smaller technical problems with the definition of the local coordinate system based on the inertia tensor arose. In fact, its three axes are well-defined with the use of equation 6.7, but not, which of the three directions corresponds to the x, y, or z axis. This was solved by defining specific criteria adjusted for every one of the nine studied functional groups.

In addition to the clustering with genetic algorithms, also clustering based on k-means and on a dendrogram were checked. Figure 6.12 shows clusters obtained with the three different algorithms using polarizabilities of the CH_3 group in various chemical environments. The k-means clustering uses an algorithm that minimizes the same overall function as equation 6.12, except that the difference is squared. Indeed, the resulting clusters are often similar to the ones obtained with genetic algorithm clustering. However, with the k-means algorithm the solution is more prone to fall into local minima that are not the overall optimum solution, because k-means lacks the random element introduced in the genetic algorithm. Clustering based on a dendrogram very often results in one cluster containing almost all entries while the others have just one entry, and therefore it is clearly not suitable for this problem.

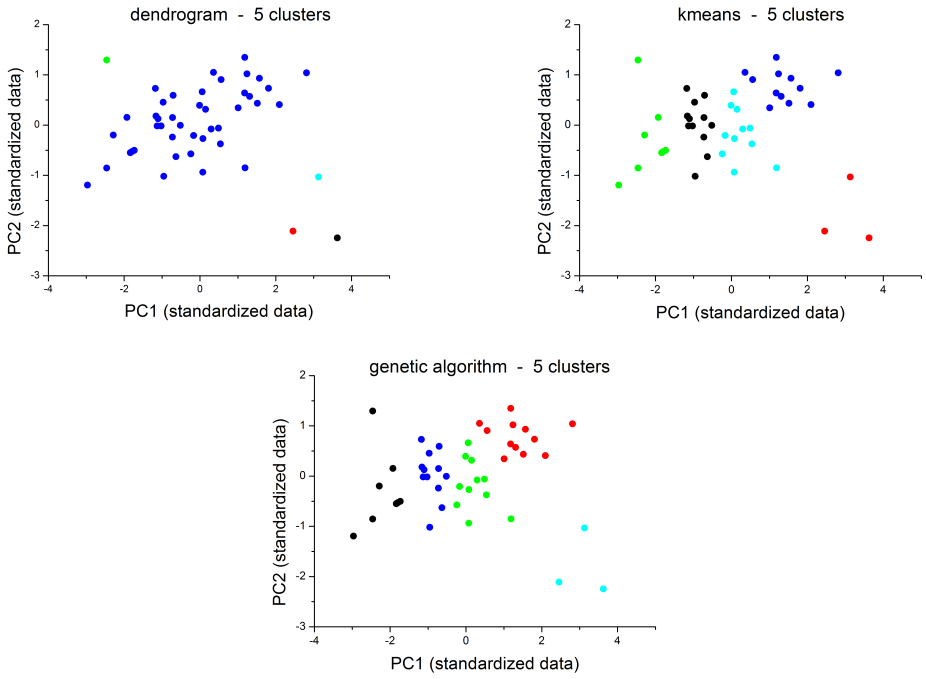


Figure 6.12 – Three algorithms sorting the CH_3 polarizabilities into five clusters. For each clustering algorithm, the colors distinguish the five obtained clusters.

6.6.2 Related work of other researchers

It was shown that for hydrogen bonded synthons, like for example a dimer connected through carboxylic acid groups (COOH-COOH interaction), the prediction of the refractive index can be improved compared to the experimental value when the synthons are treated explicitly. This means that the polarizability of the entire dimer is calculated and this value is used to obtain the crystal refractive index. In contrast, when only the polarizability of the monomer is explicitly calculated the agreement with the experiment is worse. [191] The database can without problems be expanded to contain also the most common synthons.

In another study [192], the molecular electrostatic potential (MEP) and the polarizability of four dipeptides consisting of L-alanine and glycine were analysed. Four different protonation states (anionic, cationic, neutral zwitterionic and neutral non-zwitterionic) of the dipeptides were examined as well as different conformers. Dipeptides were chosen because they are well suited to understand how amino acids and small biomolecules in general behave in different chemical environments. The ultimate goal would be the correlation of structure and properties of small building blocks with the properties of large biomolecules. As previously written, the polarizablity informs about the reaction of the electron density towards an external stimulus such as the proximity of other molecules or a solvent, whereas the MEP is important to highlight sites for possible nucleophilic and electrophilic attacks or interactions.
The authors found that both the MEP and the polarizability of the peptide bond do not change a lot among different conformers. The polarizabilities of the atoms are more variable highlighting that for the purpose of transferability, functional groups are better suited. The protonation state instead is important for both properties. Thus, when including dipeptide bonds into a database, it seems necessary to have a different entry for all four protonation states.

Wilkins *et al.* [193] recently set up a database for atomic polarizabilities, which they made accessible online. Their goals were very similar to the ones that led to the work presented in this chapter: enabling a fast prediction of molecular polarizabilities and allowing for its rationalization in terms of building blocks. In our case atomic polarizabilities are obtained from a partitioning of the electron density, while with their approach, they result from a machine learning of molecular polarizabilities. In both cases the molecular polarizability is the sum of the atomic polarizabilities. For the machine learning they used a training set consisting of more than 7000 small organic molecules containing H, C, N, O, S, and Cl. The polarizabilities of those molecules were obtained with linear response CCSD calculations. Once the so-obtained polarizabilities were machine-learnt with their algorithm, they used them to predict the polarizabilities of molecules not contained in the training set. The authors showed that the so-obtained molecular polarizabilities are better than explicitly calculated DFT ones, in the sense that the error compared to *ab-initio* calculated CCSD polarizabilities is smaller.

Atomic polarizabilities have also been used to improve multipolar pseudoatom databases by

Leduc *et al.* [194] In the ELMAM2 database described in section 4.3 multipole parameters are stored for various atom types. These are averaged values of a particular atom in different environments. The parameters can be transferred to different systems, and a multipolar electron density of those system is obtained quasi immediately. In order to define an atom type, only information about its covalent bonds is considered. Therefore, the multipole parameters lack information about the non-covalent environment. However, the non-covalent environment also has an influence, because it polarizes the electrons. Leduc *et al.* tried to retrieve this information by taking into account the local electric field at every nucleus and modifying the dipole parameters. From the initial multipole parameters the electrostatic potential is calculated, which in turn gives rise to the electric field. With this field and the atomic polarizabilities from a database as described below, the dipole moment of every atom is calculated as in equation 6.1. Thus, one obtains improved dipole moments. With those, a new electric field was calculated, and the dipole moments were improved by repeating this cycle to convergence. The authors used the improved dipole moments to calculate intermolecular electrostatic interaction energies.
In order to obtain the atomic polarizabilities, a dataset of interacting dimers was taken and the electron densities of the dimers and monomers calculated with and without electric field. With a QTAIM partitioning, dipole moments were obtained, but for the calculation of atomic polarizabilities they considered only the atomic polarization term $\boldsymbol{\mu}_p(\Omega)$ and not the charge translation term $\boldsymbol{\mu}_c(\Omega)$ whose effect they claimed to be weak. They found that using the polarizabilities from the monomer calculation led to slightly better results than using the ones from the dimer calculation where some polarization is already in the polarizabilities themselves and thus double-counted.

As mentioned in this chapter, there are many different ways to obtain polarizabilities of molecular building blocks and transfer them to other compounds. Virtually all of the approaches described in literature rely on atomic polarizabilities, although functional group polarizabilities are actually more transferable because they are more constant.[176] The amount of common functional groups is not enormously large in organic chemistry, which justifies the establishment of a database with functional group polarizabilities.

7 Pressure-Induced Polymerization and Electrical Conductivity of a Polyiodide

Apart from the addendum, this chapter originates from the previously published article "Pressure-Induced Polymerization and Electrical Conductivity of a Polyiodide" in *Angewandte Chemie International Edition*, **2019** [142] and *Angewandte Chemie*, **2019** [195].
(https://onlinelibrary.wiley.com/doi/full/10.1002/anie.201901178 and https://onlinelibrary.wiley.com/doi/full/10.1002/ange.201901178)

Abstract: *We report the high-pressure structural characterization of an organic polyiodide salt in which a progressive addition of iodine to triiodide groups occurs. Compression leads to the initial formation of discrete heptaiodide units, followed by polymerization to a 3D anionic network. Although the structural changes appear to be continuous, the insulating salt becomes a semiconducting polymer above 10 GPa. The features of the pre-reactive state and the polymerized state are revealed by analysis of the computed electron and energy densities. The unusually high electrical conductivity can be explained with the formation of new bonds.*

Polyiodides (PIs) in crystal form were discovered about 200 years ago. [196] Their structural diversity, which is due to the bonding flexibility of iodine, remains the subject of continuous interest. [197, 198, 199] Nowadays, various PIs are known, ranging from I_3^- to I_{29}^{3-}, with the general formula I_{2m+n}^{n-} and n up to 4. Their building blocks consist of I_3^- as a donor and I_2 as an acceptor forming a charge transfer (CT) complex. CT complexes can interact further to form extended structures with various topologies, up to cubic networks. As the aggregation of iodine units is somewhat unpredictable, the design of PIs is not easy. Moreover, the classification of higher units is, thus far, pragmatically based on the distance between iodine atoms, but universal criteria are still lacking.
Different theoretical approaches to study bonds in PIs exist, for example, calculations of the potential energy surface and bond order, [200] energy decomposition analysis (EDA), [201] analysis of the electron density and its Laplacian at bond critical points (BCP), [202] the electron localization function, and the one-electron potential. [203] Furthermore, the energy

densities at BCPs [204] and the interacting quantum atom (IQA) method [205] were used to analyze halogen bonds. [206]
Depending on the forces exerted by the organic counterions (chemical pressure), PIs exhibit useful electrochemical properties, such as charge-carrier transportation, high electrolyte energy densities, high redox reaction reversibility, [207, 208] and a wide range of electrical conductivity properties, from insulating to metallic behavior. [209, 210] Hence, PIs have found technical applications in electronic and electrochemical devices such as flow batteries, fuel cells, dye-sensitized solar cells, and optical devices. [207, 211, 212]

Compared with chemical pressure, an external mechanical pressure more substantially affects the crystal inter- and intramolecular landscape. In fact, a large lattice strain may induce phase transformations and even chemical reactions. The achievable static high pressure (HP) in standard diamond anvil cells (DACs) is on the order of tens of gigapascals. Such a pressure significantly changes the Gibbs energy, increasing the internal energy and making the *PV* term dominant. Similarly large energy changes are, however, not accessible in solids by temperature alteration, which induces sublimation or melting instead because of the dominant entropic term. The volume contraction is the main driving force for the addition reaction to atoms linked by unsaturated bonds and the formation of extended assemblies, which are otherwise unstable at ambient pressure. One can apply this concept also to polyvalent iodine, which is typically able to expand its coordination in heptaiodides or organic iodanes. [213] Although compression may produce exotic chemical bonding situations and properties, only few spectroscopic studies have focused on PI at HP. It was demonstrated by Raman spectroscopy that solid organic PIs can catenate under HP. Pentaiodides form I_3^- chains in iodine-doped poly(vinyl alcohol) films at pressures as low as 0.2 GPa, [214] while a pressure between 0 and 3.4 GPa is needed for bond formation in PI confined in carbon nanotubes. [215] Such a catenation does not occur in aqueous solution, advocating solid phases as the preferred matrix for a reaction between iodine units.

In this study, we investigated the response of tetraethylammonium diiodine triiodide (TEAI) to compression by using powder and single-crystal X-ray diffraction, electrical conductivity measurements, and first principle calculations. TEAI was chosen after extensive screening of the Cambridge Structural Database [216] (entries HALTAP [217] and HALTAP01 [218]) in order to find a moiety with appropriate packing, ready for catenation between donor (I_3^-) and acceptor (I_2) upon compression. Quaternary ammonium salts are often used as model cations for PI studies. [219, 220, 221, 222] Under ambient conditions, TEAI crystallizes in the orthorhombic *Aba2* space group. I_2 and the apical iodine atoms of I_3^- anions form, via contacts d_1 and d_2 (Figure 7.1), zig-zag chains extending along the crystallographic *c* axis, while oscillating in the *a* direction. The same contacts form another zig-zag motif propagating along the *a* axis and oscillating mainly along the *b* axis. These supramolecular chains encapsulate the isolated cations. Parallel chains form bilayers normal to the *b* axis owing to the linkage with I_3^- spacers (Figure A.2). The distances between one I_2 motif and its neighboring two I_3^- units are unequal (d_1 and d_2 ; Figure 7.1 and Figure 7.2b). They approach the acceptor at angles α and β, respectively, which are favorable for the formation of higher units. This

occurs when the I_3^- and I_2 molecules are close enough for their frontier molecular orbitals to overlap. In extended systems, this interplay gives rise to valence and conduction bands. [223] When the interaction is stronger, the bands broaden, and the band gap decreases, leading to a semiconducting behavior, as observed in many species. [224, 225, 226]

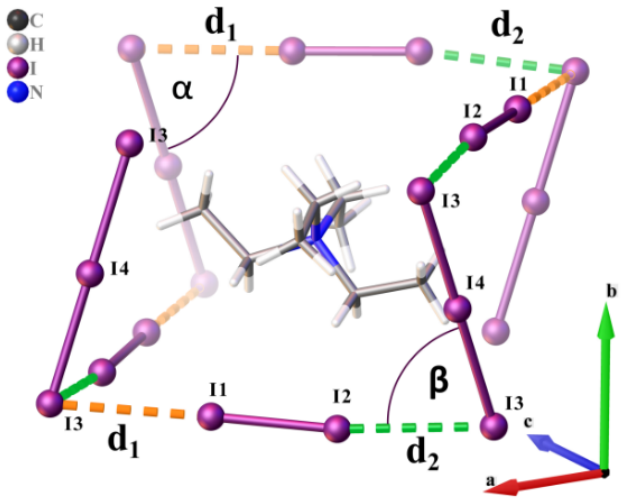


Figure 7.1 – The iodine units around the tetraethylammonium cation in TEAI at ambient pressure. The shortest (d_1 = 3.479(4) Å) iodine-iodine intermolecular distances are marked in orange and the second shortest (d_2 = 3.494(4) Å) in green.

We performed X-ray diffraction experiments on TEAI in various pressure transmitting media (PTM). Herein, we report data obtained using nitrogen as the PTM, known to be hydrostatic up to about 10 GPa [227] and chemically inert. The effects of other, less hydrostatic PTM will be discussed in a forthcoming full paper. Upon pressurization up to 12.8 GPa, the crystal volume of TEAI contracted by 35 %. The *b* axis, which is normal to the corrugated iodine bilayers, shows the highest linear compressibility because of a lack of strong intermolecular interactions in this direction. Molecular volume compression trends (Figure 7.2a) do not show any discontinuity, ruling out the occurrence of phase transitions. The lengths of the unit cell axes *a* and *b* are well predicted by periodic density functional theory (DFT) calculations over the entire pressure regime whereas the length of the *c* axis is slightly overestimated. The length of contact d_2 is well simulated whereas the underestimation of d_1 at low pressure is due to temperature effects neglected in the simulations, whereas all high-pressure experiments have been carried out at ambient temperature. The temperature effects on the lattice parameters were estimated by an additional diffraction experiment at 110 K and ambient pressure (Table A.4). Moreover, in the vicinity of the equilibrium geometry, the calculated energy surface is very flat (see Figure 7.3 for I_5^-), which is the reason why some DFT functionals or basis sets may fail in predicting the geometry of the adducts, especially when perturbed by the electric field inside a crystal.

The compression induces a continuous decrease of α and β, perpendicular to the iodine

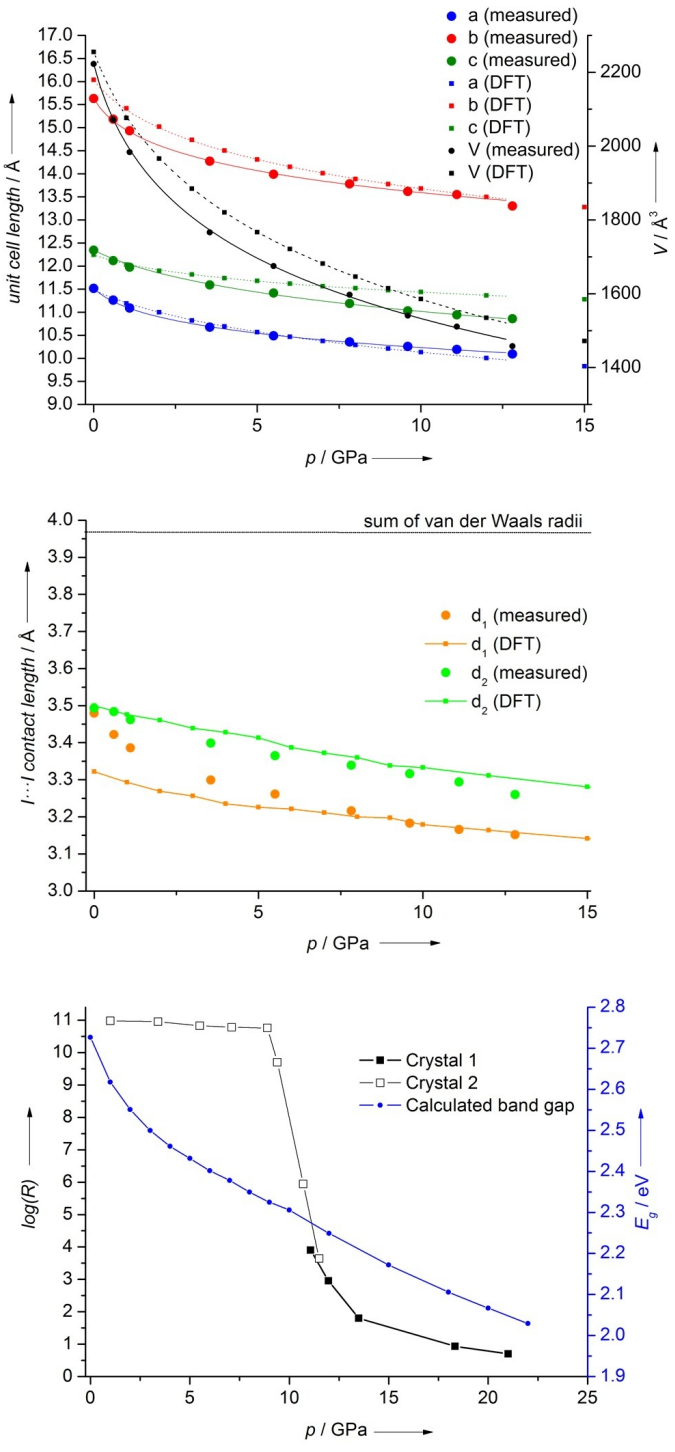


Figure 7.2 – a) Unit cell axes and volume of TEAI as a function of pressure. Lines represent third-order Vinet equations of state [228] (fitted using EoSfit7c [229]). The calculated bulk modulus is 7.7(5) GPa, which coincides with that of pure I_2. [230] b) Evolution of d_1 and d_2 (see Figure 7.1) with pressure. c) Electric resistance (measured on two different samples) and calculated band gap as a function of pressure.

bilayers (Figure 7.1 and Figure A.3). The contact d_2, subtending the β angle, is less compressible than d_1, subtending α (Figure A.3). As the pressure increases, differences between the respective distances d_1 and d_2 as well as α and β broaden. This forces the iodine zig-zag motif to bend out of its initial shape. This process is illustrated by a gradual increase in torsion between the triiodides (Figure A.3). Compression of the intermolecular iodine bonds does not cause any notable distortion of the intramolecular geometry of the I_2 and I_3^- units.
To gain more insight into the nature of the bonds in PI, we used theoretical tools for bonding analysis. In TEAI, various interactions of PI units occur among themselves and with the cation, which complicates the analysis. For simplification and abstraction, we took I_5^- calculated in vacuo as benchmark. The stable structure of I_5^- has C_{2v} symmetry, which can be seen as two I_2 molecules symmetrically interacting with a central I^- (at distances d_1 and d_2) in a bent structure. In the solid state, this symmetry is distorted because of the cations present and the interactions with other I_3^- and I_2 moieties. We explored the potential energy surface of the system (Figure 7.3) and analyzed the electron density indicators along the transformation path leading to I_5^-. Figure 7.3 shows that there is a similarity between the in vacuo potential energy surface and the one retrieved from the experimental crystal structures within the framework of structure correlation, [231] although, as expected, the distances are typically shorter in the solid than in vacuo. In fact, one can easily observe an ideal translation of the crystal potential energy surface along the diagonal of the diagram in Figure 7.3.

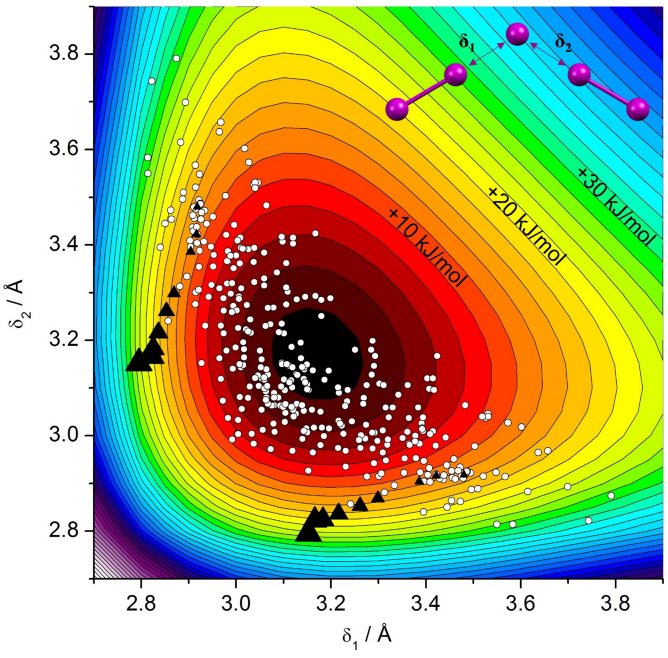


Figure 7.3 – The electronic energy of I_5^- calculated in vacuo, as a function of δ_1 and δ_2. The color indicates the destabilization from the symmetric I_5^-. The white circles represent entries in the Cambridge Structural Database containing I_5^-. The black triangles represent TEAI at different pressures (the triangle size is proportional to the pressure).

The effect of the mechanical pressure is completely different: The compressed structures of

TEAI significantly deviate from the transformation paths of the in vacuo or crystal potential surfaces at ambient pressure. It seems that the HP structures would not converge towards a symmetric I_5^-. This reflects well the multiple I_3^- – I_2 interactions in the crystals, which significantly perturb the picture of a localized I_5^- moiety (when compressed), and anticipates the catenation.

If only atomic positions are analyzed, the formation of a genuine covalent bond may not be identified. Even the electron density itself would provide only partial indication of the nature of the chemical bonding, especially under compression. Therefore, we used a decomposition of both the electron density (through quantum theory of atoms in molecules [232]) and the energy (through EDA and IQA [42]) to analyze the nature of this interaction as a function of δ_1. EDA and IQA are different schemes to partition the bond energy into classic, exchange-correlation, and kinetic terms. For I_5^-, the two methods provide the same picture (Figure 7.4): The classic (electrostatic) contribution always dominates and increases with δ_1, whereas the non-classic (exchange) contribution never exceeds 25% even at short δ_1 distances. The exchange is larger than in KI, but substantially smaller than in I_2. Thus the I–I bond in CT complexes is simultaneously of ionic and covalent nature, [233] although the ionic one always dominates.

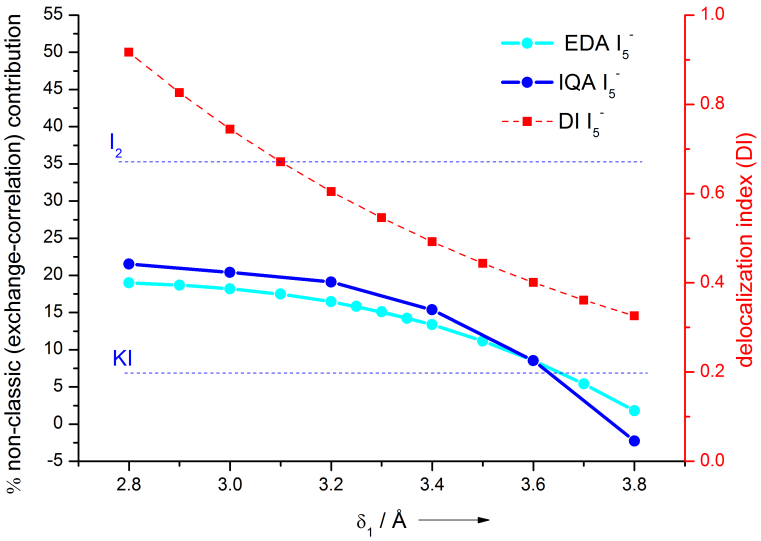


Figure 7.4 – Bonding contributions in I_5^- as a function of δ_1; the delocalization index is given in red, and the non-classic contributions to the binding energy (IQA or EDA) are shown in blue. The ionic KI and the covalent I_2 are shown as benchmarks. The reference states for the decomposition calculations are: I^-+K^+ → KI, $I^\cdot$+$I^\cdot$ → I_2, I_2+I_3^- → I_5^-. IQA and EDA decomposition analysis for I_3^- is shown in Figure A.6 of the Supporting Information.

Figure 7.5 highlights the behavior of the total energy density at the I_3^--I_2 BCP, proposed as a criterion to identify a covalent bond [234]: A negative value implies that the potential energy locally dominates over the kinetic energy and thus that the electron density contributes to

stabilization. Because of the homeomorphism between electron density and potential energy density, [235] a negative energy density at BCP implies that the energy density is negative along the whole path interconnecting the valence shells of the interacting atoms. As shown in Figure 7.5, the energy density at the BCP turns positive for $\delta_1 > 3.4$Å. For the experimental structures at different pressures, the trend mimics I_5^- in vacuo both for d_1 (forming I_7^-) and d_2 (forming the extended chain). Thus, as a weak I···I contact shrinks, the shared electron density increases, as also shown by the delocalization index in I_5^- (Figure 7.4), and becomes more stabilizing. It should be noted that while the delocalization approaches one electron pair for I–I distances close to that in I_2, the energy decomposition analysis assigns the dominant role to the electrostatic term anyway. Even in I_2 it exceeds 60 % of the total bond energy, in keeping with calculations for other homopolar diatomic molecules. [236]

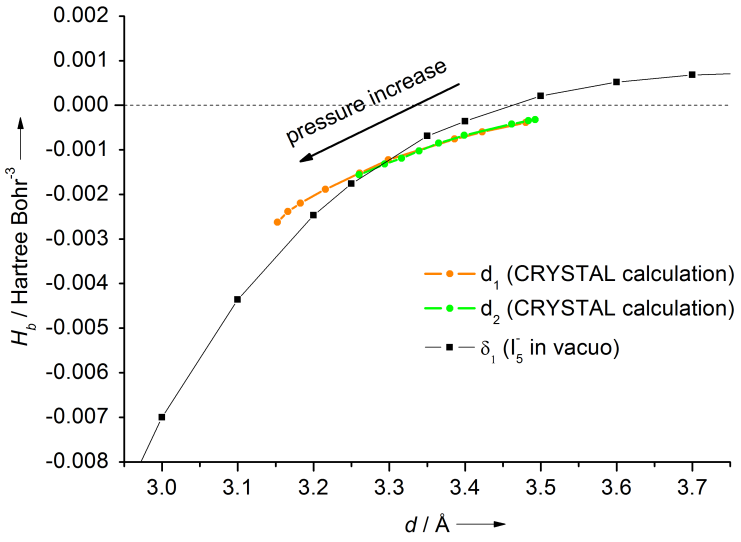


Figure 7.5 – Green and orange: Energy density at the bond critical points (BCPs) of the intermolecular contacts d_1 and d_2 calculated using the experimental structures. Black: Energy density at the BCP for the contact between I_3^- and I_2 of I_5^- in vacuo calculated at different distances δ_1.

In TEAI, the pressure-induced addition reactions apparently occur in a continuous way, without changing the electronic states of the constituting moieties. However, the formation of an extended polymer raises the question as to whether this may affect the transport properties of the compound. For this reason, we carried out measurements of the electric resistance in TEAI and observed a dramatic drop between measurements at 9.4 and 11.1GPa. This semiconductive state is established for $d_1 \approx 3.18$ Å and $d_2 \approx 3.32$ Å. In this range of distances, the percentage of covalency contribution to the bond has reached about 20 %, and upon further shortening, the increase is small (Figure 7.4).
At variance from X-ray diffraction experiments, the resistivity was measured in nano-alumina powder, a non-hydrostatic PTM. Nevertheless, powder diffraction measurements performed

under the same conditions show that TEAI behaves in nano-alumina as in a hydrostatic medium and therefore the structure–property correlation (although not determined in situ) is reliable. The van der Pauw method [237] was employed to estimate the resistivity of TEAI, which is 1.36 Ωcm at the highest pressure. Temperature scans at 25-55°C (see Figure A.7) showed that the variation of the electrical conductivity (inversely proportional to the resistivity) in TEAI under high pressure obeys the usual semiconductor equation, $\sigma = Ae^{-E_g/2kT}$ where E_g is the band gap and A is a constant. Calculations confirmed the occurrence of a continuous band gap closure upon compression, in accordance with the measured decrease in resistivity.

As reported above, the semiconducting behavior is triggered at I-I distances for which the calculated covalency contribution to the bond is 20 % (Figure 7.4), which does not grow significantly upon further compression. On the other hand, the electron sharing continuously grows as the I-I distance decreases, which causes the valence band broadening and consequent band gap reduction, which seems to be continuous along the compression path. It should be noted that only the gradual changes induced by pressure enable us to appreciate the structure-property correlation in such detail.

Recently, it was shown that PIs also exhibit photoconductive properties characteristic for CT complexes. [212] This phenomenon, however, was not detected in TEAI under high pressure, neither upon irradiation with a 1 mW 532 nm laser, nor with a 250 W tungsten lamp.

In conclusion, we have studied the effects of applying pressure on a solid-state polyiodide. The compression stimulates an approach between complementary I_3^- and I_2 units, which are clearly separated and interacting mainly electrostatically at ambient pressure, whereas the covalency contribution increases upon compression, as revealed by theoretical calculations. Ultimately, this leads to the formation of CT chains, which explains the drastically increased conductivity.

These features make TEAI a tunable pressure-sensitive electric switch. Structural studies at high pressure can rationalize the synthesis and search for future organic and hybrid semiconductors based on PIs. This includes structures with soft I···I contacts, which are the most prone for charge transport across the crystal under increased pressure. The results of this study indicate that solid PIs may be used as solid electrolytes in dye-sensitized solar cells, which eliminates the need for organic gelators [238] and ionic liquids in general.

Experimental Section

TEAI was synthesized according to literature procedures. [218]

Single-crystal diffraction measurements: A prismatic crystal of TEAI of 200×100×30 μm^3 was loaded in a Merrill–Bassett DAC, equipped with 0.5 mm diamond and steel gaskets. Nitrogen was used as a pressure-transmitting medium, and the pressure was calibrated with ruby fluorescence. [239] X-ray diffraction was measured at the Material Science Beamline at Swiss Light Source (λ = 0.49471 Å). [240] Structures were refined with Shelxl [241] incorporated in OLEX2. [232] All non-iodine atoms were refined isotropically. Hydrogen atoms attached to carbon atoms were geometrically placed and refined using a riding model. CCDC 1890310, 1890311, 1890312, 1890313, 1890314, 1890315, 1890316, 1890317, and 1890330 (HP structures,

low-temperature structure) contain the supplementary crystallographic data for this paper. These data can be obtained free of charge from The Cambridge Crystallographic Data Centre. Electrical conductivity measurements: Two crystals of TEAI of 60×60 ×80 μm^3 were loaded subsequently in DAC. A mixture of α-nano-alumina and epoxy glue 4:1 was used as the pressure-transmitting medium. The pressure was calibrated by the ruby fluorescence method. Four leads of golden wire were pressed against the sample (Figure A.1), and the electrical resistivity was measured using the four-probe method with a Keithley 2450 multimeter.
CRYSTAL14 [23, 242] was used for geometry optimizations and single-point calculations in the crystal, using the B3LYP functional and Grimme dispersion correction [243] (for all atoms except iodine). The basis sets used were of POB_TZVP quality. [244, 245] No imaginary frequency was calculated. Properties in the crystal were obtained with TOPOND. Gaussian16 [183] was used for geometry optimizations and single-point calculations in vacuo at the B3LYP/6–311G* level of theory. [246] AIMAll 14.11.23 [32] was used for analysis of the electron density descriptors and IQA. ADF [40] with B3LYP/TZP, without frozen core, and including scalar relativity was used for EDA.

Acknowledgements We thank Dr. Michał Andrzejewski for support during the electrical measurements and the Frankfurt project of the DFG research unit FOR 2125 CarboPaT (Project WI1232) and the Swiss National Science Foundation (Project Nr. 162861 and NCCR Marvel) for financial support.

7.2 Addendum

After the studies on TEAI, another polyiodide, N-propylurotropinium heptaiodide ($UrPrI_7$), was examined. Its structure was determined at six pressure points up to 4.02 GPa with X-ray diffraction measurements. Furthermore, geometry optimizations at the same pressure points were made and the band gap was calculated, leading to interesting insight in the general structure-property relationship of polyiodides.

Structural aspects of $UrPrI_7$

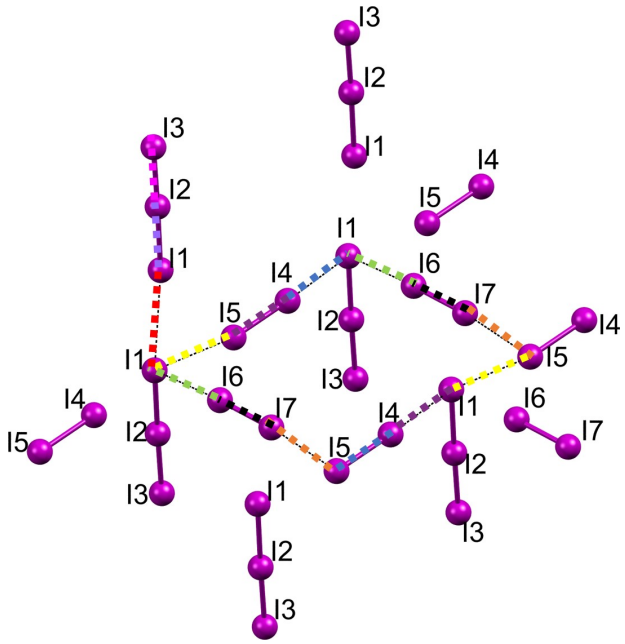


Figure 7.6 – I···I contacts in $UrPrI_7$ with color code. The cation is omitted for clarity.

Like TEAI, this polyiodide consists of I_2 interacting with negatively charged iodide units, but in contrast to TEAI, several I_2 are interacting with the same iodine atom I1. The three I_2 units, which can be identified at ambient pressure are I2-I3, I4-I5, and I6-I7 (see Figure 7.6). The evolution of the intramolecular I_2 distances with pressure is plotted in Figure 7.7. The I_2 units do not shorten a lot, or they even elongate with increasing pressure.

The diiodines approach I1, such that contacts I1-I4, I1-I5 shorten, while I1-I2 remains more or less constant because it was already relatively short at ambient pressure. The experimental points in Figure 7.8 demonstrate that especially the contact I1-I4 shortens a lot. As a consequence of that, the contacts I1-I2 and I1-I4 are both around 3.0 Å long at the highest measured pressure, and therefore a kind of pentaiodide is produced, similar to what happens in TEAI. However, in contrast to TEAI, this pentaiodide becomes very symmetric with increasing pressure. This can be observed in Figure 7.9 showing the same as Figure 7.3 but with superimposed data from $UrPrI_7$. The grey triangles representing $UrPrI_7$ tend towards the diagonal and thus

towards a symmetric pentaiodide.
Both distances I1-I2 and I1-I4 are shorter than d_1 in TEAI (orange line in Figure 7.2b). Even at 12.6 GPa, d_1 is still longer than the two distances in $UrPrI_7$ at 4.02 GPa.
Thus, $UrPrI_7$ forms a more symmetric pentaiodide with shorter bond lengths than it is the case in TEAI. However, while d_2, which is the distance of the other I_2 unit to I_3^- in TEAI, also shortens considerably with pressure leading to continuous iodine chains, the geometry in $UrPrI_7$ is different and continuous chains with bond lengths below 3.4 Å are not observed. This aspect will be discussed again together with the analysis of the band gap.

Figure 7.9 also illustrates that at ambient pressure both TEAI and $UrPrI_7$ are positioned among the white dots, which represent entries in the Cambridge Structural Database (CSD) for $I_3^- \cdots I_2$ contacts. Both compounds diverge from this trend with increasing pressure because the contacts are compressed. Due to this compression geometries are obtained, which are different from the ambient pressure structures representing the clear majority in the CSD.

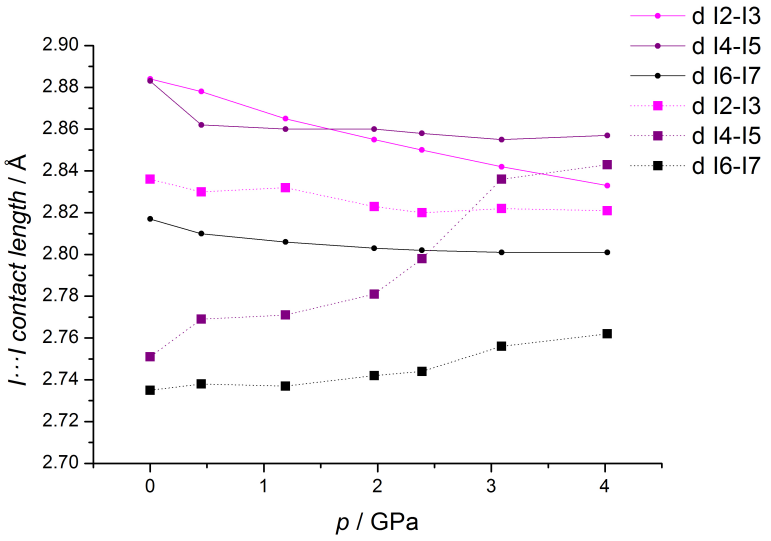


Figure 7.7 – Intramolecular I_2 distances plotted with respect to pressure. Squares represent experimental data points and and circles refer to geometry optimized structures.

To finish the structural analysis, it is to be mentioned that the distance I1-I1, which is one of the two contacts determining the interlayer spacing, shortens considerably. Figure 7.6 shows that the I_3 unit consisting of I1-I2-I3 is establishing the connection between two layers of iodine. The other such contact is I3-I6.
At 4.02 GPa, the highest measured pressure, the iodine atom I1 forms five interactions, whereof two are around 3.0 Å, one is 3.3 Å and two are around 3.45 Å in the experimental structure. Therefore, it can be considered almost pentavalent.

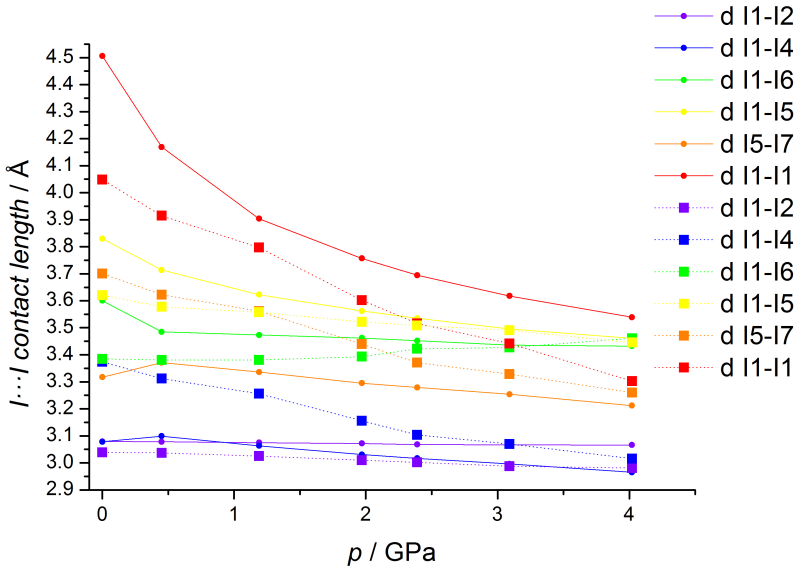


Figure 7.8 – Intermolecular distances. Squares represent experimental data points and circles refer to geometry optimized structures.

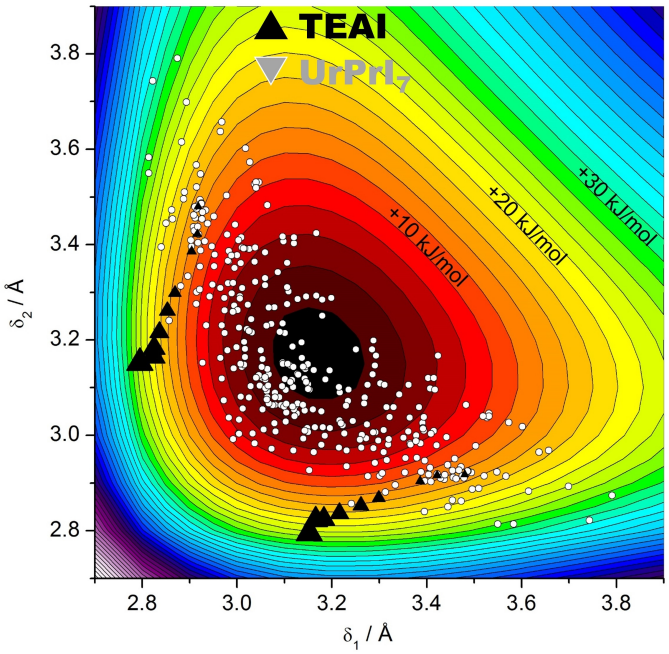


Figure 7.9 – The colors indicate the destabilization from the symmetric I_5^- obtained through calculation *in vacuo*. The white circles represent entries in the CSD containing I_5^-. The black triangles represent TEAI and the grey triangles UrPrI_7 at different pressures (the triangle size is proportional to the pressure). For TEAI, d_1 is plotted against the intramolecular I_3^- length. For UrPrI_7, I1-I2 is plotted against I1-I4.

Band gap and correlation with structure

Figure 7.10 shows the band gap calculated for the geometry optimized structures of $UrPrI_7$. The level of theory is the same as used for TEAI so that the results can directly be compared. The band gap is considerably lower than in TEAI. When considering that TEAI becomes semiconductive around 10 GPa, corresponding to a band gap of around 2.3 eV, one could expect that $UrPrI_7$ is semiconductive already at ambient pressure, where the band gap is already lower than 2.3 eV.

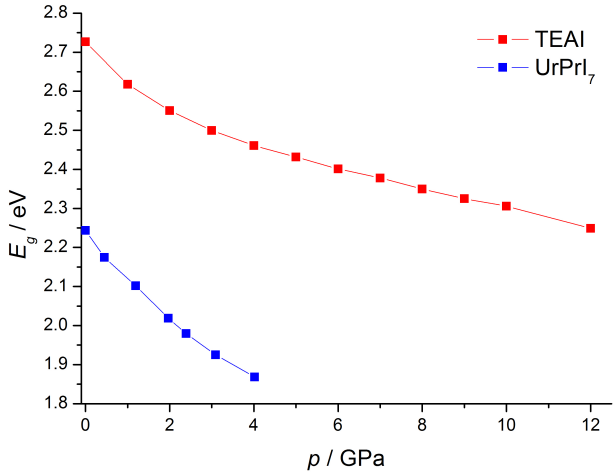


Figure 7.10 – Band gap of TEAI and $UrPrI_7$ for the geometry optimized structures.

In TEAI contacts d_1 and d_2 are needed to form a continuous chain. In $UrPrI_7$ there are several possible pathways, the shortest one being through I1-I4-I5-I1 (violet, blue, yellow in Figure 7.6). While I1-I4 and I4-I5 are below 3.4 Å at 0 GPa and below 3.0 Å at 4.02 GPa, this is not the case for I1-I5. In fact, I1-I5 of $UrPrI_7$ is considerably longer than d_2 of TEAI at ambient pressure (always referring to geometry optimized structure). Therefore, one could expect that $UrPrI_7$ is less conductive and has a larger band gap. This is obviously not the case. Likely the reason is that in $UrPrI_7$ there are many different possible pathways to form an iodine chain. Although each individual one of them has at least one bond length, which is longer than in TEAI, the sum of all these pathways causes a smaller band gap.

Therefore, the band gap cannot be correlated one to one with the bond distances or the bond character like for example its covalency. This is confirmed by I_2, where the shortest distances between different I_2 units in the crystal are 3.5 Å long [247]. Nevertheless, the crystal has a band gap of 1.3 eV [248], which is much lower than in TEAI and $UrPrI_7$.
Based on this reasoning, it can be deduced that the I···I distance and the bond character are important criteria for the conductivity in one particular structure, but when comparing several compounds, the number of possible options for a polymeric chain is equally important.

Unfortunately, it is very challenging to measure the conductivity of $UrPrI_7$ since it decomposes very quickly by loosing iodine, and then transforms into a different structure. In general, the measurement of polyiodides, be it with X-rays for diffraction or with electrodes for conductivity, is difficult because polyiodides are often not very stable. This issue highlights the necessity for a theoretical understanding of their bonding characteristics in order to predict properties where they cannot easily be measured.

Yet, also this theoretical analysis in not trivial, since, as discussed above, a certain I···I distance does not simply correspond to a given conductivity. Therefore, establishing a clear structure-property relationship in polyiodides is difficult. It is evident that a continuous polyiodide chain can be conductive due to the hypervalency of iodine and the sharing of electrons between atoms. But it is also clear from a search in the CSD that, at ambient pressure, polyiodides usually consist of I^-, I_2 and I_3^- units, which are farther apart from each other than a covalent I-I bond, whereas longer polyiodides are rare. And therefore, most of the compounds are not conductive. High pressure studies are a very powerful tool to shed light on the relationship between I···I interaction lengths and conductivity. They allow to analyse various I···I distances and properties for the same compound and to observe the transition from a structure with clearly separated units to a continuous polyiodide chain as it is the case in TEAI.

In both examined polyiodides, the structural changes with increasing pressure seem continuous. The intermolecular I···I contacts shorten, and well separated units are gradually transformed into polymers. The band gap is continuously decreasing with increasing pressure, and all the examined bond properties behave like that as well. For TEAI, it was demonstrated that the bonds transform from dominantly ionic ones into more covalent ones. All these indicators reveal that there is not one specific point where molecules stop to exist and a polymer is formed. However, the experimental measurement of the resistivity (inverse of conductivity) pointed out a more abrupt change, suggesting that TEAI undergoes an electronic phase transition. Since the calculated results do not support any non-continuous behaviour that would explain the sudden drop of resistivity, more experiments are needed, especially also on other compounds.

8 Different Aspects of X-ray Constrained Wavefunctions

This chapter is about X-ray constrained wavefunction calculations (XCW), thereof derived energies and electron densities, and their ability to include crystal field effects under various conditions. Chapter 9, instead, will focus on molecular orbitals obtained from X-ray constrained wavefunction calculations. Section 8.1 deals with the two opposed functions minimized in the XCW procedure (total energy and χ^2) and how to transfer the latter to have units of energy by using the self-electrostatic energy of molecules. Section 8.2 provides a more in depth analysis of the self-electrostatic energy calculated from structure factors and from a wavefunction. The matter of section 8.3 is the ability of X-ray constrained WFNs to include crystal field effects. Finally, section 8.4 is about the use of X-ray constrained wavefunctions for frozen density embedding.

8.1 X-ray constrained wavefunction fitting using the self-electrostatic energy

8.1.1 Theoretical background

In a conventional wavefunction calculation the minimized quantity is the electronic energy. In a crystallographic least-squares refinement the minimized quantity is the difference of the calculated structure factors F_k^{calc} to the targeted structure factors. These are usually experimentally measured ones but can be also simulated. Since this chapter treats basic and fundamental questions rather than applications, the structure factors were simulated and therefore represent ideal conditions. They will subsequently be abbreviated $F_k^{"exp"}$. In an XCW calculation both terms, energy E and structure factor difference χ^2, are minimized simultaneously (equation 8.1). However, these two quantities have different units and thus the weighting factor λ is not dimensionless. In fact, λ has units of energy but it depends on the volume and thus its meaning changes when the volume of the studied molecules changes.

$$J = E + \lambda\chi^2 \tag{8.1}$$

In the *traditional* XCW formula χ^2 is defined as:

$$\chi^2 = \frac{1}{N_r - N_p} \sum_{k}^{N_r} \frac{(F_k^{calc} - F_k^{exp})^2}{\sigma_k^2} \tag{8.2}$$

σ_k are the standard uncertainties of the measured structure factors, N_r is the number of measured reflections and N_p is the number of parameters of the model. In practice this value is often set equal 1, assuming that the scale factor is the only freely refined parameter while the wavefunction parameters themselves are not considered.
If the two parts of equation 8.1 shall have the same unit, equation 8.2 has to be modified such that χ^2 also has units of energy.

The self-electrostatic energy corresponds to the electrostatic interaction of a charge density distribution with itself. When it is expressed based on the structure factors, the formula is: [249]

$$E_{elstat-self} = \frac{1}{2\pi V_m} \sum_{\mathbf{H}} \frac{|F_{\mathbf{H}}|^2}{\mathbf{H}^2} \tag{8.3}$$

with $\mathbf{H} = |2 * \sin(\theta)/\lambda|$ and V_m the molecular volume.
Thus, if χ^2 shall be an expression with units of energy, it can be expressed with the self-electrostatic energy:

$$\chi_E^2 = \frac{1}{N_r - N_p} \frac{1}{2\pi V_m} \sum_{H} \frac{(F_H^{calc} - F_H^{exp})^2}{\sigma_H^2 |\mathbf{H}|^2} \tag{8.4}$$

And therefore the *new* XCW formula becomes:

$$J = E + \lambda' \chi_E^2 \tag{8.5}$$

λ is primed because it is different from λ in the traditional formula. The implementation of the new formula in *Tonto* is described in appendix B.

The aim of this section is to test two things: (**I**) whether using χ_E instead of χ leads to better results and (**II**) how the self-electrostatic energy, the electronic energy of the X-ray constrained wavfunction, and their differences to the reference value develop upon varying λ or λ'. The reference is an initial wavefunction used to simulate the structure factors.
Thus, (**I**) should answer the question how the use of the modified formula changes crystallographic agreement values (e.g. the R-value), the electron density (e.g. the RSR value), and the electronic energy; especially when comparing the results from the traditional and new formula obtained with corresponding λ-value.
The real space R (RSR) value is used to quantify the similarity of two electron densities and is calculated on a grid of n_p points ([64, 250]):

$$RSR(\rho_1, \rho_2) = \frac{\sum_{i=1}^{n_p} | \rho_1(\mathbf{r}_i) - \rho_2(\mathbf{r}_i) |}{\sum_{i=1}^{n_p} | \rho_1(\mathbf{r}_i) + \rho_2(\mathbf{r}_i) |} \tag{8.6}$$

(**II**) deals with the self-electrostatic energy, which can be calculated using the structure factors according to equation 8.3 and the electronic energy as obtained from *Tonto*. The self-electrostatic energy can be compared for the calculated and the simulated structure factors $F^{"exp"}$. The question is how these energies and their difference to the reference value behave. In principle there could even be an optimum value of λ, considering that the electronic energy (going away from the reference value with increasing λ) and the self-electrostatic energy (approaching the value of the simulated structure factors with increasing λ) behave oppositely.

8.1.2 Calculation details

For the three molecules thioacetone ($(CH_3)_2CS$), alanine ($C_3H_7NO_2$), and hydrogen fluoride initial wavefunctions were calculated at PBE/cc-pVTZ(6d) level of theory using *Gaussian*. The so-obtained wavefunctions are the references and were used to simulate structure factors with the structure factor analysis module of *XD* [65] (*XDsfra*). Structure factors were simulated for the full sphere up to resolution 1.0 $Å^{-1}$ and a 10×10×10 $Å^3$ unit cell. This choice is not unproblematic and section 8.2 presents some more details about it. Subsequently, the simulated structure factors were used in *Tonto* as "experimental" structure factors and the σ_k was set to 1 for all structure factors. XCW fittings with the conventional and the new formula were done at different values of λ or λ', respectively. The calculations were of Hartree-Fock type with the 6-311++G(2d,2p) basis set. The geometry and the unit cell were obviously the same as the one for the structure factor simulation and the scale factor was optimized.

Due to the additional factor $2\pi V$ as well as the factor 2 (squared) of **H** in the new formula (equation 8.4) compared to the traditional one (equation 8.2), the corresponding λ' value is:

$$\lambda' = 2\pi V * 4 * \lambda \tag{8.7}$$

With this formula the value $\lambda = 1$ becomes $\lambda' = 169604$ for a volume of 10×10×10 $Å^3$ and transformed into atomic units. Although λ is depending on the volume and cannot have an absolute meaning when comparing different structures, the correct transformation from λ to λ' matters when comparing the performance of the traditional and new formula.

Analysis of the results was done for the electronic energy, the self-electrostatic energy, the R-factor (how good is the fit in reciprocal space?), and the RSR-value (how good is the fit in real space?). In addition, the difference between the self-electrostatic energies obtained with simulated structure factors and with structure factors obtained during the XCW calculation can be analysed, both in reciprocal and real space. The values presented refer to the difference in real space: $E^{\lambda=x}_{elstat-self} - E^{ref}_{elstat-self}$.

8.1.3 Results

Tables 8.1-8.6 show an overview of the results for the three examined compounds. All energies are given in atomic units. The numbers in Tables 8.1, 8.3, and 8.5 were obtained with the

traditional XCW formula while the numbers in Tables 8.2, 8.4, and 8.6 were obtained with the new formula. In the first row, the electronic energy (E_{el}) plus the constant nucleus-nucleus potential (V_{NN}) resulting from the XCW calculation is given as printed in the *Tonto* output. This energy corresponds to the first term in equation 8.1 or 8.5.
In the second row, the self-electrostatic energy is shown calculated according to equation 8.3. There are results for different λ values from 0 to 100 and in addition there is the column 'ref'. In this column the electronic energy of the reference DFT wavefunction used to simulate the structure factors is written as well as the self-electrostatic energy of these simulated structure factors.
The third row gives the RSR value calculated according to equation 8.6, an agreement measure in direct space and the fourth row the R-factor as printed by *Tonto*, an agreement value in reciprocal space. When increasing λ or λ' these agreement factors decrease designating a better agreement.
The subsequent two rows in the tables show the difference of the XCW electronic energy and self-electrostatic energy to the one obtained with $\lambda = 0$. This indicates how far the energy is away from the one of an unconstrained calculation.
Finally, the last row shows the sum of the absolute difference of the XCW electronic energy ($E_{el}^{\lambda=x}$) to the DFT electronic energy (E_{el}^{ref}) and the difference of the self-electrostatic energy obtained with the calculated structure factors ($E_{elstat-self}^{\lambda=x}$) to the one obtained with simulated structure factors ($E_{elstat-self}^{ref}$):
$\left|E_{el}^{\lambda=x} - E_{el}^{ref}\right| + \left|E_{elstat-self}^{\lambda=x} - E_{elstat-self}^{ref}\right|$.

Table 8.1 – Thioacetone $(CH_3)_2CS$. XCW calculation done with traditional formula

	λ-value				
	ref	0	0.01	0.1	0.5
Electronic energy + V_{NN}	-515.8236	-514.6495	-514.6495	-514.6495	-514.6495
Self-electrostatatic energy	249.6786	248.2052	248.2059	248.2120	248.2388
RSR $*10^{-3}$		10.7161	10.71372	10.6927	10.6009
R-factor		0.00282	0.00282	0.00281	0.00277
$E_{el}^{\lambda=x}$ - $E_{el}^{\lambda=0}$		0.0000	0.0000	0.0000	0.0000
$E_{elstat-self}^{\lambda=x}$ - $E_{elstat-self}^{\lambda=0}$		0.0000	0.0007	-0.0068	-0.0336
sum of differences		2.6476	2.6469	2.6407	2.6140
λ-value	1	3	5	10	100
Electronic energy + V_{NN}	-514.6494	-514.6493	-514.6491	-514.6484	-514.6381
Self-electrostatatic energy	248.2708	248.3858	248.4830	248.6708	249.3753
RSR $*10^{-3}$	10.082	9.728	9.006	5.247	3.396
R-factor	0.00273	0.00256	0.00243	0.00216	0.00093
$E_{el}^{\lambda=x}$ - $E_{el}^{\lambda=0}$	0.0000	-0.0001	-0.0004	-0.0011	-0.0114
$E_{elstat-self}^{\lambda=x}$ - $E_{elstat-self}^{\lambda=0}$	-0.1806	-0.2778	-0.4657	-1.1701	-1.3663
sum of differences	2.4671	2.3701	2.1830	1.4888	1.3023

Table 8.2 – Thioacetone $(CH_3)_2CS$. XCW calculation done with new formula

	λ'-equivalent value (actual value / 169604)				
	ref	0	0.01	0.1	0.5
Electronic energy + V_{NN}	-515.8236	-514.6495	-514.6495	-514.6494	-514.6488
Self-electrostatatic energy	249.6786	248.2052	248.2131	248.2791	248.4955
RSR $*10^{-3}$		10.7161	10.6919	10.4886	9.8086
R-factor		0.002818	0.00281	0.002761	0.002578
$E_{el}^{\lambda=x}$ - $E_{el}^{\lambda=0}$		0.0000	0.0000	0.0000	-0.0007
$E_{elstat-self}^{\lambda'=x}$ - $E_{elstat-self}^{\lambda'=0}$		0.0000	-0.0079	-0.0739	-0.2903
sum of differences		2.6476	2.6397	2.5737	2.3580
λ-value	1	3	5	10	100
Electronic energy + V_{NN}	-514.648	-514.6434	-514.6401	-514.6340	does not
Self-electrostatatic energy	248.6637	248.9562	249.0680	249.1778	converge
RSR $*10^{-3}$	9.2517	8.1402	7.6211	7.0274	
R-factor XCW	0.00241	0.00203	0.00181	0.00152	
$E_{el}^{\lambda'=x}$ - $E_{el}^{\lambda'=0}$	-0.0018	-0.0061	-0.0094	-0.0154	
$E_{elstat-self}^{\lambda'=x}$ - $E_{elstat-self}^{\lambda'=0}$	-0.4586	-0.7510	-0.8628	-0.9726	
sum of differences	2.1908	1.9026	1.7941	1.6904	

Table 8.3 – Alanine. XCW calculation done with traditional formula

	λ-value				
	ref	0	0.01	0.1	0.5
Electronic energy + V_{NN}	-323.4982	-321.9707	-321.9707	-321.9707	-321.9707
Self-electrostatatic energy	228.2702	228.8772	228.8769	228.8744	228.8634
RSR $*10^{-3}$		11.7578	11.7553	11.7326	11.6332
R-factor		0.00682	0.00682	0.00679	0.00672
$E_{el}^{\lambda=x}$ - $E_{el}^{\lambda=0}$		0.0000	0.0000	0.0000	0.0000
$E_{elstat-self}^{\lambda=x}$ - $E_{elstat-self}^{\lambda=0}$		0.0000	0.0003	0.0028	0.0138
sum of differences		2.1345	2.1342	2.1317	2.1208
λ-value	1	3	5	10	100
Electronic energy + V_{NN}	-321.9706	-321.9704	-321.9701	-321.9688	-321.9488
Self-electrostatatic energy	228.8501	228.8005	228.7563	228.6642	228.2248
RSR $*10^{-3}$	11.5126	11.0675	10.6724	9.8501	5.4439
R-factor	0.00662	0.00627	0.00596	0.00532	0.00210
$E_{el}^{\lambda=x}$ - $E_{el}^{\lambda=0}$	0.0000	-0.0002	-0.0006	-0.0019	-0.0219
$E_{elstat-self}^{\lambda=x}$ - $E_{elstat-self}^{\lambda=0}$	0.0271	0.0767	0.1209	0.2130	0.6524
sum of differences	2.1074	2.0581	2.0142	1.9234	1.5948

Table 8.4 – Alanine. XCW calculation done with new formula

	λ'-equivalent value (actual value / 169604)				
	ref	0	0.01	0.1	0.5
Electronic energy + V_{NN}	-323.4982	-321.9707	-321.9707	-321.9707	-321.9702
Self-electrostatatic energy	228.2702	228.8772	228.8769	228.8744	228.8634
RSR $*10^{-3}$		11.7578	11.7378	11.5648	10.9212
R-factor		0.00682	0.00681	0.00675	0.00648
$E_{el}^{\lambda=x}$ - $E_{el}^{\lambda=0}$		0.0000	0.0000	0.0000	0.0000
$E_{elstat-self}^{\lambda=x}$ - $E_{elstat-self}^{\lambda=0}$		0.0000	0.0012	0.0133	0.0777
sum of differences		2.1345	2.1342	2.1317	2.1208
λ'-equivalent value	1	3	5	10	100
Electronic energy + V_{NN}	-321.9694	-321.9654	-321.9618	-321.9549	does not
Self-electrostatatic energy	228.7192	228.5009	228.3855	228.2530	converge
RSR $*10^{-3}$	10.2981	8.6856	7.7139	6.3460	
R-factor	0.00616	0.00514	0.00445	0.00340	
$E_{el}^{\lambda=x}$ - $E_{el}^{\lambda=0}$	-0.0013	-0.0053	-0.0089	-0.0158	
$E_{elstat-self}^{\lambda=x}$ - $E_{elstat-self}^{\lambda=0}$	0.1580	0.3763	0.4917	0.6242	
sum of differences	1.9779	1.7636	1.6517	1.5605	

Table 8.5 – Hydrogen fluoride. XCW calculation done with traditional formula

	λ-value				
	ref	0	0.01	0.1	0.5
Electronic energy + V_{NN}	-100.3840	-100.0541	-100.0541	-100.0541	-100.0541
Self-electrostatatic energy	44.4676	44.8516	44.8514	44.8500	44.8437
RSR $*10^{-3}$		9.3004	9.2985	9.2828	9.2143
R-factor		0.00499	0.00499	0.00497	0.00491
$E_{el}^{\lambda=x}$ - $E_{el}^{\lambda=0}$		0.0000	0.0000	0.0000	0.0000
$E_{elstat-self}^{\lambda=x}$ - $E_{elstat-self}^{\lambda=0}$		0.0000	0.0002	0.0016	0.0078
sum of differences		0.7139	0.7137	0.7123	0.7060
λ-value	1	3	5	10	100
Electronic energy + V_{NN}	-100.0541	-100.0541	-100.0540	-100.0538	-100.0509
Self-electrostatatic energy	44.8361	44.8083	44.7837	44.7336	44.4986
RSR $*10^{-3}$	9.1325	8.8383	8.5879	8.0977	5.9646
R-factor	0.00484	0.00458	0.00437	0.00394	0.00208
$E_{el}^{\lambda=x}$ - $E_{el}^{\lambda=0}$	0.0000	-0.0001	-0.0001	-0.0004	-0.0033
$E_{elstat-self}^{\lambda=x}$ - $E_{elstat-self}^{\lambda=0}$	-0.1806	-0.2778	-0.4657	-1.1701	-1.3663
sum of differences	0.6985	0.6706	0.6462	0.5962	0.3642

Table 8.6 – Hydrogen fluoride. XCW calculation done with new formula

	λ'-equivalent value (actual value / 169604)				
	ref	0	0.01	0.1	0.5
Electronic energy + V_{NN}	-100.0541	-100.0541	-100.0541	-100.0539	-321.9702
Self-electrostatatic energy	44.8516	44.8510	44.8459	44.8229	228.8634
RSR $*10^{-3}$		9.3004	9.2845	9.1547	8.7928
R-factor		0.00499	0.00498	0.00492	0.00471
$E_{el}^{\lambda=x}$ - $E_{el}^{\lambda=0}$		0.0000	0.0000	0.0000	-0.0002
$E_{elstat-self}^{\lambda=x}$ - $E_{elstat-self}^{\lambda=0}$		0.0000	0.0006	0.0057	0.0287
sum of differences		0.7139	0.7133	0.7082	0.6854
λ'-equivalent value	1	3	5	10	100
Electronic energy + V_{NN}	-100.0536	-100.0522	-100.0515	-100.0508	-100.0473
Self-electrostatatic energy	44.7954	44.7121	44.6613	44.5954	44.39
RSR $*10^{-3}$	8.6081	8.4846	8.4143	8.0597	4.8723
R-factor	0.00452	0.00406	0.00371	0.00323	0.00149
$E_{el}^{\lambda=x}$ - $E_{el}^{\lambda=0}$	-0.0006	-0.0019	-0.0026	-0.0033	-0.0069
$E_{elstat-self}^{\lambda=x}$ - $E_{elstat-self}^{\lambda=0}$	0.0561	0.1395	0.1902	0.2562	0.4077
sum of differences	0.6583	0.5763	0.5263	0.4610	0.3605

Figure 8.1 shows on the left side plots of the electronic energy E_{el} plus the constant V_{NN} term with respect to the λ value. In order to plot the results from the XCW calculations with both, the traditional and new formula in the same graph, the λ' values were divided by $2\pi V * 4 = 169604$ to bring them back on the same scale. The plots go from $\lambda = 0$ to $\lambda = 10$. The reference DFT energy E_{el}^{DFT} is always a lot (several hundredths of kilojoules per mole) below the lines of $E_{el}^{\lambda=x}$ and therefore not shown. However, it is to be noted that the DFT energy is calculated in a different way. Nevertheless, also the Hartree-Fock energy of the DFT wavefunction is still lower in energy than the X-ray constrained wavefunctions, because in the HF case only the energy is minimized and not also the structure factor agreement. On the right side are plots of the self-electrostatic energy $E_{elstat-self}$. The blue value corresponds to $E_{elstat-self}$ from the simulated structure factors $F^{"exp"}$.

8.1.4 Discussion

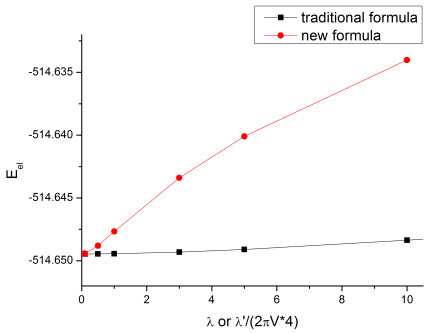
Part I - Comparison of results obtained with traditional and new formula

At λ=λ'=0 the results from the traditional and new formula are always identical since this calculation is a pure Hartree-Fock calculation without any constraints. Upon increasing λ (or λ') the electronic energy obviously increases due to the additional constraint and the consequently reduced flexibility in the variational procedure. For all three compounds thioacetone, alanine, and hydrogen fluoride the electronic energy increases faster with the new XCW formula using χ_E^2 than with the traditional one using χ^2. On the other hand, the self-electrostatic energy of the calculated structure factors approaches the one of the simulated structure factors faster with the new formula. As visible in Tables 8.1-8.6, also the R-value and the RSR value are lower, that is in better agreement with the simulated experiment, when using the new formula for comparable λ values. Therefore, the fitting with the new formula is better in terms of the self-electrostatic energy and the statistical agreement while the electronic energy increases faster. In other words, a lower λ'-equivalent value is needed to obtain a fitting, which is as good as one obtained with a higher λ value. This can also explain why at $\lambda = 100$ the calculations for alanine and thioacetone still converge, whereas they do not for λ'-equivalent = 100.

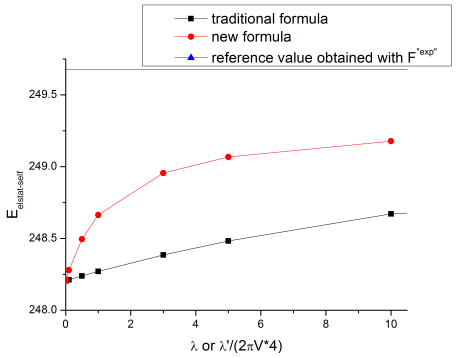
Part II - Analysis of electronic and self-electrostatic energy. Best λ-value

For thioacetone the self-electrostatic energy from the calculated structure factors (obtained during the XCW calculation) is lower than the self-electrostatic energy from the simulated structure factors as visible in Figure 8.1b. When increasing λ from 0 to higher values, both curves approach the reference value in blue. For alanine and hydrogen fluoride it is the other way around. In those cases the self-electrostatic energy from the calculated structure factors $F^{\lambda=x}$ is higher than from the simulated $F^{"exp"}$, and the red and black curves approach the blue line by going down.

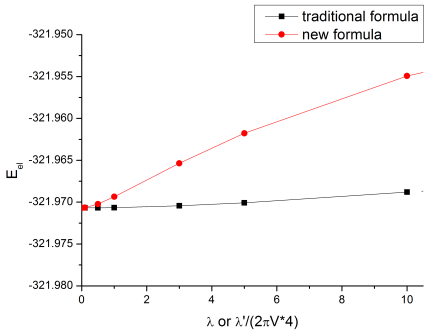
At a given value of λ (or λ'), the $E_{elstat-self}^{\lambda=x}$ curves crosses the blue line from the simulated



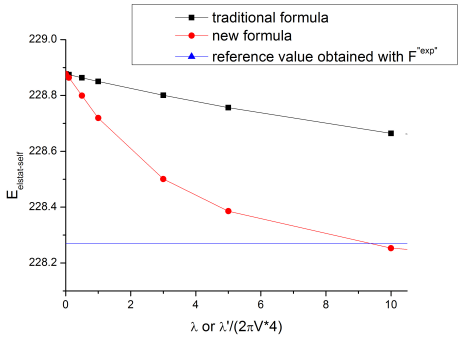
(a) Electronic energy of thioacetone



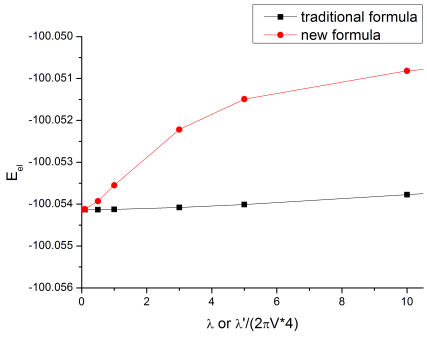
(b) Self-electrostatic energy of thioacetone



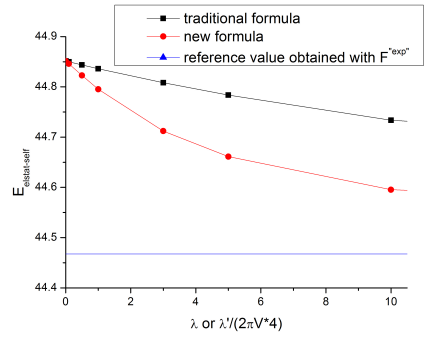
(c) Electronic energy of alanine



(d) Self-electrostatic energy of alanine



(e) Electronic energy of hydrogen fluoride

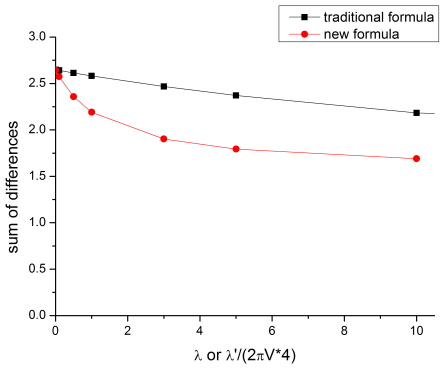


(f) Self-electrostatic energy of hydrogen fluoride

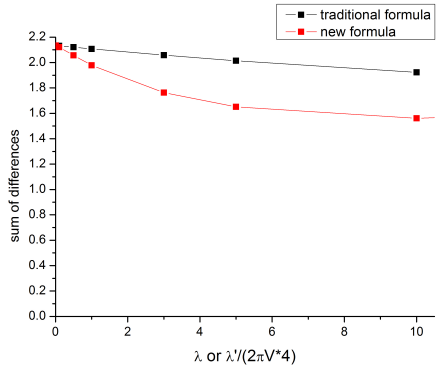
Figure 8.1 – Energy terms (in atomic units) with respect to λ and λ'-equivalent values.

experimental structure factors. For alanine and with the new formula this happens already before 10, for hydrogen fluoride with the new formula approximately around λ'-equivalent=50 and for thioacetone not until λ'-equivalent=100 but it also approaches it. After this point the individual structure factors are still improving as indicated by the decreasing R-value but the agreement of the overall self-electrostatic energy actually worsens. The difference of the self-electrostatic energy to its targeted value is a measure of the overall agreement (like it can also be done for the electron density). This means that at the point where the line crosses, an optimum fitting is reached and afterwards one might be overfitting. However, this happens at

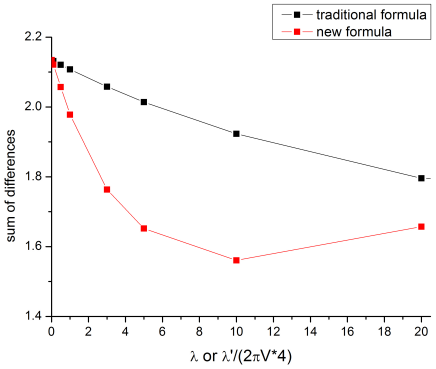
very high λ, too high to be reached experimentally.



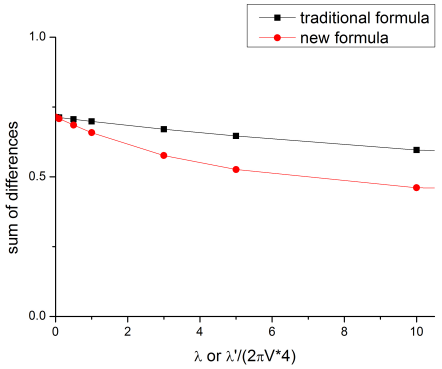
(a) Thioacetone



(b,I) Alanine



(b,II) Alanine



(c) Hydrogen fluoride

Figure 8.2 – Sum of absolute differences (in atomic units). Left: from λ=0 to 10, right: from λ=0 to 20. Only alanine is shown until 20 because in this case the increase of the red line can be observed. For the other two compounds this was not observed in the λ' range for which calculations can be brought to convergence.

For a fair comparison between traditional and new method, both absolute differences, the one of the electronic and the one of the self-electrostatic energy to the reference value have to be analysed at the same time. They represent the behaviour of both parts of equation 8.1, which develop in opposite directions. Therefore, the two differences were summed up and the results are shown in the last row of Tables 8.1-8.6 and in Figure 8.2. The summation of the absolute differences can only be done reasonably because they are both energies.

For all three compounds the sum of the absolute differences initially decrease with increasing λ. However, after the self-electrostatic energy crossed the reference value, the absolute difference increases again as visible in Figure 8.2b for alanine around λ'-equivalent=10. Therefore, the energy difference to the targeted value reaches a minimum and this is the best λ' value in terms of energy. This could serve as a new criterion for the choice of λ although it is practically not reachable since a calculation with real experimental data would not converge at such high λ-values.

8.2 Electronic energy from structure factors and from wavefunctions

The self-electrostatic energy $E_{elstat-self}$ is the interaction energy of the electrons with themselves. It can be calculated from the structure factors according to equation 8.3. [249]
A more detailed analysis of the self-electrostatic energy $E_{elstat-self}$, as done in this section, shows that large unit cells and high angular resolutions for the structure factors are needed to obtain a reasonable value. Whether a value is reasonable can be recognized by comparison to the electron-electron interaction energy V_{ee} from the corresponding wavefunction.

In order to do this comparison, the energy terms for two small molecules (difluorine F_2 and hydrogen fluoride) were examined. A reference wavefunction for both molecules was calculated at HF/6-311++G(2d,2p) level of theory. From this wavefunction a set of structure factors $F^{"exp"}$ was simulated for the full sphere and used for XCW calculations at the exact same level of theory.

Initially a constant resolution of 1 $Å^{-1}$ was used for the structure factor simulation and the size of the unit cell was varied. Results are shown in Tables 8.7 and 8.8.

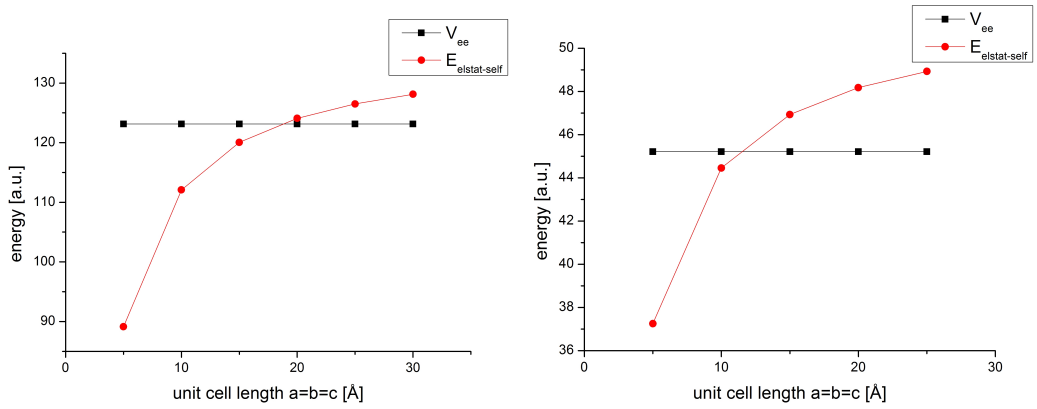
Table 8.7 – Electronic energies of F_2 obtained from direct and reciprocal space. The column on the left indicates the size of the unit cell used for the structure factor simulation. The energies are in atomic units.

a=b=c [Å]	$E_{elstat-self}$ from $F^{"exp"}$	$E_{elstat-self}$ from F^{calc}	V_{ee} from XCW
5	89.138	89.138	123.153
10	112.084	112.084	123.153
15	120.602	120.602	123.153
20	125.907	125.907	123.153
25	126.505	126.505	123.153
30	128.121	128.121	123.153
V_{ee} of the reference Hartree-Fock wavefunction is 123.153			

Table 8.8 – Electronic energies of hydrogen fluoride obtained from direct and reciprocal space. The column on the left indicates the size of the unit cell used for the structure factor simulation. The energies are in atomic units.

a=b=c [Å]	$E_{elstat-self}$ from $F^{"exp"}$	$E_{elstat-self}$ from F^{calc}	V_{ee} from XCW
5	37.248	37.248	45.215
10	44.456	44.456	45.215
15	46.933	46.933	45.215
20	48.786	48.786	45.215
25	48.925	48.925	45.215
V_{ee} of the reference Hartree-Fock wavefunction is 45.215			

V_{ee} from the X-ray constrained wavefunctions is constant because the reference wavefunction was calculated at Hartree-Fock level and therefore the XCW calculation does not need to fit any additional effect. For the same reason, the $E_{elstat-self}$ calculated with the simulated structure factors $F^{"exp"}$ and with the structure factors calculated from the XCW F^{calc} are identical.



(a) Difluorine F_2 (b) Hydrogen fluoride

Figure 8.3 – Electronic energy of the X-ray constrained wavefunction V_{ee} and self-electrostatic energy obtained from the calculated structure factors $E_{elstat-self}$ plotted against the size of the unit cell used for the structure factor simulation.

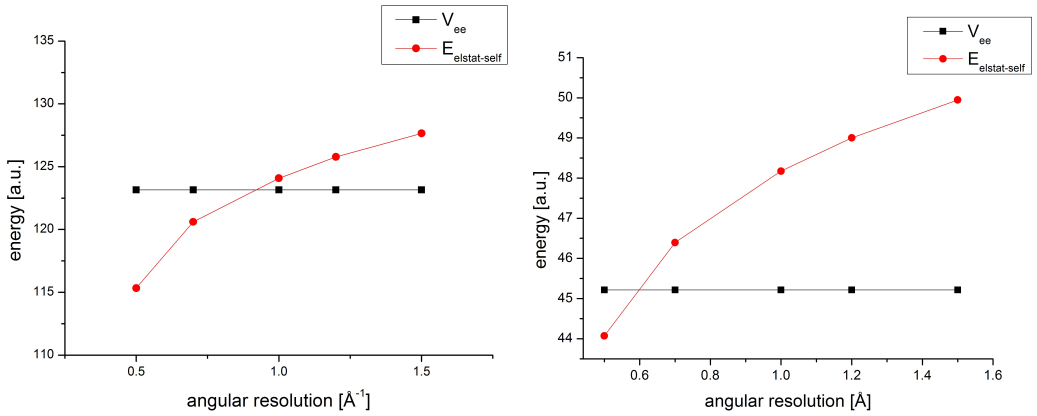
Figure 8.3 demonstrates that the self-electrostatic energy strongly depends on the size of the unit cell. In fact, $E_{elstat-self}$ is underestimated with small unit cells and approaches V_{ee} with increasing size of the unit cell. This can be explained by the fact that in a larger unit cell the electron-electron interaction energy of one molecule becomes more similar to the gas-phase value, which constitutes here the reference value. However, the curves rise so much that they overtakes the V_{ee} values. More analysis will be needed to explain this behaviour.

Equation 8.3 has a singularity for **H**=0, which is why this reflection was excluded from the summation. For non-polar unit cells this reflection gives no contribution to the electronic energy but for polar cells the correct formula would include a constant term considering the dipole moment D [249]:

$$E_{elstat-self} = \frac{1}{2\pi V_m} \sum_{\mathbf{H}} \frac{|F_{\mathbf{H}}|^2}{\mathbf{H}^2} + \frac{2\pi}{3V_m} D^2 \tag{8.8}$$

This constant term was not considered because it has no effect on the non-polar crystal with F_2 and for HF it is also just a constant term. Since, the same behaviour is observed in both cases, it can also not explain the difference between V_{ee} and $E_{elstat-self}$.

In Figure 8.4 the self-electrostatic energy for a constant unit cell of 20×20×20 Å^3 is plotted against varying resolution of the structure factor data set. Like when changing the unit cell size, $E_{elstat-self}$ initially approaches V_{ee} but then becomes even higher.



(a) Difluorine F_2 (b) Hydrogen fluoride

Figure 8.4 – Electronic energy V_{ee} of the X-ray constrained wavefunction and self-electrostatic energy $E_{elstat-self}$ obtained from the calculated structure factors plotted against the angular resolution of the structure factor data set.

After this more detailed analysis it becomes clear that the self-electrostatic energies vary a lot with the unit cell and the resolution used for the simulation of the structure factors. Therefore, the absolute numbers in the previous section 8.1 need to be taken with care. However, the goal of that section was a comparison between the traditional XCW formula and a new formula where χ^2 has units of energy. For the purpose of comparison the numbers are adequate. Nevertheless, more research is needed to better understand the behaviour of the self-electrostatic energy.

8.3 How well can X-ray constrained wavefunction calculations retrieve electric field effects?

Using the XCW method, one can combine quantum chemical calculations with experimental information of crystals in the form of structure factors. The wavefunction calculation itself is often a simple Hartree-Fock type calculation although also DFT is possible. The structure factor part compensates for what was neglected in the calculation, in particular electron correlation and the effect of the crystal field.
Bytheway *et al.* [251] calculated the effect of dynamic correlation on structure factors and predicted that it is often just above the 1% assumed detectability threshold and thus much smaller then the effect of the asphericity of the electron density. Furthermore, the information about correlation is only in few data points that often have low intensity. Therefore, the authors concluded that very careful experiments are needed to be able to model correlation effects and that for small molecules it is even impossible.
Genoni *et al.* [64] made a theoretical study where they were simulating structure factors from wavefunctions at a correlated level of theory (CCSD or CISD) and then used these structure factors for XCW calculations at Hartree-Fock level. Subsequently, the resulting electron density was compared to the initial CCSD or CISD one. They found out that the angular resolution at which the structure factors are calculated plays a crucial role. The effect is slightly counter-intuitive at first glance because a higher resolution with more simulated structure factors does not lead to a better reconstruction of the electron density but to one which differs more from the initial CCSD or CISD one. This was explained with the fact that if the high-resolution reflections, which arise mainly from the scattering of the core, become predominant they mask the effect of electron correlation, which has the largest influence on the valence electrons and therefore the low-angle reflections. Furthermore, they found an oscillating behaviour of the difference between simulated and calculated structure factors with respect to the resolution. The authors concluded that correlation effects are indeed visible in the XCW electron density provided that only low and medium-angle reflections are used and the λ-parameter is sufficiently high.

In this section the focus lies instead on the effect of electric fields. We will start with a linear, homogeneous field and subsequently also study the effect of point charges mimicking the electrostatic potential of surrounding molecules in a crystal. The exemplary molecule chosen for this study is the ubiquitous urea. Subsequently, the effect of the crystal field was studied using structure factors from periodic boundary calculations with *CRYSTAL17* for L-alanine.

In all cases studied here, only the *reference wavefunction* contains explicit information about the electric field and charges or about the entire crystal environment. From the reference wavefunction structure factors are simulated and fitted with X-ray constrained wavefunctions. In the quantum chemical part of this wavefunction calculation there is no field or charges. Their effect is only accounted for by the structure factor part. The resulting X-ray constrained wavefunction are compared to the reference WFNs to evaluate how well the effect of the electric field is included.

8.3.1 The effect of linear electric fields on the electron density from XCW calculations

Electric field along dipolar axis of urea

A wavefunction was calculated at HF/6-311++G(2d,2p) level of theory with an electric field applied along the carbon-oxygen bond with a strength of 0.01 au.

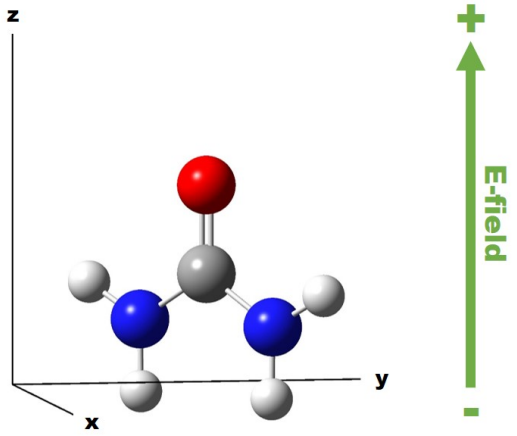


Figure 8.5 – Urea molecule with the C=O bond along the z-axis. The electric field is also oriented in this direction.

Structure factors were simulated for the resolutions $\sin(\theta)/\lambda$ = 0.3, 0.5, 0.7, 0.9, 1.2, 1.5, and 2.0 $Å^{-1}$ and a 10 × 10 × 10 $Å^3$ unit cell. *Tonto* was used for the XCW calculations at same level of theory as used for the reference wavefunction (HF/6-311++G(2d,2p)) and with the same unit cell as used for the structure factor simulation. The scale factor was not optimized and all the uncertainties on the structure factors (σ) were set to 1 e. The electron density was calculated with *Multiwfn* [252]. Subsequently, the electron density resulting from the X-ray wavefunction was compared to the reference density by analysing bond critical points and the RSR value as defined in equation 8.6.

Figures 8.6 and 8.7 show the same data represented differently. In Figure 8.6 the RSR value is plotted against λ. The RSR value describes the agreement between the XCW density and the density of the reference wavefunction calculation with an applied electric field. With increasing λ-values the RSR is decreasing, i.e. the electron density from the XCW calculation is more in agreement with the reference wavefunction. The lower the resolution of the calculated set of structure factor is, the faster the decrease happens.
Conversely, Figure 8.7 depicts the RSR value plotted against the angular resolution. For a given λ, the RSR is lower the lower the resolution is. This holds true down to 0.3 $Å^{-1}$, but at a resolution of 0.1 $Å^{-1}$ the agreement is worse. With the highest lambda value (λ = 20) this reversing trend is already visible at resolution 0.5 $Å^{-1}$ because the curve starts to go less steeply down, which eventually leads to an increase again.
Actually, a higher resolution signifies that more structure factors were calculated (to higher

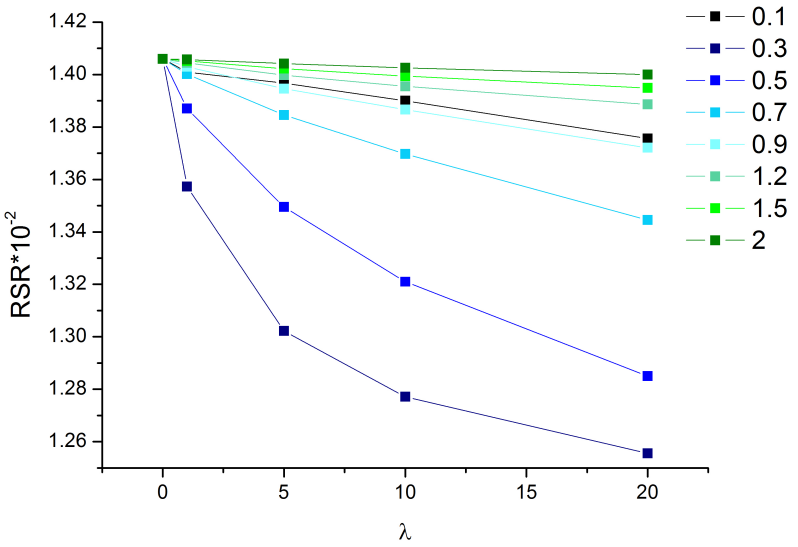


Figure 8.6 – RSR plotted against λ for different angular resolution $(\sin(\theta)/\lambda)$.

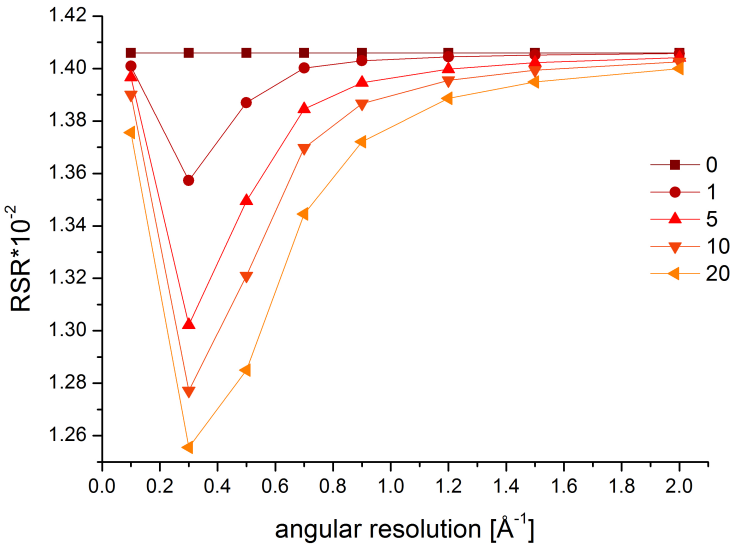


Figure 8.7 – RSR plotted against angular resolution $(\sin(\theta)/\lambda)$ for different λ.

scattering angles) and one would expect better results. However, the effect of the electric field is largest on the valence electrons because they are the most polarizable ones. The valence electrons scatter at low angles while at high scattering angles the influence of scattering from the core is clearly dominant. Therefore, including more high angle data just conceals the effect on the valence electrons by giving the important low-angle reflections a lower relative weight compared to the total dataset. Yet, below a certain resolution the amount of data becomes too

low, which explains the steep increase from 0.3 to 0.1 $Å^{-1}$.

Figure 8.8 shows the electron density at the BCP of the C=O bond along which direction the electric field was applied. Initially, only points at angular resolutions 0.1, 0.3, 0.5 $Å^{-1}$, and higher were calculated and the point at 0.3 $Å^{-1}$ looked like an outlier. However, the calculation of points at 0.2, 0.25, and 0.35 $Å^{-1}$ confirmed the correctness of this point and the curve indeed drops down followed by a steep increase. This behaviour resembles the oscillating behaviour found by Genoni *et al.* [64]. The point at 0.3 $Å^{-1}$ is much lower and therefore it is doubtful whether one can trust the results at this low resolution despite the RSR value being the best at that resolution.

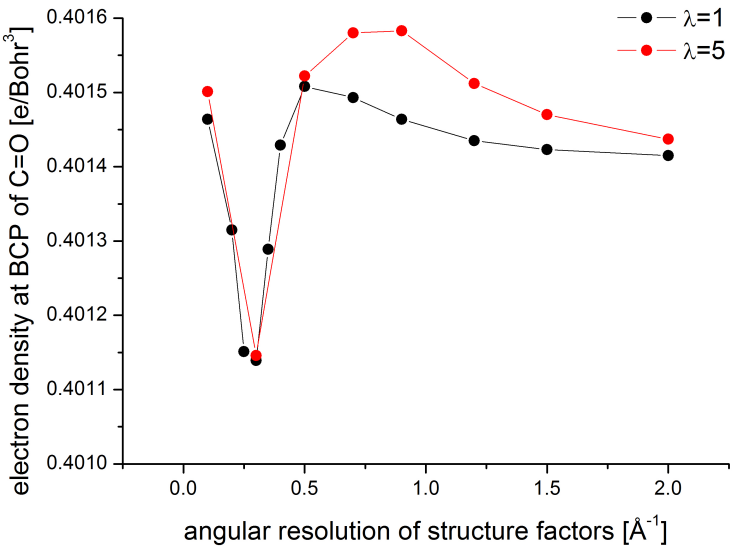


Figure 8.8 – Electron density at the BCP of C=O plotted against angular resolution $(\sin(\theta)/\lambda)$ for different λ.

Figure 8.9 shows the differences between the structure factors calculated from the wavefunction with and without field. In contrast to the findings of Genoni *et al.* [64] for the effect of correlation, there is no obvious oscillating behaviour in the structure factor difference itself. The effect of the field is more continuous on all resolutions. All the reflections which change by more than 2 % on a relative scale, are h,k,0 reflections and thus in the direction of the applied field.

"Randomly" oriented linear electric field

A Hartree-Fock wavefunction was calculated under an electric field of: x=0.01 au, y=0.02 au, z=0.03 au as shown in Figure 8.10. As before, XCW calculations were fitting structure factors obtained from this wavefunction and a hypothetical $10 \times 10 \times 10$ $Å^3$ unit cell.

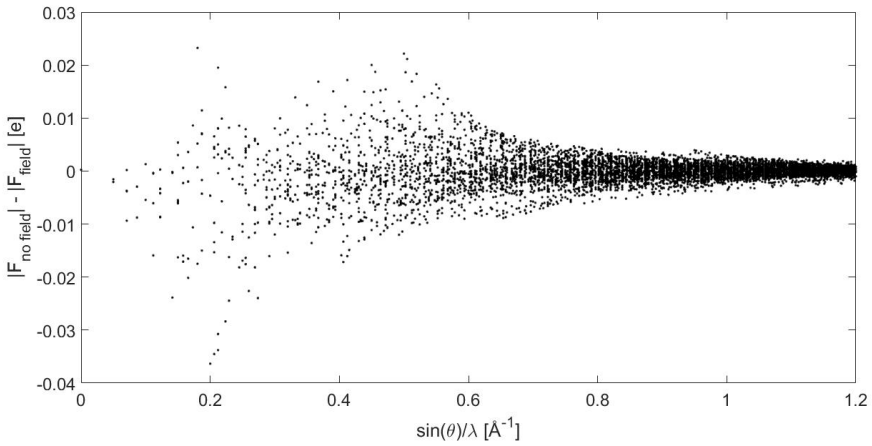


Figure 8.9 – Difference between structure factors from urea wavefunction calculated with and without field between 0 and 1.2 $Å^{-1}$.

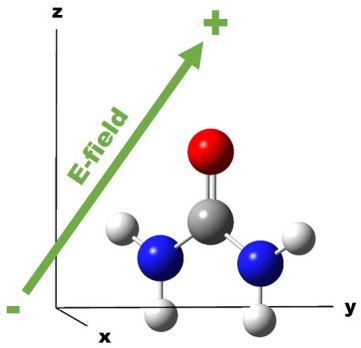


Figure 8.10 – Urea molecule with randomly oriented electric field.

The same behaviour of the RSR value as with the linear electric field along C=O can be seen as illustrated in Figure 8.11. Since the simulated field is larger, also the RSR value is larger, meaning that there is less agreement between the XCW density and the reference density of the wavefunction calculated with the field. It does not seem to matter in which direction an electric field is applied, truncation of the structure factor data at a lower resolution anyway leads to electron densities which are better in agreement with the one from the reference wavefunction.

8.3.2 The effect of point charges around urea simulating neighbouring molecules

The first coordination sphere of urea in the crystal was simulated with ten urea molecules placed around the central one as illustrated in Figure 8.12. *Gaussian* was used to calculate the charges that fit the electrostatic potential of these ten molecules. Subsequently, the point charges placed at the position of every atom were used to calculate a wavefunction. This wavefunction simulates the electric field due to the influence of the first coordination sphere

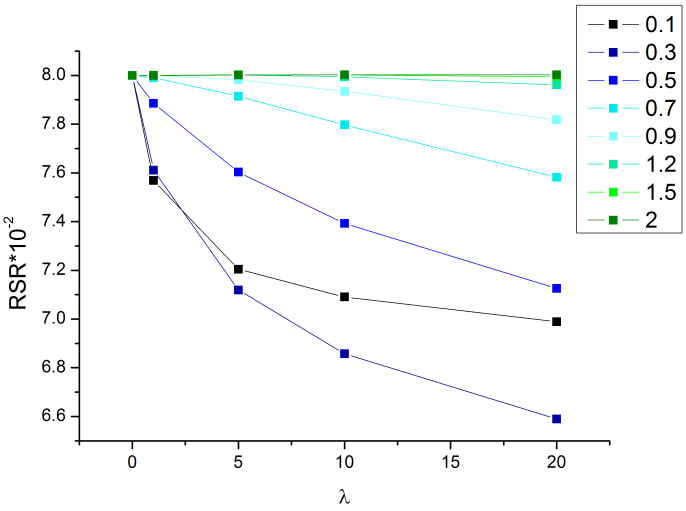


Figure 8.11 – RSR plotted against λ for different $\sin(\theta)/\lambda$. The RSR represents the similarity between the XCW density and the density from a wavefunction calculated with a linear, randomly oriented field.

in the crystal but does not take into account long range interactions.
The effect of point charges fitting the electrostatic potential used in *Gaussian* and point charges automatically assigned in *Tonto* for a cluster radius of 8 Å is similar and detailed in Figure B.1 in the appendix.

From the wavefunction calculated with charges, structure factors were simulated and subsequently used for XCW calculations.

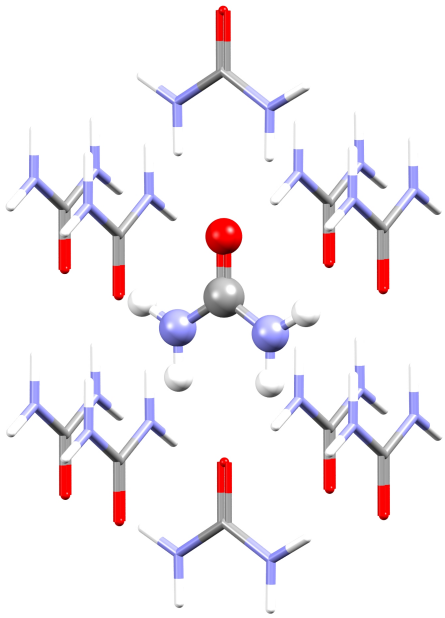


Figure 8.12 – The central molecule is shown with balls and sticks and the ten surrounding urea molecules with sticks. For the calculation of the wavefunction the surrounding molecules were represented as point charges at every atom position fitting the electrostatic potential.

The same trends with respect to the λ-parameter and the angular resolution as for the two linear fields can be seen in Figure 8.13. When judged based on the RSR value, the strength of the effect is between the effect of the two linear fields studied before and thus in absolute magnitude between 0.017 au and 0.037 au.

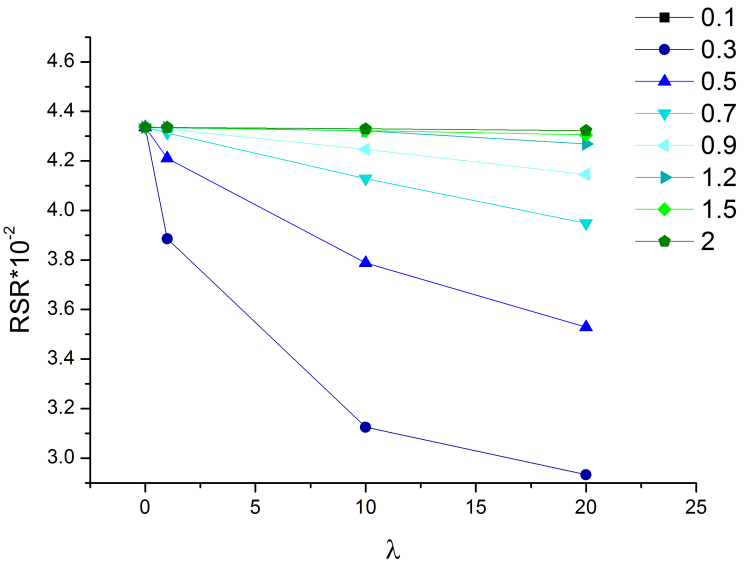


Figure 8.13 – RSR plotted against λ for different resolutions $\sin(\theta)/\lambda$. The RSR represents the similarity between the XCW density and the density from a wavefunction calculated with point charges around the central urea.

While the RSR value is just one number for the comparison of the entire electron densities, one can also plot a density difference map to see more in detail where these changes happen. Figure 8.14 shows that due to the effect of point charges at the positions of surrounding oxygen atoms, electron density moves away from the hydrogen atoms to the nitrogen causing the red lobes. Contrariwise electron density moves from the C=O bond towards the oxygen atom due to the positive surrounding charges.

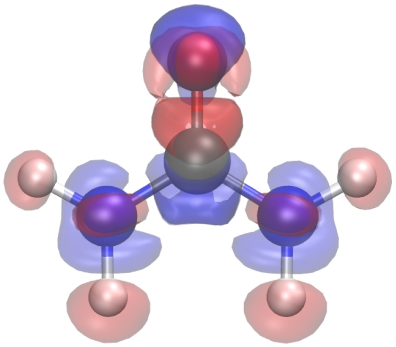


Figure 8.14 – Difference between density from wavefunction without field and from XCW at λ=10, resolution = 0.7 $Å^{-1}$. The isovalue is 0.0008 e/Bohr^3, with red for positive and blue for negative contour levels. Therefore, blue signifies more density in the XCW case and red less electron density.

Table 8.9 shows the total molecular dipole moment of urea obtained from the different wave-

functions. The one from the X-ray constrained wavefunction is in between the one from the wavefunction with charges and the one without any charges. However, it is closer to the one without charges and thus the XCW procedure does not recover the entire effect of the electric field due to the charges. It is worth mentioning that the calculated dipole moments are in agreement with higher level calculations of Santos *et al.* [253] who obtained 4.67 D for the dipole moment of urea *in vacuo* and 7.54 D in the same solid phase of urea as used here.

Table 8.9 – Total molecular dipole moment in Debye

Calculation of wavefunction	dipole moment
WFN without charges	5.254
X-ray constrained WFN with $\lambda = 10$ and resolution = 0.7 Å^{-1}	5.511
WFN with charges simulating crystal environment	7.900

8.3.3 Comparison of electric field and correlation effects on the electron density from XCW calculations

As written previously, electron correlation and the crystal field are not considered in the Hartree-Fock part of an XCW calculation but are present only in the structure factor part. This justifies studies examining how well these effects are included in XCWs.
A brief comparison between these two effects is made here. The correlated wavefunction is a CCSD one obtained with *Gaussian*. The urea molecule is in a slightly different geometry than for the studies of the electric field but this is of no relevance here.

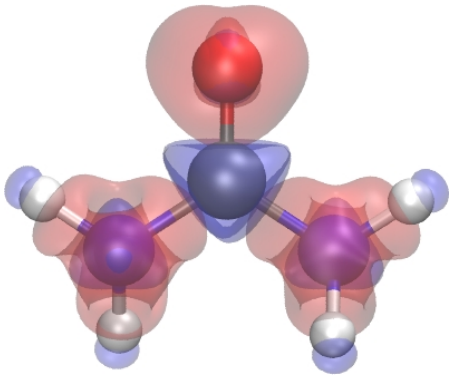
Table 8.10 shows the RSR agreement value for different XCW densities with respect to the respective reference density.
The effect of an electric field depends on its strength. Depending on that it can have a stronger or weaker influence on the electron density than correlation has. Increasing λ always leads to an improved agreement.

Table 8.10 – RSR values for various XCW calculations using structure factors simulated up to resolution 0.7 Å^{-1}.

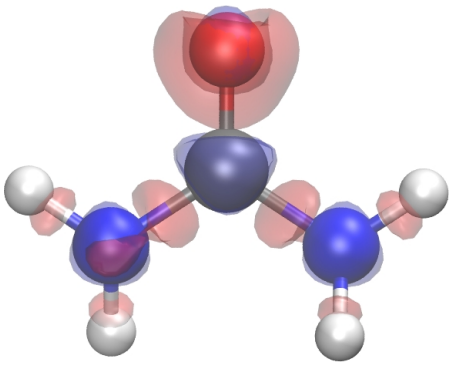
Structure factors simulated using wavefunction including:	RSR*10^{-2}	
	λ=1	λ=10
Correlation	4.683	3.950
E-field linear along C=O	1.400	1.370
E-field random linear	7.990	7.797
Point charges	4.313	4.128

Figure 8.15 visualizes the different effect of correlation and the electric field due to point charges around urea. While the charges shifts the outermost electrons in the positive direction of the field (as with a linear field along C=O) and produce a striped pattern, correlation moves

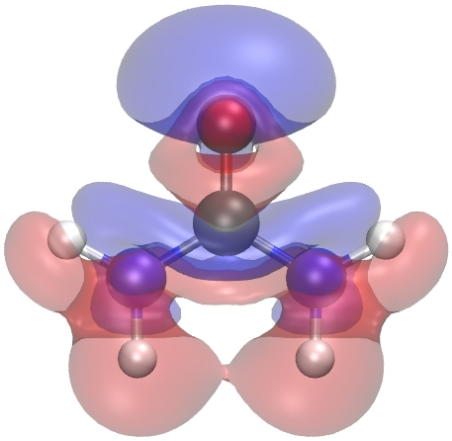
electron density 1) away from the bonds to the nuclei and 2) generally outwards.
The comparison of the XCW density to a normal Hartree-Fock one always provokes the same effect as the comparison between the reference density compared to the normal Hartree-Fock one, but the effect is damped.



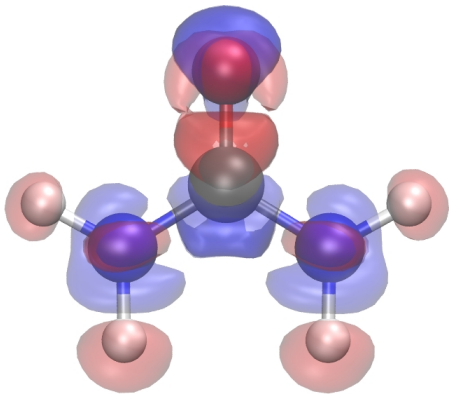
(a) HF - CCSD in direct space



(b) HF - XCW (fitting structure factors including correlation)



(c) HF (no charges) - HF (point charges)



(d) HF - XCW (fitting structure factors including effect of charges)

Figure 8.15 – Density differences for the effect of correlation and point charges. The isosurface level is 0.002 e/Bohr^3 for a) and b) and 0.0008 e/Bohr^3 for c) and d). Red represents positive and blue negative contour levels. The XCW used here were done with λ=10 and resolution 0.7 Å^{-1}.

With higher λ the agreement between the XCW and the reference wavefunction improves as well as generally with lower angular resolution of the structure factors.
Point charges simulating the electric field have in total magnitude a similarly large effect on the electron density as electron correlation as can be seen from the RSR value and the difference densities.
While the definition of an optimum λ-value is rather difficult (see section 8.1), it is possible to say which resolution is best for a specific case, namely the one giving the lowest RSR without producing any peculiarity.

8.3.4 Effect of the crystal field on the electron density from XCW calculations

Besides studying the effect of an electric field or point charges on structure factors and the ability of X-ray constrained wavefunction calculations to include these effects, also the crystal field can be investigated with similar methods. In chapter 9 the effect of the crystal field on the X-ray constrained density and especially also on X-ray constrained extremely localized molecular orbitals (XC-ELMOs) of different molecules is examined. In this section the total electron density of L-alanine is analysed after shortly introducing the independent molecule model (IMM).

Independent molecule model for structure factor simulation

A wavefunction calculated with periodic boundary conditions (crystal calculation) can be used to simulate structure factors. When standard crystal calculations are carried out, these structure factors include long-range crystal field effects and short-range intermolecular interactions. However, it is also possible to perform crystal calculations where these effects are explicitly excluded using the *MOLSPLIT* option of *CRYSTAL17*. This avoids molecule-molecule interactions by removing the molecules very far apart from each other. After the wavefunction calculation has been done, the molecules are placed again in their normal crystal geometry, and the superposition of the individual molecular densities gives rise to the crystal one. We call this approach using a non-interacting system the independent molecule model (IMM).
Both the 'normal' and the IMM structure factors can be used for XCW calculations. The densities of the so-obtained X-ray constrained wavefunctions can be subtracted to highlight differences. Therefore, the influence of the crystal field and intermolecular interactions going via reciprocal space, on the density of a molecule can be studied. The same holds for XC-ELMOs, apart from the fact that the orbitals are squared before comparison to get rid of the phase.
This method to investigate the influence of a crystal environment on the XCW results works only for simulated structure factors since in experimental ones, interactions cannot be artificially excluded. However, in that case, one can instead analyse the electron density differences resulting from the comparison of an unconstrained calculation ($\lambda = 0$) and of an XCW computation ($\lambda > 0$ and experimentally measured structure factors as restraints).

L-alanine in the crystal

A periodic wavefunction of L-alanine based on its real crystal structure (CSD name LALNIN23, Figure 8.16) was calculated with *CRYSTAL17* at HF/TZVP [245] level of theory. In addition, an IMM wavefunction with the same specifications was computed. Structure factors were simulated for the full sphere up to resolution 1Å^{-1}. Hartree-Fock was used because correlation is not of interest for this study and the basis set was changed with respect to the previous section because the one used in *Tonto* did not lead to convergence for the periodic calculations with

CRYSTAL17. Nevertheless, the 6-311++G(2d,2p) basis set was used for the XCW calculations.

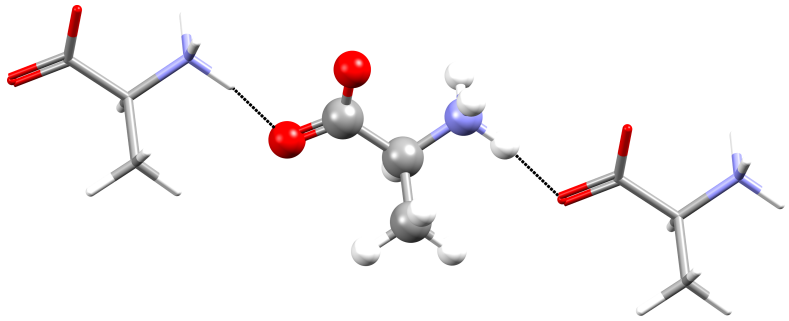
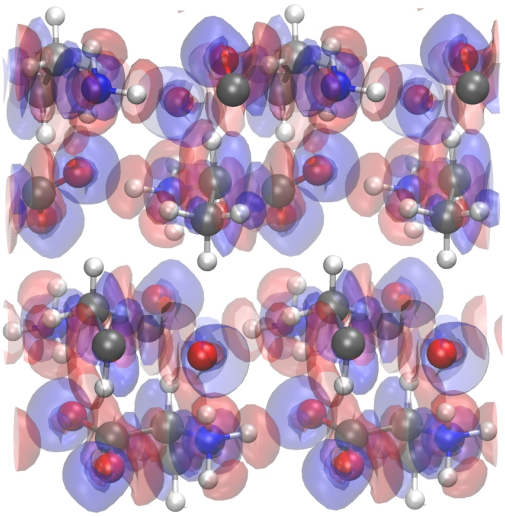
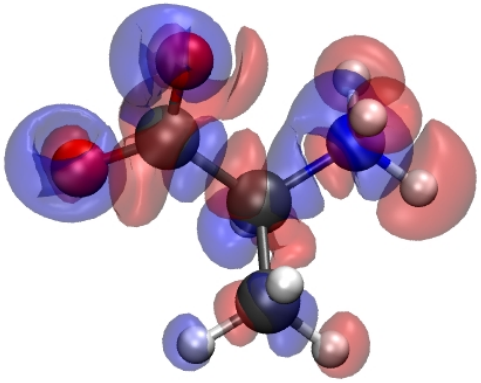


Figure 8.16 – Strong hydrogen bonds formed by an L-alanine molecule in the crystal.

Figure 8.17a depicts the direct space electron density difference between the normal *CRYSTAL17* calculation and the IMM one and is therefore illustrating the effect of interactions in the crystal. Figure 8.17b instead shows the density difference map of an XCW fitting to normal and to IMM structure factors of a *CRYSTAL17* calculation. The features in Figures 8.17a and b are qualitatively very similar although the effect is weaker when going via reciprocal space fitting and therefore the isovalue in the XCW picture is chosen lower.
The blue lobes around the oxygen atoms represent an accumulation of electrons when using structure factors including the crystal field effect. This can be explained by the interaction of the oxygen atoms with partially positively charged neighbouring atoms, especially the strong hydrogen bond shown in Figure 8.16. Due to this interaction, electrons are polarized from the C-O bond towards the oxygen atom causing the observed density difference. On the hydrogen atoms of the amine the opposite side of this effect becomes visible. Due to the strong hydrogen bond, electrons from the oxygen atoms of a neighbouring molecule push electrons on the hydrogen atoms into the N-H bond. Thus, there is a local depletion of electron density explaining the red lobes on these atoms. It is to be pointed out that, with this method, one highlights the rearrangement of electron density within the molecule itself as a consequence of the crystal field and intermolecular interactions. But one does not see accumulation of electron density in the regions of intermolecular interactions. It is an intrinsic part of XCW calculations, which use molecular wavefunctions to describe molecules in crystal, that this is so. Nevertheless, the example of L-alanine demonstrates that effects due to intermolecular interactions in the crystal are recovered in XCW calculations and density difference maps can be used to visualize them.



(a) Direct subtraction of densities from *CRYSTAL17* calculation $\rho_{IMM} - \rho_{crystal}$. The picture appears chaotic because cubes of periodically repeated units are subtracted and not a molecular density. The isovalue is 0.002 e/Bohr^3. Blue regions indicate an accumulation of electron density for normal crystal calculations with respect to IMM calculations.



(b) Density differences obtained by comparing XCW (λ=5) calculations with structure factors including and excluding the electric field. The isovalue is 0.0008 e/Bohr^3. Blue regions indicate an accumulation of electron density for XCW calculations with structure factors including the crystal field effects.

Figure 8.17 – Electron density difference maps for L-alanine.

8.4 Frozen density embedding using X-ray constrained wavefunctions

The calculation of a wavefunction is limited to systems without too many atoms and electrons. For the case of a solute molecule in a solvent, for example, it is usually not possible to compute an explicit wavefunction for the entire system. Calculating only the solute alone, instead, is not accurate enough for most problems. For this reason, strategies had to be developed that allow for the calculation of an explicit wavefunction of the solute while taking into account the environment. The embedding approach by Wesołowski *et al.* [254, 255] divides the total system into two subsystems S_A and S_B, where the total density is the sum of the two densities:

$$\rho(\mathbf{r}) = \rho_A(\mathbf{r}) + \rho_B(\mathbf{r}) \tag{8.9}$$

Only for subsystem S_A (for example the solute molecule) a wavefunction is explicitly calculated. From the wavefunction the electron density can be calculated. For HF and DFT it is:

$$\rho_A(r) = 2\sum_{i=1}^{N/2} \varphi_i^* \varphi_i \tag{8.10}$$

The influence of subsystem S_B is considered via a constant density $\rho_B(\mathbf{r})$. This approximation requires that the interactions between S_A and S_B are not too strong and not of dominantly covalent nature. Due to the use of a constant density in which an explicitly calculated molecular wavefunction is embedded, this method was termed Frozen-density embedding theory (FDET). With Kohn-Sham equations, the effect of the frozen density is taken into account via an effective potential $V_{eff}^{\#}(\mathbf{r})$, which replaces the conventional potential:

$$\left[-1/2\nabla^2 + V_{eff}^{\#}(\mathbf{r})\right]\varphi_i = \varepsilon_i \varphi_i \tag{8.11}$$

Subsequently, the equations can be solved in the usual self-consistent way. To obtain the embedding potential, the electron density $\rho_B(\mathbf{r})$ represented in any desired form and the potential from the nuclei $\nu_B(\mathbf{r})$ are needed.

Usually $\rho_B(\mathbf{r})$ is obtained with conventional quantum mechanical calculations. However, it is also possible to use an experimentally derived electron density such as for example the ones obtained from X-ray diffraction experiments with charge-density quality data. In that case the electron density can either be represented by a multipolar model or it can come from X-ray constrained wavefunctions. Obviously, the potential from the nuclei is also available in these cases, since the atomic coordinates must be known in both methods.

For this study, X-ray constrained wavefunction calculations of glycylglycine were performed to obtain experimental densities, which were subsequently used for embedding. The embedded molecules were acrylic acid, acrolein, and acetone in different orientations with respect to glycylglycine. The excitation energies of the embedded molecules were studied focusing on the effect of changing the λ weighting parameter in the X-ray constrained wavefunction

calculations.

It was found that the excitation energies obtained with $\rho_B(\mathbf{r})$ from X-ray constrained wavefunctions are comparable to the ones obtain with $\rho_B(\mathbf{r})$ from conventional quantum wavefunctions. In contrast to those, XCWs are experimental and implicitly take into account correlation effects even when calculated at Hartree-Fock level of theory. In addition, XCWs represent the situation in the crystal and not in gas-phase. Therefore, they might provide a useful alternative to conventional ones in cases where the embedding system itself should already include intermolecular interactions.

9 Analysis of Crystal Field Effects and Interactions with X-ray Restrained ELMOs

Apart from the addendum, this chapter is based on the paper "Analysis of crystal field effects and interactions with X-ray restrained ELMOs", which was submitted to the special issue "Charge density studies towards structure-property correlation" of the *Journal of Molecular Structure*.

Abstract: *The purpose of this study is to assess the capability of X-ray restrained molecular wavefunction methods to reproduce the perturbation due to the crystal field.*
The effect of the crystal environment on the molecular charge densities has been thoroughly investigated experimentally and theoretically. However, the same analysis in terms of molecular orbitals is unprecedented. We specifically analysed the extremely localized molecular orbitals (ELMOs), because they allow for a valence bond like interpretation that is closer to the typical chemical reasoning.
For a set of test molecules, we calculated first principles wavefunctions (called primary wavefunctions) and the corresponding scattered intensities (primary structure factors), that we used for X-ray restrained Hartree-Fock and ELMO calculations. This mimics the typical procedure of modelling wavefunctions from experimental X-ray diffraction, having the primary wavefunction as benchmark.
This study follows a previous investigation concerning the effects of electron correlation, where it was shown that those subtle features could be retrieved only to a minor extent. The electric field generated by surrounding molecules is instead more substantial. Therefore, retrieving this information should be more successful and, in case, extremely useful for supramolecular chemistry and crystal engineering. The work also provides useful indications to experimentalists who want to exploit the potentialities offered by the X-ray restrained wavefunction methods.

9.1 Introduction

Long-range crystal field and short-range intermolecular interactions significantly affect the structure and properties of molecules in crystals. Long-range forces are mainly due to atomic/molecular charges whose strength is decaying with the reciprocal square power of the distance, whereas other forces are decaying more rapidly. On the other hand, short-range forces are a blend of different effects, including partially covalent interactions, charge transfer, the apparent repulsion due to the Pauli exclusion principle, London-type forces, as well as the electrostatic forces mentioned above. The interplay among these different factors is extremely complex. However, understanding these interactions in crystals is fundamental for the purpose of rationally designing materials featuring specific properties and, more generally, for crystal engineering. Various techniques are being used to study interactions in crystals, from the traditional examination of crystal structures to the equally important analysis of electron densities and their derivatives [256, 257, 258, 259] obtained from purely theoretical calculations or from experiments (especially from X-ray diffraction measurements and subsequent modelling). Since the correlation of electron densities with properties is crucial, various descriptors based on the electron density itself were derived [25, 260, 33, 261, 262] and used to analyse interactions .

In this context, the possibility to compute so-called X-ray constrained wavefunctions (XCWs) [106, 263, 128, 264, 265, 64, 266, 127] has emerged as a powerful approach that enables scientists to retrieve not only one-electron density distributions, but also tentative wavefunctions from experimental X-ray diffraction data. Since these wavefunctions are actually restrained (not exactly constrained) to structure factors (SF), we will subsequently refer to them as X-ray restrained wavefunctions (XRWs) in contrast to the generally used term in literature. These experimental wavefunctions can be obtained through the simultaneous minimization of the electronic energy and of the difference between observed and calculated structure factor magnitudes for a molecular wavefunction (WFN) ideally embedded in a crystal. X-ray restrained wavefunctions have been used in diverse studies for the analysis of crystal interactions and properties, e.g. for refractive indices [129, 267] or for the study of bonds [268, 136, 269, 270, 271], etc. Although so far these studies mainly focused on the electron density and its related properties in crystals, the XRW technique also provides molecular orbitals. As we will see in this study, by exploiting orbitals from XRW calculations, one would be able to exploit the full arsenal of this new technique.

Although extended to different strategies of quantum chemistry [272, 273, 130, 274, 275, 276, 277, 278], the XRW approach was originally developed in the framework of the Hartree-Fock formalism and this is the level of theory that is still mainly used when X-ray restrained wavefunction calculations are performed. However, when the Hartree-Fock technique is exploited, both in the traditional way and within the XRW method, the resulting canonical orbitals generally spread over the entire molecule, thus hampering a straightforward chemical interpretation. On the other hand, more frequently, chemical-bonding analyses based on MOs takes advantage of some localization strategies, reminiscent of valence bond treatments, or of

some perturbation theory-based approaches (as for example the analysis of frontier molecular orbitals or orbital diagrams widely adopted also in textbooks). Concerning the localization techniques, the goal is achieved by obtaining molecular orbitals mainly localized on small fragments within the molecule (generally, atoms and bonds). For example, traditional localized MOs are those obtained through unitary transformations of the canonical Hartree-Fock molecular orbitals [279, 280, 281, 282, 283, 284, 285] that maximize or minimize chemically/physically sound functionals (such as, the orbitals self-extension [279, 280], the orbitals auto-repulsion energy [281, 282], the charge density overlap [283], the local orbital populations [284, 285]). Obviously, being unitary transformations of the canonical Hartree-Fock orbitals, they provide a wavefunction equivalent to the canonical Hartree-Fock one (they differ only for a phase factor) and, consequently, the same properties (e.g. total energy and electron density) for the system under exam. However, they are not exactly localized, but they have tails extending over other regions of the molecule than the specified fragment. This partially reduces their chemical interpretability.

In order to build molecular orbitals that are strictly localized on molecular subunits, we have to resort to the so-called extremely localized molecular orbitals (ELMOs). [37, 286] They are molecular orbitals exactly localized on any of the desired fragments, defined by a localization scheme chosen a priori according to the chemical intuition or some computational needs. In this case, due to a reduction in the variational space (i.e., the ELMOs are expanded on a subset of the available atomic basis functions, see section 9.2.1), the Hartree-Fock representation is not preserved and, hence, the total energy associated with the ELMO wavefunction is higher than the corresponding Hartree-Fock one. This is the price to pay to have a strict localization and a better chemical interpretability.

Similar to the XRW approach with canonical Hartree-Fock MOs, an X-ray restrained wavefunction technique in the framework of the ELMO theory was also developed (XC-ELMO strategy in literature, subsequently termed XR-ELMO) [272, 287, 130, 274]. As in the unrestrained case, the strict localization allows for a more straightforward chemical interpretability of the molecular orbitals. An inherent and immediate advantage of obtaining localized functions is the possibility to define them for specific fragments, such as atoms, bonds or functional groups, and to consequently store them as electronic LEGO building blocks in proper libraries [288, 289, 290] with the final goal of reconstructing electron densities and related properties [291, 292], including structure factors. This allows for the refinement of crystallographic structures of large systems (e.g. proteins and peptides) with undoubted advantages. [293]

In this study, we investigate another possibility offered by the X-ray restrained ELMOs: exploiting them to interpret supramolecular interactions in crystals within an orbital framework. In this way, the covalent interactions (visualized through orbitals) as well as the electrostatic ones (visualized through densities) are equally accounted for. The goal of using ELMOs is to pinpoint orbitals that are mostly perturbed by an interaction and, consequently, to get insights into their roles (and the roles of the corresponding molecular fragments) in the stabilization of a crystal packing. Furthermore, it is interesting to find out whether the change in the extremely

localized molecular orbitals is due to short-range or orbital interactions or, otherwise, to long-range effects due to the crystal field (see Figure 9.1).

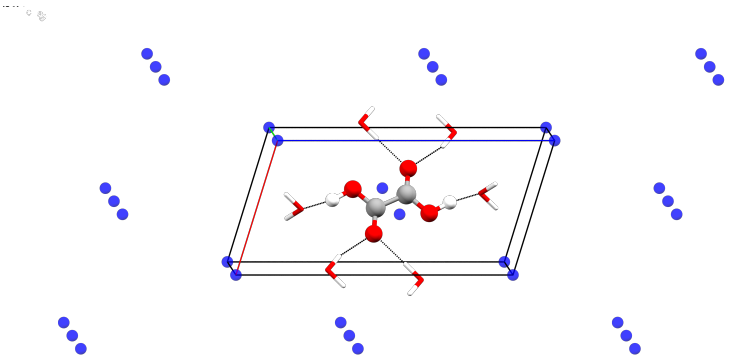


Figure 9.1 – Oxalic acid molecule in a crystal. The strongest short-range interactions with the surrounding water molecules are shown with black lines. Blue dots represent other oxalic acid molecules in the crystal responsible for the long-range crystal effects.

The paper is organized as follows. In the next section, we briefly describe the theoretical and computational details of our investigation, also presenting the fundamentals of the ELMO and XR-ELMO methods and describing the systems taken into account. Afterwards, we present and discuss the obtained results and finally, in the last section, we draw general conclusions.

9.2 Theory and Methods

9.2.1 The ELMO and XR-ELMO Techniques

The ELMO strategy originally proposed by Stoll [286] requires a predefined localization scheme that subdivides the molecule into non-exclusive subunits. This implies assigning local basis-sets $\beta^i = \{|\chi^i_\mu\rangle\}_{\mu=1}^{M^i}$ to the different fragments, each of them given by the basis functions (or atomic orbitals) centred on the atoms that belong to the considered subunit. Afterwards, the molecular orbitals corresponding to the different fragments are expanded on these local basis-sets. For example, for the i-th subunit, the α-th ELMO can be expressed as:

$$|\varphi^i_\alpha\rangle = \sum_{\mu\in\beta_i} C^i_{\mu\alpha}|\chi^i_\mu\rangle \tag{9.1}$$

The molecular wavefunction is the single Slater determinant constructed with orbitals defined in equation 9.1:

$$|\Phi_{ELMO}\rangle = \frac{1}{\sqrt{(2N!)}det\,[\mathbf{S}]}\hat{A}\left[\varphi^1_1\bar{\varphi}^1_1\cdots\varphi^1_{n_1}\bar{\varphi}^1_{n_1}\cdots\varphi^f_1\bar{\varphi}^f_1\cdots\varphi^f_{n_f}\bar{\varphi}^f_{n_f}\right] \tag{9.2}$$

with $\hat{A}$ as the antisymmetrizing operator, n_i as the number of occupied ELMOs for the i-th fragment, $\bar{\varphi}^i_\alpha$ as a spin-orbital with spatial part φ^i_α and spin part β, and $det\,[\mathbf{S}]$ as the

determinant of the overlap matrix between the occupied ELMOs.

The coefficients of the linear combination 9.1 that define the ELMOs are obtained by minimizing the energy associated with the ELMO wavefunction and are obtained by solving modified Hartree-Fock equations for each fragment in which the molecule under exam was initially subdivided:

$$\hat{F}^i|\varphi^i_\alpha\rangle = \epsilon^i|\varphi^i_\alpha\rangle \tag{9.3}$$

with $\hat{F}^i$ as the Fock operator for the i-th fragment, which is defined as

$$\hat{F}^i = \left(1-\hat{\rho}+\hat{\rho}^{i^\dagger}\right)\hat{F}\left(1-\hat{\rho}+\hat{\rho}^i\right) \tag{9.4}$$

where $\hat{F}$ is the traditional Fock operator of quantum chemistry, $\hat{\rho}^i$ is the local density operator for the i-th subunit, which depends only on the occupied ELMOs of the fragment, and $\hat{\rho}$ is the global density operator, which depends on all the occupied ELMOs of the molecule and, for this reason, couples the eigenvalue-equations associated with the different subunits.

The only difference between the unrestrained and the X-ray restrained ELMO techniques is that the former are obtained with a traditional variational method, whereas the latter minimize the sum of the total energy of the system and the disagreement with respect to a set of X-ray structure factors weighted by their experimental uncertainties (if coming from a measurement). This means that the desired X-ray restrained ELMOs are those extremely localize molecular orbitals that always respect the expansion constraints imposed by equation 9.1, but that also minimize the functional:

$$J[\Psi_{ELMO}] = \langle\Psi_{ELMO}|\hat{H}|\Psi_{ELMO}\rangle + \lambda(\chi^2[\Psi_{ELMO}]) \tag{9.5}$$

where $\hat{H}$ is the non-relativistic Hamiltonian operator for the reference molecule in the unit-cell of the crystal under exam, λ is an external multiplier giving the strength of the experimental restraint, and χ^2 is the measure of the statistical disagreement between computed and experimental values. In particular,

$$\chi^2 = \frac{1}{N_r - N_p}\sum_h \frac{\left(\eta|F^{calc}_h| - |F^{exp}_h|\right)^2}{\sigma^2_h} \tag{9.6}$$

with N_r as the number of considered X-ray diffraction data, N_p as the number of adjustable parameters, **h** as a triad of Miller indices labelling each reflection, σ_h as the experimental uncertainty corresponding to the observed structure factor amplitude $|F^{exp}_h|$, and η as a scale factor that is determined by minimizing χ^2. Although the method is conceived to use experimental structure factor magnitudes, also theoretically simulated structure factors can be used in equation 9.6. This is the procedure that we actually adopted for most of the X-ray restrained wavefunction calculations (including the XR-ELMO ones) carried out in the present study (see section 9.2.2 and the Results section).

Looking for the ELMOs that minimize functional 9.5 corresponds to solving a new set of modi-

fied Hartree-Fock equations (one for each subunit), [272, 287] which are formally analogous to those proposed by Stoll:

$$\hat{F}^{i,exp}|\varphi_\alpha^i\rangle = \epsilon^i|\varphi_\alpha^i\rangle \tag{9.7}$$

but where the modified Fock operator for the generic i-th fragment has this form:

$$\begin{aligned}\hat{F}^{i,exp} = &\left(1-\hat{\rho}+\hat{\rho}^{i^\dagger}\right)\hat{F}\left(1-\hat{\rho}+\hat{\rho}^{i}\right)\\ &+\lambda\sum_h K_h \mathrm{Re}\left\{F_h^{calc}\right\}\left(1-\hat{\rho}+\hat{\rho}^{i^\dagger}\right)\hat{I}_{h,R}\left(1-\hat{\rho}+\hat{\rho}^{i}\right)\\ &+\lambda\sum_h K_h \mathrm{Im}\left\{F_h^{calc}\right\}\left(1-\hat{\rho}+\hat{\rho}^{i^\dagger}\right)\hat{I}_{h,C}\left(1-\hat{\rho}+\hat{\rho}^{i}\right)\end{aligned} \tag{9.8}$$

where

$$K_h = \frac{2\eta}{N_r - N_p}\frac{\eta|F_h^{calc}| - |F_h^{exp}|}{\sigma_h^2|F_h^{calc}|} \tag{9.9}$$

and $\hat{I}_{h,R}$ and $\hat{I}_{h,C}$ are the real and imaginary parts of the structure factor operator given by

$$\hat{I}_h = \hat{I}_{h,R} + i\hat{I}_{h,C} = \sum_{k=1}^{N_m} e^{i2\pi(\mathbf{R}_k r+\mathbf{t}_k)\cdot(\mathbf{Bh})} \tag{9.10}$$

with $\mathbf{B}$ as the matrix of the reciprocal-lattice and $\{\mathbf{R}_k,\mathbf{t}_k\}_{k=1}^{N_m}$ as the N_m unit-cell symmetry operations.

9.2.2 Computational protocol

Compounds

We studied two different compounds starting with the very simple molecule HCN. First this molecule was examined in an electric field and second in a crystal. In both cases the same geometry of the HCN molecule was used, which was optimized *in vacuo* at the B3LYP/cc-pVDZ level of theory. In order to study HCN in a crystal, we considered a hypothetical crystal of HCN with no symmetry elements except translation (space group P1) and, consequently, with only one molecule per unit-cell. The distance between the nitrogen atom of one molecule and the hydrogen of its closest neighbour was chosen to be 2.77 Å, which corresponds to the mean lengths of this particular hydrogen bond as retrieved from the CSD [216]. The search was made for nitrogen triply bonded to carbon interacting with a hydrogen atom bonded to a carbon (C≡N···H-C). The angle (CNH) was constrained to be between 170 and 180° in the search to obtain only linear contacts. By chance, the unit-cell length in the direction along the molecule becomes exactly 5 Å. Therefore, the unit-cell dimensions were chosen to be 5 Å in each direction (a=b=c= 5 Å).
Afterwards, we considered oxalic acid dihydrate, a hydrogen-bonded system, for which we took into account the geometries of three studied pressure points, all of them taken from a study conducted by Casati *et al.* [294] and left unchanged.

Periodic ab-initio calculations and structure factor simulation

For the case of HCN in an electric field, the molecule was placed in a hypothetical 10×10×10 $Å^3$ unit-cell in the absence of any roto-reflection symmetry element and fixed molecular geometry in order to simulate structure factors with XD [65].
For all the examined crystal systems, single point periodic calculations with *CRYSTAL17* [23] were performed on the described structures. The calculations were done at the restricted Hartree-Fock (RHF) level of theory with a 6-31G(d,p) basis set. The RHF/6-31G(d,p) combination was consistently used for all the single-point wavefunction calculations *in vacuo* or with periodic boundary conditions, as well as for all the X-ray restrained calculations for sake of consistency. XR-ELMO calculations are time-intensive, which is the reason why we adopted a rather low level of theory and a rather contracted basis set. Noteworthy, a precise description of electron correlation is not of importance for the purpose of this study. At the moment, we are rather interested to test the machinery of XR-ELMOs and its ability to retrieve electric field induced changes to the wavefunction, not in a very accurate wavefunction itself.
For the example of HCN in a hypothetical crystal and α-OAD at 0, 3.6, and 5.3 GPa, structure factors were simulated with *CRYSTAL17*. In addition, the MOLSPLIT option was used to calculate a periodic wavefunction where no intermolecular interactions are present for reasons explained in section 9.3.1. The software automatically recognizes oxalic acid and water as individual molecules and separates them far apart. Table 9.1 gives an overview about the three studied compound, their unit-cell, symmetry, and resolution for which structure factors were simulated.

X-ray restrained calculations

For all the XRW calculations with canonical Hartree-Fock MOs (subsequently called XR-RHF) we exploited the *Tonto* [126] software. All computations were performed with a weighting factor λ equal to 5, if not otherwise specified. Different λ-values between 0.5 and 50 were also tested but they did not lead to qualitatively different results from the λ=5 case (see Figure 9.4). Since in this study we mainly used only theoretical structure factors as restraints, all the uncertainties σ_h were set equal to 1.0 *e* and the obtained χ^2 values were generally very low. All the XR-ELMO calculations were performed using a modified version for the *GAMESS-UK* quantum chemistry package [295] where the XR-ELMO equations have been implemented. Also for these computations, we set λ=5 and σ_h=1.0 *e*. The lattice and molecular geometry always corresponded to the calculation of the primary wavefunction.
For the X-ray restrained wavefunction fitting calculation of OAD, the asymmetric unit was chosen such that it contains one oxalic acid and two water molecules. For this purpose the inversion centre was removed and the atoms duplicated exactly at the symmetric coordinates such that the space group changed from $P2_1/n$ to $P2_1$.

Table 9.1 – Overview of different compounds used with corresponding resolution for structure factors.

System	unit-cell dimensions (Å^3)	symmetry	resolution for SF (Å^{-1})
HCN in electric field	10×10×10	P1	0.0-0.3, 0.0-0.7, 0.0-1.2, 0.0-1.8
HCN in crystal	5×5×5	P1	0.0-0.7
Oxalic acid dihydrate (periodic wavefunction calculation)	experimental	$P2_1/n$	0.0-0.8
Oxalic acid dihydrate (XRW calculation)	experimental	$P2_1$	0.0-0.8

Differences between electron densities

We analysed the differences between the electron densities resulting from the various computations (from now on, density difference maps or, more simply, density differences). In particular, electron densities obtained from XRW calculations based on structure factors including electric field or crystal interactions were subtracted from electron densities from XRW calculations using structure factors without interactions ($\rho^{XR}_{no\ field} - \rho^{XR}_{field}$ or respectively: $\rho^{XR}_{IMM} - \rho^{XR}_{crystal}$, see Table 9.2), as it will be described with more details in the Results section for each specific case. For the direct comparison of the obtained XR-ELMOs ($|\varphi^{XR}_{no\ field}|^2 - |\varphi^{XR}_{field}|^2$ and $|\varphi^{XR}_{IMM}|^2 - |\varphi^{XR}_{crystal}|^2$), we decided to consider and subtract the squared orbitals instead of the simple orbitals for two reasons; because the former are more directly connected to electron density distributions and, above all, because they do no depend on a phase factor.

9.3 Results

9.3.1 HCN

Electric field

A molecule embedded in a crystal interacts with its first neighbours through short range intermolecular contacts, but it also feels the long range electric field generated by the outer molecules. In order to simulate the latter, we made the simplest approximation, which consisted in applying a homogenous and linear electric field along the linear molecule HCN in both orientations of the main inertial axis. The applied field was 0.01 a.u., which is of the order of magnitude typically experienced by a molecule in a crystal with intermediately strong packing forces. From now on, we will indicate these calculations (and the resulting electron densities) as “direct space calculations” (and “direct space electron densities”) to distinguish them from computations (and electron densities) associated with the use of structure factors

Table 9.2 – Definition of various density differences. IMM refers to the independent molecule model (see the text for more details). Instead of (*IMM-crystal*), in all the cases mentioned below, the differences can also be (*no electric field – field*). XR refers to wavefunctions or orbitals obtained from X-ray restrained calculations. The primary wavefunction is the initially calculated wavefunction used to simulate the structure factors.

Symbol	difference between	source in direct space	source in reciprocal space
$\rho_{IMM} - \rho_{crystal}$	molecular electron densities	primary wavefunction (Hartree-Fock or ELMO calculation)	
$\|\varphi_{IMM}\|^2 - \|\varphi_{crystal}\|^2$	squared ELMOs	primary wavefunction (ELMO calculation)	
$\rho^{XR}_{IMM} - \rho^{XR}_{crystal}$	molecular electron densities	XR-Hartree-Fock or XR-ELMO calculations	primary structure factors
$\|\varphi^{XR}_{IMM}\|^2 - \|\varphi^{XR}_{crystal}\|^2$	squared ELMOs	XR-ELMO calculations	primary structure factors
$\rho^{XR}_{\lambda=0} - \rho^{XR}_{\lambda>0}$	molecular electron densities	XR-Hartree-Fock or XR-ELMO calculations	primary or measured structure factors
$\|\varphi^{XR}_{\lambda=0}\|^2 - \|\varphi^{XR}_{\lambda>0}\|^2$	squared ELMOs	XR-ELMO calculations	primary or measured structure factors (but not recommended for simulated primary ones)

(namely, reciprocal space information) as restraints. Subsequently, the direct space densities obtained from the wavefunctions calculated with and without the external electric field were subtracted as shown in Figure 9.2.

At the terminal sites of the molecule, the electrons are shifted according to the electrostatic field direction. In the central part of the molecule (around the carbon atom), instead, the opposite happens to compensate for the effect at the extremities. This illustrates a classical through-bond reorganization of the electron density in response to the external electric field. [111]

This is the behaviour obtained for the direct-space calculations. If the XRW procedure works, the same behaviour should be observed in XRW electron densities. For this reason we computed structure factors for the three wavefunctions (no-field, positive electric field corresponding to the left side of Figure 9.2, and negative electric field corresponding to the right side of

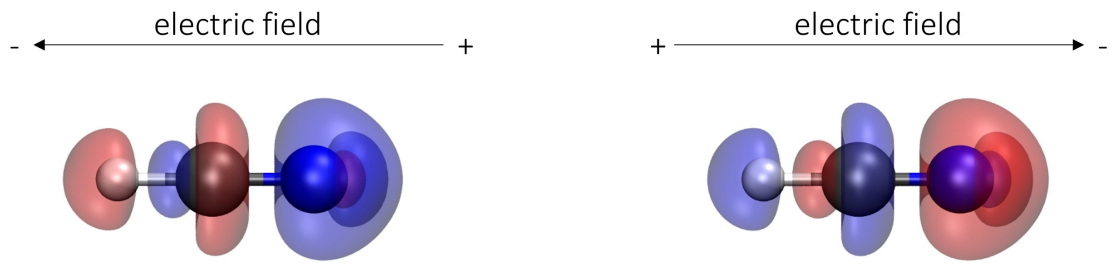


Figure 9.2 – Electron density differences in direct space. The compared densities correspond to a wavefunction calculated without any field and to wavefunctions resulting from computations with an electric field applied along the molecular axis ($\rho_{no\ field} - \rho_{field}$). The plotted isosurface corresponds to 0.001 e/bohr^3. Blue regions indicate an accumulation of electron density when the electric field is applied, whereas red regions indicate depletion of electron density when the electric field is applied.

Figure 9.2). As shown by Genoni *et al.* [64], the diffraction resolution of the structure factors plays an important role. Therefore, structure factors were simulated up to $\sin\theta/\lambda = 1.8\,\text{Å}^{-1}$ and X-ray restrained wavefunctions were calculated using four different resolution shells (0.0-0.3, 0.0-0.7, 0.0-1.2, 0.0-1.8 Å^{-1}).

Figure 9.3 shows electron density differences $\rho^{XR}_{no\ field} - \rho^{XR}_{field}$ obtained from X-ray restrained wavefunctions at Hartree-Fock and ELMO levels. These XRWs were calculated by fitting to primary structure factors that include or exclude the effects of the electric field. In analogy to the direct space calculations, the electric field was also applied in both directions.

In general, the effect is comparable to the one obtained through the direct-space calculations, although it is damped (the isosurface level being 0.0002 e/bohr^3 in Figure 9.3 compared with 0.001 e/bohr^3 in Figure 9.2). This is due to the fact that XRW calculations by definition only partially include the effect of the electric field. In fact, while the second part of equation 9.5 contains this effect (through reciprocal space fitting), the first part (energy minimization) does not. This holds true not only for the electron density but also for quantities derived from it, such as for the dipole moment shown in the appendix in Figure C.1.

Furthermore, the diffraction resolution plays an important role for the electrostatic field as it does also for including electron correlation [64]. In microscopy, a higher resolution is synonymous of higher accuracy in the image formation. However, in diffraction mode, where the image comes from optimizing a model against the measured structure factors, the higher accuracy is non-homogeneously spread. Core electrons dominate the high scattering angle, while valence electrons, on which the effect of the electric field (as well as of electron correlation) is the largest, scatter only at low angles. Therefore, including higher angle data reduces the weight of valence electrons and of their density. This pitfall is more pronounced for weaker deformations as those produced by an external electric field or by electron correlations, whereas it is less harmful for electron polarizations due to chemical bonding, because the effects are much larger. This explains the decreased magnitude and clarity of the density

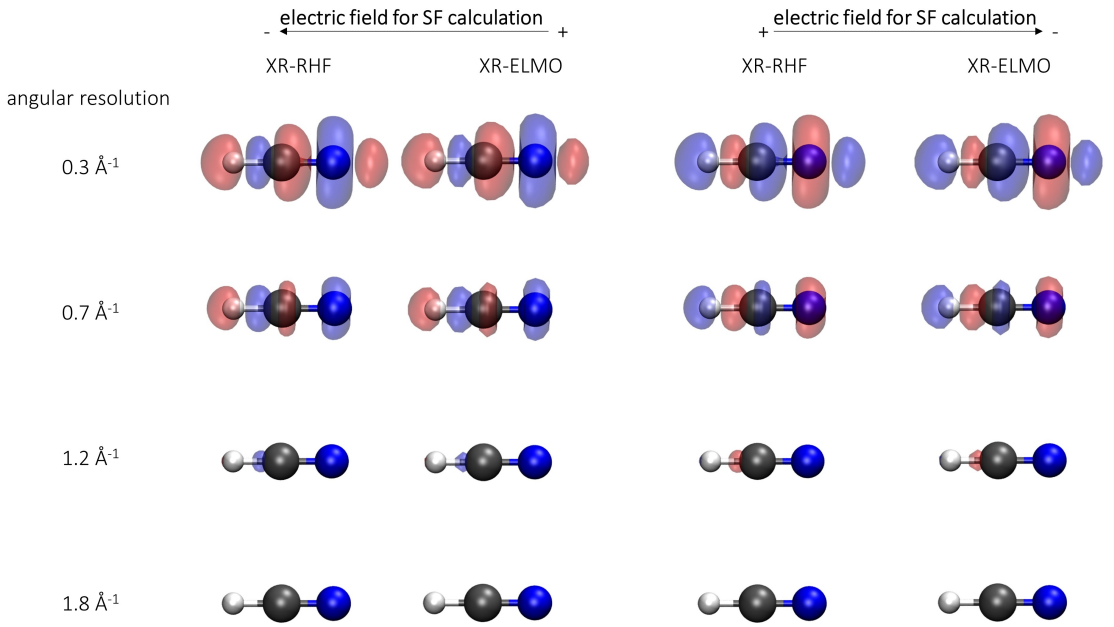


Figure 9.3 – Density differences obtained by comparing XRW calculations with SF including and excluding the electric field ($\rho^{XR}_{no\ field} - \rho^{XR}_{field}$), both at Hartree-Fock and ELMO level; blue regions indicate an accumulation of electron density for XRW calculations with SF including the electric field effects. For all the plots, the isovalue is 0.0002 e/bohr^3.

differences in Figure 9.3 with an increasing resolution. Increasing the strength of the simulated electric field from 0.01 a.u. to 0.1 a.u. qualitatively leads to similar results, as shown in the appendix in Figure C.2. An interesting feature is the lobe on the right side of the nitrogen nucleus, which, in Figure 9.2, is embedded in a larger and opposite-sign lobe. This seems to be an artefact arising from the transformation through reciprocal space and back to direct space. It is likely due to a poor description of the very diffuse density that contributes to the scattering only at very low resolution (hardly sampled by a medium/small-size unit-cell in direct space and therefore invisible).

Figure 9.4 shows the results of X-ray restrained Hartree-Fock calculations using a set of structure factors up to only 0.3 Å^{-1} for different values of λ. When the isosurface levels are chosen as in Figure 9.4, it becomes very clear that different λ-values affect the results only quantitatively, but not qualitatively. It is interesting to notice that, above a certain λ-value, the isovalue does not need to be further increased. It seems that, at that point, increasing the weight of the reciprocal space fitting does not significantly change the result.

In Figure 9.3, it is easy to observe that the density differences resulting from XR-ELMO calculations are completely analogous to those obtained through traditional X-ray restrained Hartree-Fock calculations. Therefore, the XR-ELMO results are trustable and we can analyse the orbital variations in the same way as those of the electron densities, as shown in Figure 9.5. As already mentioned in the section dedicated to the description of the computational protocol, the obtained orbitals were always squared before being compared (subtracted).

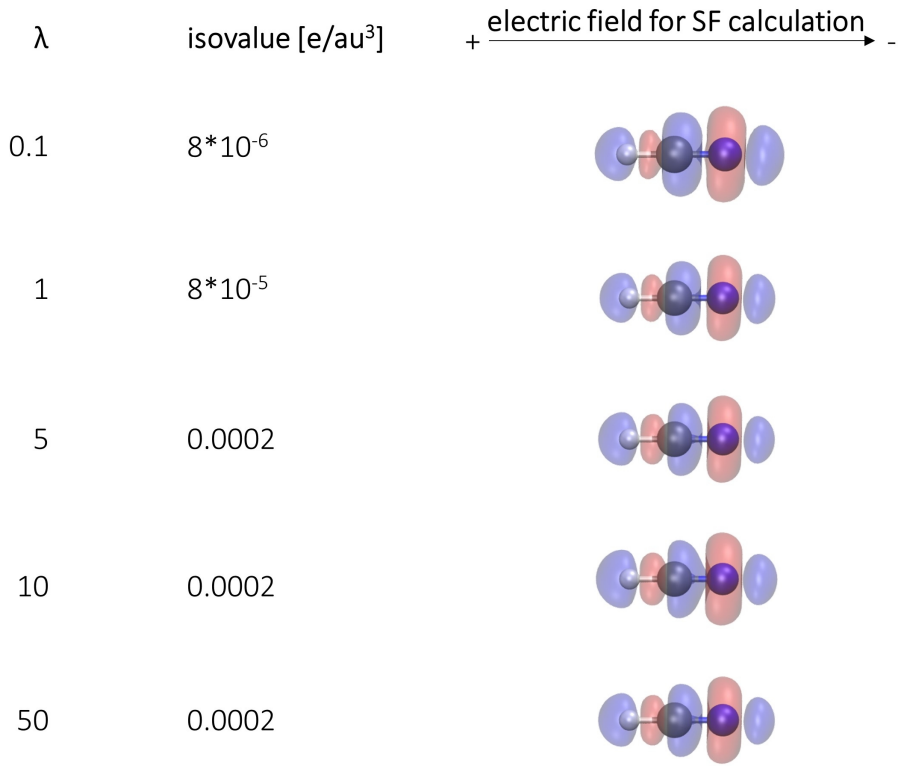


Figure 9.4 – Density differences obtained by comparing XR-RHF calculations with SF including and excluding the effects of the negative electric field. Blue regions indicate an accumulation of electron density for XR-RHF calculations with SF including the effects of the electric field. The isovalue is adjusted to different λ-values in order to highlight the qualitative similarity among the different cases.

In Figure 9.5, within each set of four, the uppermost picture always refers to the nitrogen lone pair orbital, the second one to the C-H bond orbital, the third one to the C-N σ bond orbital and the lowest one to one of the two π orbitals for the C-N bond. While the two π orbitals behave in the same way, the σ one only slightly changes because it describes electrons that are less polarizable. For all the three resolutions and for both fields, the electrons in the C-N π orbital move towards the positive direction of the electric field. The electrons in the C-H bond also move in an analogous way, whereas the ones in the N-lone pair orbital show an opposite behaviour, probably for compensation effects. Noteworthy, this dual behaviour would not be revealed so straightforwardly by canonical Hartree-Fock orbitals.

From this first and very simple example on HCN in an electric field, we can draw already some conclusions: a) the electron density on the terminal atoms N and H complies with the electric field while the electron density on the carbon atom behaves oppositely for a compensation effect (see Figure 9.2 and 9.3); b) as for correlation, lower diffraction resolution is better; c) XR-ELMOs analysis allowed to reveal the effects of the electric field from a localized molecular orbital-based point of view and, consequently, to rationalize these effects in terms of molecular fragments. This would be impossible from an analysis on the overall electron density.

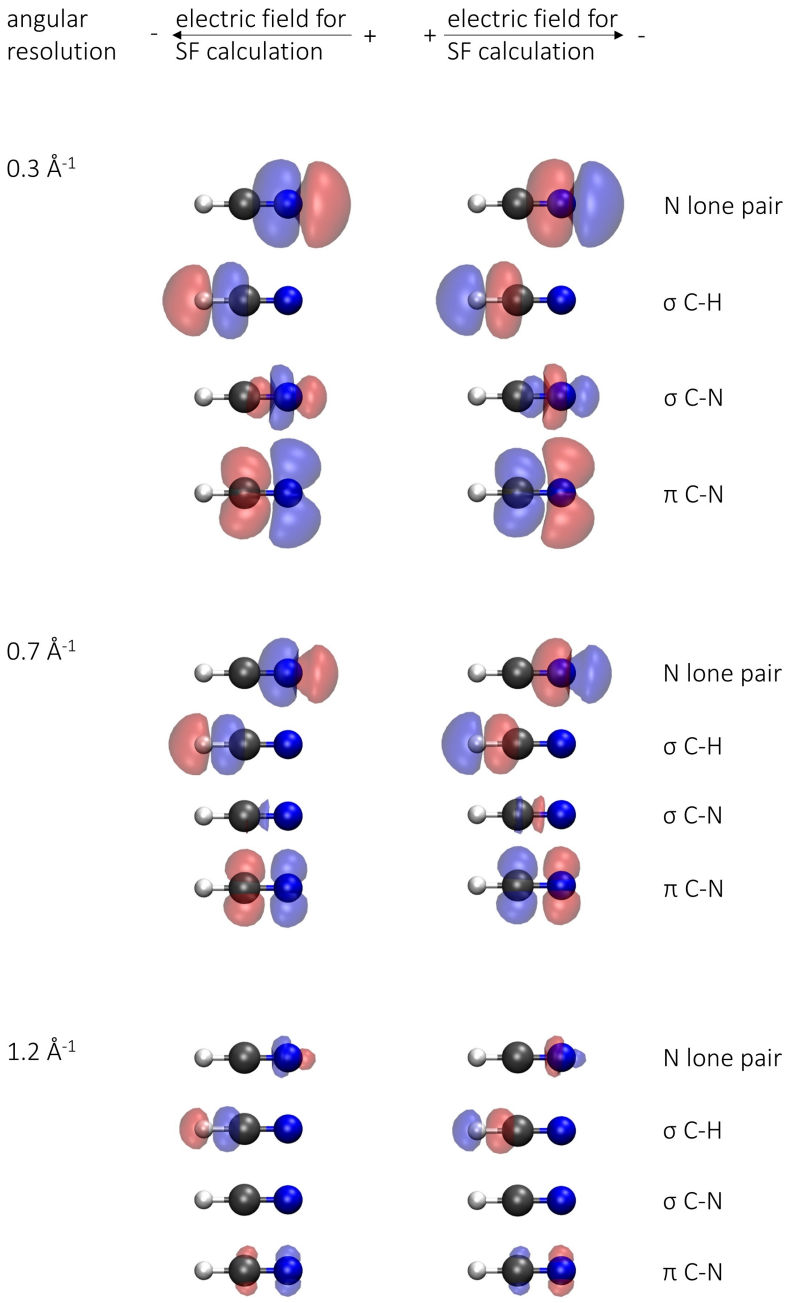


Figure 9.5 – Orbital differences resulting from the comparison of XR-ELMO calculations with SF including and excluding the electric field effects: $(|\varphi^{XR}_{no\ field}|^2 - |\varphi^{XR}_{field}|^2$; blue regions indicate an accumulation of electron density for XRW calculations with SF including the effects of the electric field. The isovalue is set to 2×10^{-5} e/bohr^3).

HCN - crystal field

The next step of our investigation consisted in simulating an entire crystal, where both long and short-range interactions were simultaneously present. A hypothetical crystal based on this molecule was constructed where all molecules are aligned as described in the Compounds section 9.2.2. As subsequently explained in the Computational protocol, we computed structure factors based on a periodic wavefunctions representing this crystal. Simulated structure factors have several advantages compared to experimental ones: a) thermal effects, which increase the complexity of the problem, are not included, b) any packing can be simulated, even those that are not observed in practice, c) calculated structure factors are not affected by experimental errors.

There is an additional motivation of using theoretical simulations: one can explicitly include and exclude some of the effects that we want to study. In particular, we performed calculations for a crystal with normally interacting molecules and one excluding the mutual interaction of molecules (using the MOLSPLIT option in *CRYSTAL17*) in a sort of "independent molecule model" (hereinafter called IMM). For XR-ELMO calculations, this option is useful because the simulated structure factors used as constraints are obtained from periodic Hartree-Fock wavefunctions. Therefore, an XR-ELMO calculation at a given $\lambda > 0$ would also fit (compensate for) the error introduced by the use of ELMOs. In other words, a calculation at $\lambda > 0$ would include the effect due to the crystal field *and* the fitting to canonical Hartree-Fock molecular orbitals with ELMOs. This makes XR-ELMO fitting to IMM structure factors important for comparison, since this also fits structure factors from canonical HF calculations but does not include intermolecular interactions. To exclude any bias due to the choice of the basis set, all the structure factors were simulated preforming periodic RHF calculations with the same set of basis functions as used for the XR-RHF and XR-ELMO calculations.

In order to judge how the crystal environment and specific interactions in the crystal influence the molecule, electron densities obtained from XR-RHF calculations based on structure factors including crystal interactions were subtracted from electron densities resulting from XR-RHF calculations fitting IMM structure factors, as shown in the difference-density plot at the bottom left of Figure 9.6. On the right side we also show the same differences for four selected XR-ELMOs (which are analogous to those considered in Figure 9.5), calculated using simulated structure factors up to 0.7 $Å^{-1}$. Overall, the picture resembles the one where an electric field with orientation from H to N was applied (physical definition, with positive charge at N and negative charge at H; see Figure 9.5 for comparison). The similarity can be well explained by the fact that the molecule indeed feels this electric field in the crystal due to the dipole moment of the neighbouring molecules. The dominating blue lobe might be due to the hydrogen bond interaction, which causes electrons on the nitrogen to shift towards the hydrogen atoms of the adjacent molecule, thus reducing the red feature. Overall, the changes in the density and in the ELMOs under the effect of the crystal electric field are very small in absolute values, thus justifying the use of theoretical ELMOs as building block in the current version of the ELMO-databases [288, 289, 290]. Nonetheless, when plotted at a low iso-contour level, differences

are visible, which point out the nature of the crystal environment. Thus, the use of structure factors from an independent molecule model (IMM) calculation can give valuable insights into these interactions.

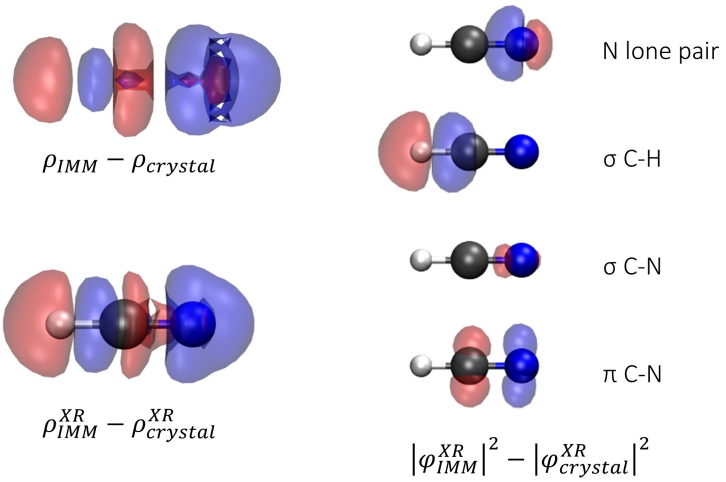


Figure 9.6 – Top left: density differences resulting from the comparison of the direct space IMM and crystal calculations ($\rho_{IMM} - \rho_{crystal}$; iso-contour level set to 0.0008 a.u.) Bottom left: density difference resulting from the comparison of the XR-RHF calculation using IMM structure factors with the XR-RHF calculation using crystal structure factors ($\rho^{XR}_{IMM} - \rho^{XR}_{crystal}$; iso-contour level set to 4×10^{-5} e/bohr^3, blue=negative and red=positive). Right: orbital differences resulting from the comparison of the XR-ELMO calculation using IMM structure factors with the XR-ELMO calculation using crystal structure factors ($|\varphi^{XR}_{IMM}|^2 - |\varphi^{XR}_{crystal}|^2$; isocontour level set to 2×10^{-5} e/bohr^3, blue=negative and red=positive).

9.3.2 Oxalic acid dihydrate

Finally, we examined the α-phase of oxalic acid dihydrate (OAD), featuring several strong hydrogen bonds, among which the strongest ones are transformed into bonds when the crystal gets compressed due to an external pressure. [294] From the six hydrogen bonds formed by one molecule (see Figure 9.7), three are symmetry-independent. The shortest hydrogen bond at ambient pressure (1.464 Å) is the one formed by the carboxyl-hydrogen with the neighboring water molecule. It shortens with decreasing temperature [296] and with increasing pressure, while the O-H bond of the hydrogen bond involved in this contact elongates until the proton shifts towards the water at 5.3 GPa. Table 9.3 shows the hydrogen bond distances for OAD measured at different pressures.

In analogy to the HCN case, periodic wavefunctions were calculated for α-OAD for the experimental structures at 0, 3.6, and 5.3 GPa and used to simulate structure factors. The IMM was used to obtain a set of structure factors where no intermolecular interactions are present.

For the X-ray restrained wavefunction fitting calculation, the setting reported in Table 9.1 was chosen in order to calculate one entire (and symmetrical) molecule of oxalic acid and,

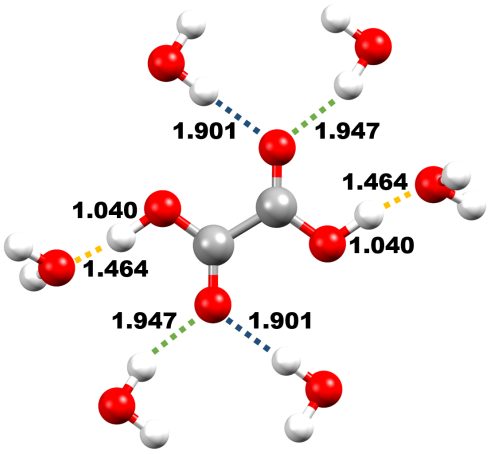


Figure 9.7 – Hydrogen bonds of oxalic acid dihydrate for the ambient pressure experimental structure. Distances are given in Ångström.

Table 9.3 – Bond and interaction distances (in Å) at different pressures

	C=O···HOH (blue)	C=O···HOH (green)	OH···OH_2 (orange)	O-H (solid line)
0 GPa	1.901	1.947	1.464	1.040
3.6 GPa	1.690	1.693	1.331	1.107
5.3 GPa	1.647	1.637	1.065	1.366

necessarily, two molecules of water. These water molecules were properly chosen far from the oxalic acid (see Figure 9.8) to avoid explicit strong hydrogen bonds, shown in Figure 9.7. Therefore, the hydrogen bonds are activated only through the fitting of the structure factors calculated from periodic calculations on the experimental crystal structure but not in the direct space calculation of the wavefunction of the supramolecular fragment (see again Figure 9.8). This choice enabled us to better highlight the influence of the structure factors on the oxalic acid not directly affected by the water molecules in the X-ray restrained wavefunction calculation.

Figure 9.9 shows the density differences obtained by comparing XR-RHF calculations using IMM structure factors to XR-RHF calculations exploiting usual crystal structure factors (always $\rho^{XR}_{IMM} - \rho^{XR}_{crystal}$). There, the influence of the water molecules only considered in the structure factors part is clearly visible. In fact, when fitting to crystal structure factors, there is less electron density on the carboxylic hydrogen because the oxygen atoms of the water molecules are binding this site and, therefore, they push the electron density away. An effect is also visible on the carbonyl group, where a polarization occurs and electrons at the carbon atom are pushed away towards the oxygen, because the oxygen atoms (an electron donor) is accepting a weak hydrogen bond from the water molecule. For the sake of comparison, in Figure C.3 in the appendix, we show the differences obtained by subtracting the electron density corresponding

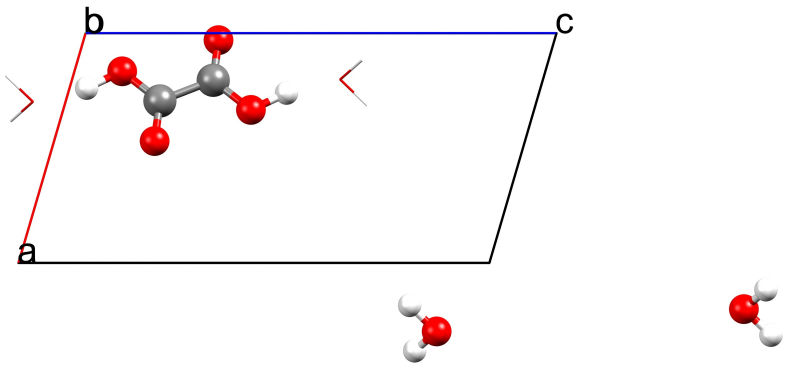


Figure 9.8 – Fragments chosen for the XR-RHF and XR-ELMO calculation (in ball-and-stick representation). The water molecules are far away from the oxalic acid, such that their interactions are not taken into account in the pure wavefunction calculation. Their new positions were obtained by shifting the original ones (here shown with sticks) one unit-cell length in all three directions.

to a simple wavefunction calculation (RHF/6-1G(d,p)) on the oxalic acid with two water molecules as close as in the crystal structure at 0 GPa, to the electron density associated with a simple wavefunction calculation (RHF/6-31G(d,p)) on the oxalic acid with two water molecules that are 5 Å apart from the oxalic acid. This was done to show that the observed density differences are analogous to those seen in Figure 9.9 and, therefore, that they are indeed due to the short-range hydrogen bonds mentioned above. At 3.6 and 5.3 GPa the difference increases as shown in Figures 9.9b and 9.9c.

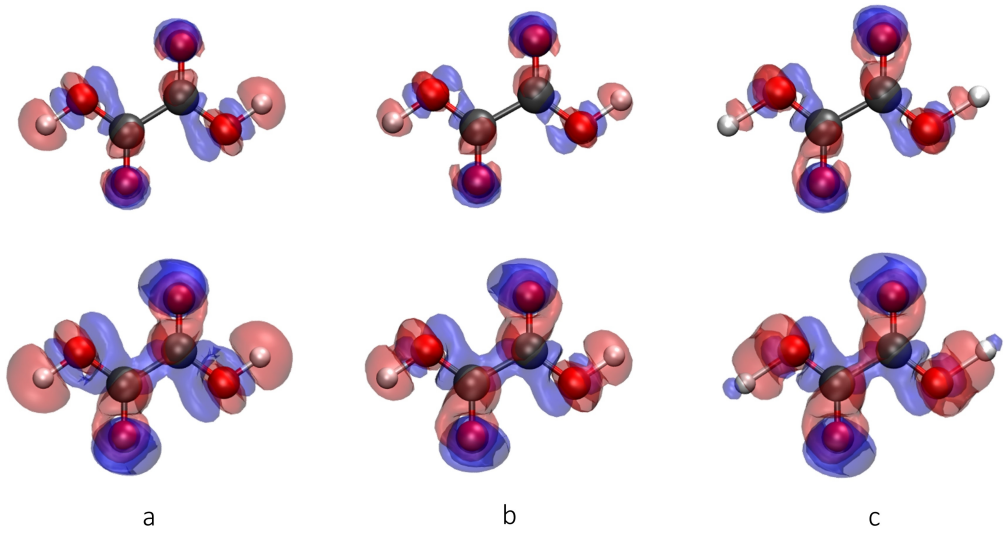


Figure 9.9 – a) Electron density differences resulting from the comparison of XR-RHF calculations using IMM structure factors wth XR-RHF computations exploiting crystal structure factors. ($\rho^{XR}_{IMM} - \rho^{XR}_{crystal}$; isosurfaces: 0.002 e/bohr^3 (top row) and 0.0008 e/bohr^3 (bottom row)). Red and blue designate positive and negative regions, respectively. b) and c) analogous results obtained for the crystal structures at 3.6 GPa and 5.3 GPa.

When analysing the XR-ELMOs of Figure 9.10, one can see that the O-H σ-type orbital shifts electrons away from the hydrogen atom towards the covalently bonded oxygen due to the influence of the hydrogen bond to the neighbouring water molecule present in the crystal. An interesting feature appears at 5.3 GPa. At this pressure, the bond to the water is established and, therefore, there are again more electrons in this region, as one can evince from the small blue lobe. On the contrary, the region between the hydrogen and oxygen molecule is characterized by a depletion of the electron density. In carbonyl orbitals one can observe a shift of electrons from carbon to oxygen due to the weakening of the double bond in favour of additional interactions that the oxygen atom is forming.

Thus, only looking at the molecule itself and not knowing the interactions in the crystal, small changes in the XR-ELMOs can indicate which parts of the molecule are involved in the interactions. From the electron density and the σ O-H bond orbital at 5.3 GPa it also seems that this ELMO is able to indicate where a bond is being formed.

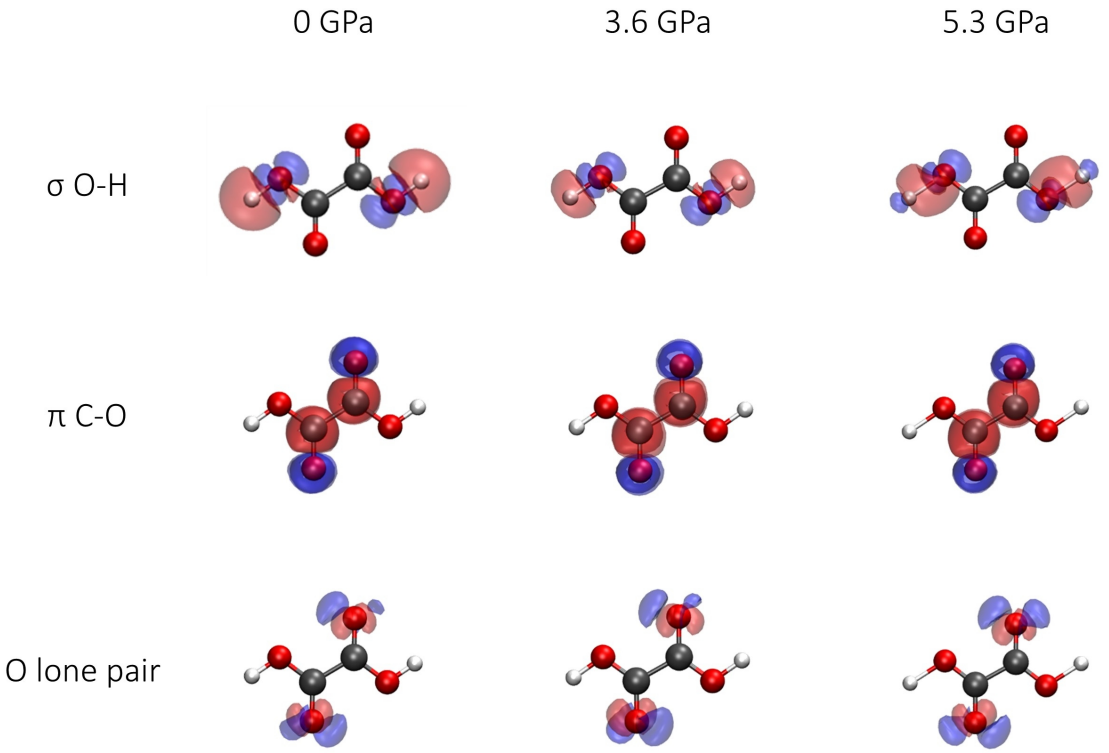


Figure 9.10 – Top row: differences for the XR-ELMOs corresponding to the two O-H bonds; middle row: differences for the XR-ELMOs associated with the two carbonyl π orbitals; bottom row: differences for the XR-ELMO corresponding to one of the oxygen atom lone pairs. In all cases we depicted the differences $|\varphi^{XR}_{IMM}|^2 - |\varphi^{XR}_{crystal}|^2$ with the iso-contour level set to 0.0002 e/bohr^3, blue=negative and red=positive. The differences are shown for the computations performed on the crystal structures at 0 GPa, 3.6 GPa, and 5.3 GPa.

For the sake of completeness, in Figure 9.11 we also show the electron density differences resulting from the comparison of an unrestrained RHF calculation (λ=0.0) and of an XR-RHF

computation (λ=0.5 and experimentally measured structure factors as restraints) on the crystal structure of the oxalic acid at ambient pressure. [294] It is possible to see that the features of the electron density are very similar to those in Figure 9.9a. They are only slightly less pronounced. This can be attributed to the fact that an experimental dataset also contains temperature effects and noise. The analysis of the orbitals is thus more difficult. Nevertheless, also in this case, for the ELMOs corresponding to O-H bonds involved in hydrogen-bond contacts, we can observe the same effects already highlighted for the simulated structure factors.

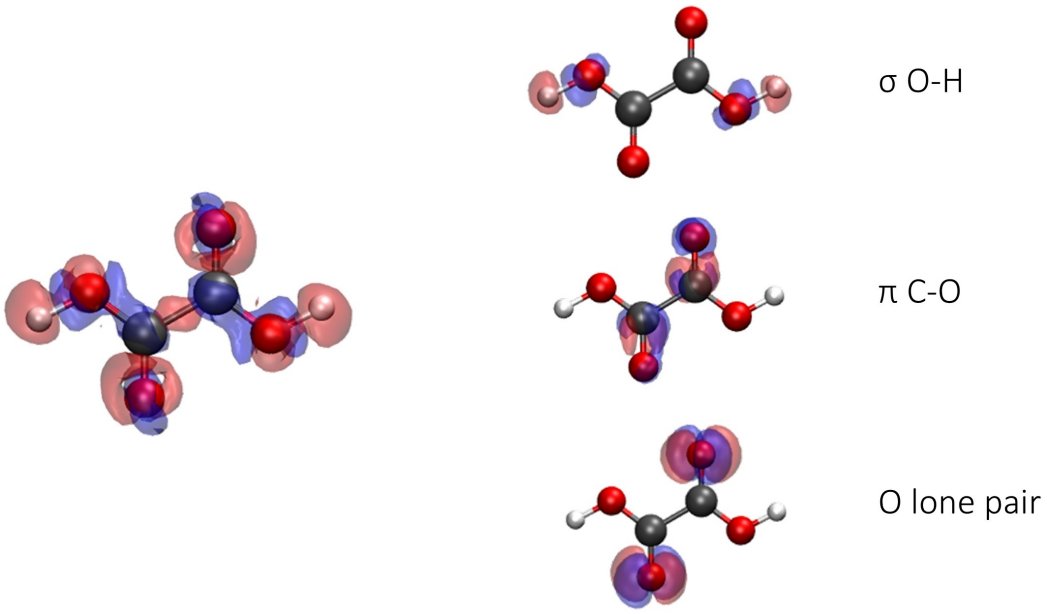


Figure 9.11 – Left: density differences resulting from the comparison of the unrestrained RHF/6-31G(d,p) calculation with the XR-RHF/6-31G(d,p) computation (with λ=0.5 and experimental structure factor magnitudes) on the crystal structure of OAD at ambient pressure ($\rho^{XR}_{\lambda=0} - \rho^{XR}_{\lambda=0.5}$; isosurface: 0.004 e/bohr^3, blue=negative and red=positive); right: corresponding ELMO differences ($|\varphi^{XR}_{\lambda=0}|^2 - |\varphi^{XR}_{\lambda=0.5}|^2$; isosurface: 0.004 e/bohr^3, blue=negative and red=positive).

9.4 Conclusions

This work focused on X-ray restrained wavefunction fittings, both with canonical orbitals and with extremely localized molecular orbitals. The main finding was that X-ray restrained wavefunction techniques are able to account for the electron density and molecular orbital deformations caused by intra-crystal electric fields. We simulated structure factors calculated with or without an applied electric or crystal field. If the crystal field is switched off, but the wavefunction is anyway periodic (for example using the option MOLSPLIT of *CRYSTAL17*), we generate an *independent molecule model*. This provides a benchmark for molecular or

orbital density differences, very useful to highlight the effects of the electric field. These density differences are generally small, thus, justifying the use of purely theoretical ELMOs in databases [274, 288, 289] constructed for improving the refinement of macromolecular crystal structures. [293]

The diffraction resolution plays a role similar to the one already observed for correlation [64]: high resolution data do not allow a better recovery of the electric field effect but a worse one. This can be explained with the down-weighting of the low-angle reflections, which are important for a good modelling of the valence electrons.

With the example of HCN under the influence of an electric field, it was shown that not all XR-ELMOs polarize in the direction of an applied electric field. Some of them are actually polarized in the opposite direction. The method was also successful in the analysis of the intermolecular interaction effects for the more complicated system of the oxalic acid dihydrate. In that case it was possible to highlight the influence of the hydrogen bond contacts on the electron density of the oxalic acid molecule, whose effects become more important as pressure increases.

Furthermore, for OAD, both experimental and simulated structure factors were used. Therefore, in light of the obtained results, the approach can also be exploited when measured X-ray data are available, although the pictures might become slightly less clear because experimental structure factors intrinsically contain more information than the theoretical ones and are affected by experimental errors. Further applications and tests of the technique, also using challenging experimental cases, are already envisaged.

9.5 Addendum

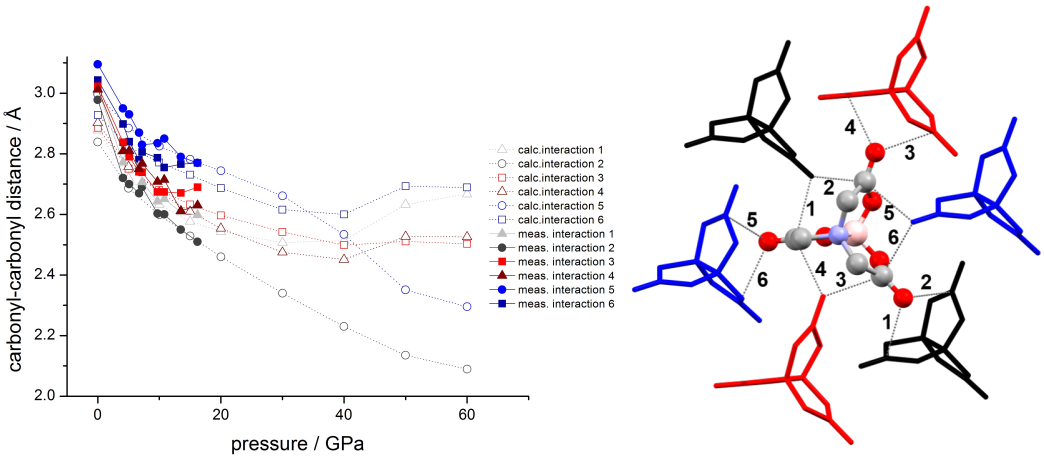
In addition to the compounds presented in the paper, the method was also applied to boron nitriolotriacetate (NTA-B). This molecule has three carbonyl groups, which interact with carbonyl groups of neighbouring molecules in the crystal. Crystals of this compound were synthesized and measured at different pressures. The synthesis and the data collection at high pressure (HP) are described elsewhere. [297] The crystal packing is dominated by short contacts between carbonyl groups, some of which have angles not far away from the Bürgi-Dunitz angle [298] and are promising candidates for HP reactivity through nucleophilic addition and thus possible polymerization. Each carbonyl O and C atom partake in two such interactions, resulting in a very efficient packing. The HP behaviour of NTA-B was studied experimentally and computationally.

9.5.1 Boron-nitrilotriacetate high pressure results

Boron-nitrilotriacetate (NTA-B) belongs to the family of boratranes [299]. Atranes are characterized by a transannular dative bond between a bridgehead nitrogen atom to a Lewis acidic metal atom such as boron, silicon, or phosphorus at the other bridgehead. The electronegativity difference between the bridgeheads atoms determines the strength of the transannular interaction.

The crystal structure of NTA-B at ambient pressure was investigated in the 1980s by Bürgi *et al.* [300, 301] The compound crystallizes in the monoclinic $Pna2_1$ space group forming transparent plates. The molecules show non-crystallographic C_{3v} symmetry, with planar five-membered rings. Each of the three carbonyl groups on every molecule interacts with two carbonyl groups of neighbouring molecules leading to six different kinds of carbonyl-carbonyl interactions as shown in Figure 9.12 (numbers in Figure 9.12b). The oxygen atoms are electronegative and the carbon atoms electropositive. Due to the carbonyl-carbonyl intermolecular interaction, electrons are polarized from the lone-pair orbitals of the oxygen atoms into the π^*-orbital of the other carbonyl group. The $\pi^* \leftarrow n$ interactions between lone-pair orbitals on the electronegative oxygen atoms and the π^*-orbitals of the neighbouring electropositive carbon atoms show promise for polymerization at HP, via classic nucleophilic addition reactions. In order to examine this, NTA-B was measured at different pressures up to 16 GPa and the structure was simulated with periodic DFT calculations up to 60 GPa (see appendix C.4).

The calculated C=O$\cdots$C=O interaction distances (shown in Figure 9.12a) are in good agreement with the measured ones. The mismatch at ambient pressure is due to the different temperature of experimental and calculated data (273 and 0 K respectively) and probably to an overcorrection of the dispersion forces, but these factors become irrelevant at higher pressure, where the *pressure*$*$*volume* term dominates. The shortest C=O$\cdots$C=O distance decreases from 2.84 Å at ambient pressure to 2.09 Å at 60 GPa. This is the only interaction where a bond critical point (BCP) was present over the entire pressure regime. However, the electron density



(a) Intermolecular distances

(b) Nomenclature for intermolecular contacts

Figure 9.12 – a) Experimental (filled symbols) and calculated (empty symbols) distances for the six carbonyl-carbonyl interactions in NTA-B. Similar colour indicates that the same oxygen atom is involved in the interaction and the same symbol indicates that the same carbon atom is involved. b) The six different carbonyl-carbonyl interactions in NTA-B are marked with dotted black lines and the nomenclature is specified.

at this BCP is much lower than it is in structures with polymerized C=O bonds. Thus, despite the distance being extremely short, the expected polymerization was not observed in this pressure range. This may be due to the fact that the two most compressible interactions in the 40-60 GPa range involve the same carbon atom and are thus competing. The absence of bond critical points over all pressure ranges, an established criterion to examine interactions, speaks for the use of other analysis tools such as the ELMOs involved in the interaction.

Other types of interactions and short contacts are also present. The oxygen atoms from the carbonyl groups are also involved in C=O···B contacts, and there are several CH···O hydrogen bonds, where the involved oxygen may be the heterocyclic or the carbonylic one, and also short O···O contacts are present.

9.5.2 XRW and XR-ELMO calculations on simulated and measured structure factors of NTA-B

The top left picture of Figure 9.13 shows the electron density difference from XRW fitting to IMM structure factors and to crystal structure factors for the 0 GPa structure of NTA-B. The influence from carbonyl groups of neighbouring molecules on the electron density is visible by the blue lobes around the oxygen and the red ones around the carbon atom of the carbonyl group. Electrons from the depicted carbonylic oxygen atoms get polarized towards the carbonylic carbon and the hydrogen atoms of neighbouring molecules. This, in turn, causes the electron density of the carbonyl group to be polarized from carbon towards oxygen,

thus the relative electron deficiency on the carbon atom and the abundance of electrons on the oxygen atom. This effect slightly increases with the pressure and thus reflects the shortening of the interaction distance, visible from the surfaces with the higher isovalue in appendix C.5. However, it is not visible in which direction the carbonyl-carbonyl interactions are exactly going, mainly due to two reasons: 1) Even when putting explicit point charges at the position of the neighbouring carbon atoms, the electron density does not simply polarize only in these directions as shown in appendix C.6. 2) Not only the carbonyl-carbonyl interactions are present in NTA-B but many interactions, thus substantially complicating the situation.

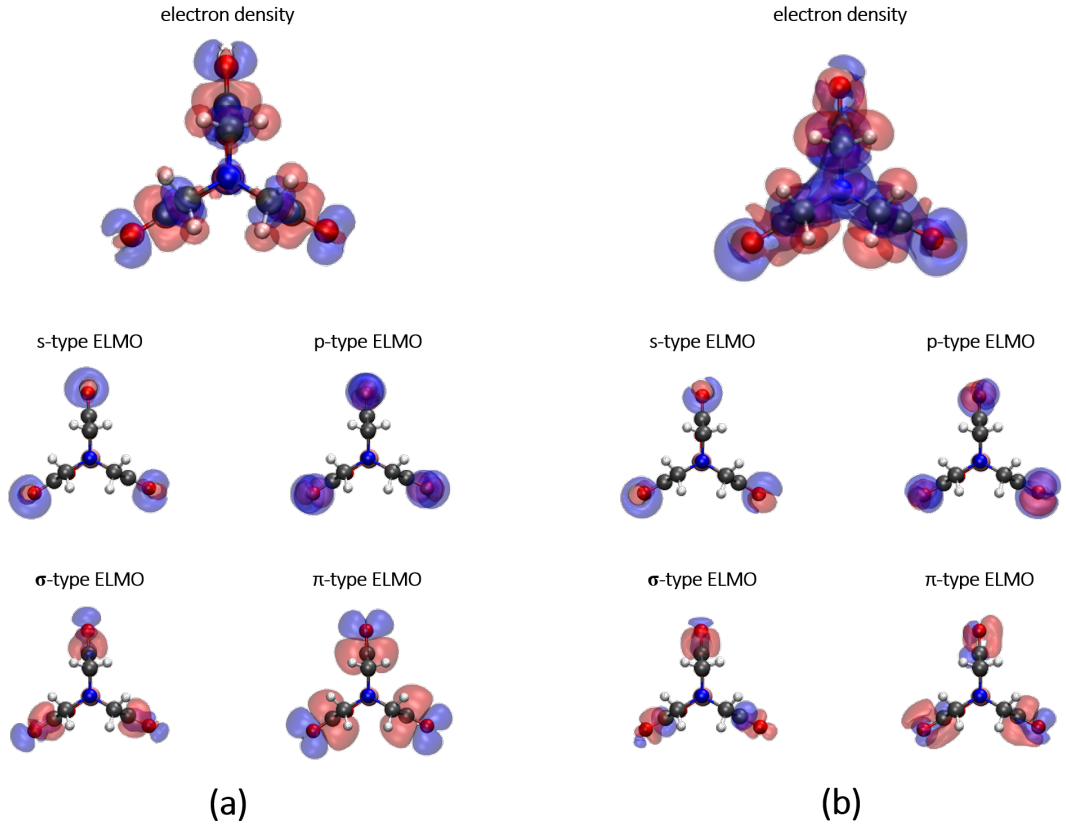


Figure 9.13 – a) Differences for calculated structure factors of NTA-B. On top is the electron density difference ($\rho^{XR}_{IMM} - \rho^{XR}_{crystal}$; iso-contour level set to 0.001 e/bohr^3) and on the bottom the XR-ELMO differences ($|\varphi^{XR}_{IMM}|^2 - |\varphi^{XR}_{crystal}|^2$; iso-contour level set to 4×10^{-5} e/bohr^3). b) Differences for measured structure factors. On top is the electron density difference ($\rho^{XR}_{\lambda=0} - \rho^{XR}_{\lambda>0}$; iso-contour level set to 0.001 e/bohr^3) and on the bottom the XR-ELMO difference ($|\varphi^{XR}_{\lambda=0}|^2 - |\varphi^{XR}_{\lambda>0}|^2$; iso-contour level set to 0.0004 e/bohr^3)

The bottom left part of Figure 9.13 depicts density difference maps for some selected squared XR-ELMOs, which were obtained with the same structure factor set as used for the calculation of the electron densities from canonical molecular orbitals. These occupied XR-ELMOs are involved in the carbonyl-carbonyl interactions (all three XR-ELMOs corresponding to the three carbonyl groups are always shown simultaneously). The lone-pair s orbitals on the oxygen

atoms get polarized outwards due to the influence of the neighbouring carbon atoms. The interpretation of the p lone pair orbitals on the oxygen atoms is more difficult but corresponds to the result obtained when taking the difference between a wavefunction where point charges were put around a carbonyl group and a wavefunction without charges (see appendix C.6). The σ-orbitals of the carbonyl bonds show a small polarization towards the oxygen atoms, whereas the π-orbitals are clearly polarized towards the oxygen atoms, thus showing the tendency to abandon the double bonds in favour of a single bond in the direction of the neighbouring carbonyl groups. The similarity of the XR-ELMOs of the same kind reflects the approximate C_{3v} symmetry of the molecule.

The pictures in Figure 9.13b result from XRW calculations with experimentally measured structure factors. Only the high-resolution data collected at ambient pressure can be used for this highly sensitive analysis, the quality of high-pressure data is not good enough. In case of experimentally measured structure factors, it is obviously not possible to use the IMM. Instead of that the density from an XRW calculation with $\lambda = 0.5$ was subtracted from the density obtained with $\lambda = 0$, which corresponds to a normal Hartree-Fock calculations. Calculations with $\lambda = 5$ did not converge. The electron density difference resembles the one obtained from the simulated structure factors, although the polarization from the carbonylic carbon atom towards the oxygen atom is less obvious and the picture is generally noisier. This can be attributed to the fact that the experimental data also contains temperature effects and noise of the measurement.
Interpretation of the repolarization within the individual orbitals is more difficult. The s and p-type orbital differences look similar to the theoretical results, whereas the shift of electrons in the π orbital from carbon to oxygen is less evident than in the case where simulated structure factors were used.

Although virtual orbitals are not directly restrained by structure factors (since they are unoccupied) they are indirectly restrained via the SCF calculation. Several unoccupied orbitals per fragment result from an XR-ELMO calculation and Figure 9.14 shows only the differences of the π^* orbitals on the three carbonyl groups. Due to the interactions of the carbon atoms (as an electron acceptor) with atoms from neighbouring molecules, this orbital polarizes in the opposite way as the π orbital.

In conclusion, one can say that NTA-B does not polymerize in the measured pressure regime and also the geometry optimized structures do not support that, despite the carbonyl-carbonyl interaction distances becoming quite short. The compound forms many different interactions with its neighbours in the crystal hindering a straightforward interpretation. Nonetheless, electron density difference maps and especially also ELMO difference maps clearly indicate a polarization within the carbonyl groups due to these interactions.

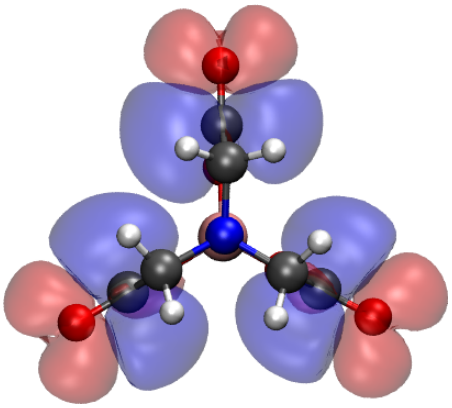


Figure 9.14 – Difference of the three carbonyl π^* orbitals resulting from the comparison of the XRW calculation using IMM structure factors with the XRW calculation using crystal structure factors ($|\varphi^{XR}_{IMM}|^2 - |\varphi^{XR}_{crystal}|^2$; iso-contour level set to 4×10^{-5} e/bohr^3).

10 Conclusions

The common motive of this dissertation is the identification of atoms, functional groups, and molecules in crystals, and how they interact with each other. If we understand how these building blocks behave individually and together, we can explain observed characteristics of crystals and we can create new crystals with desired properties. However, X-ray diffraction, as well as many other experimental techniques used on crystals, usually probe the entire bulk. Therefore, theoretical concepts are required, which allow us to define sub-crystalline and sub-molecular building blocks, as well as they enable us to study them.
But do atoms and functional groups in molecules [25, 302], bonds [303, 304, 305, 306], and molecules in crystals exist at all? These questions have been addressed by many researchers, and like all their publications, this dissertation does not provide any unique answer to them. Rather we assumed that such entities indeed exist, even if a precise definition is not possible. It is an enormous advantage to work with this concept and in fact the existence of some sub-molecular units with quite unchangeable properties can hardly be denied. Hence, we studied functional groups in molecules and molecules in crystals while carefully evaluating whether the assumption of such building blocks is sensible and justified.

In chapter 6 we examined the polarizabilities of various functional groups in different chemical environments. We found that a fully automatized clustering of these functional groups leads to very similar clusters as the ones we would define based on our chemical knowledge. This is due to the fact that functional groups keep their properties rather constant also in varying chemical environments, making them the most sensible chemical building blocks. Many chemical disciplines such as synthesis and analysis exploit this concept, but it is rarely used in the construction of databases for polarizabilities and other properties, which are often based on atoms. However, since the properties of a given atom are usually not as constant as the properties of a given functional group, it becomes necessary in those cases to define various atom types for the same element. We demonstrated that functional groups are very useful building blocks for the creation of a database, which can be used for a fast evaluation of properties. But there are also some issues such as the definition of the functional groups themselves, the definition of a proper local coordinate system, and a sensible categorization (clustering).

In chapter 7 we studied the behaviour of the polyiodide TEAI. At ambient pressure the compound consists, besides the cation, of interacting I_3^- and I_2 units, separated by around 3.5 Å. With increasing pressure the distances between them decrease until a polymer is formed. In consistency with that, the crystal was shown to become semiconductive at high pressure (around 10 GPa). However, we wanted to further understand and rationalize, how the iodine-iodine interactions change upon pressurization. For this purpose, we theoretically analysed I_5^-, the simplest case of an I_3^- unit interacting with an I_2. We found that, with decreasing distance, this interaction is transformed from a mainly electrostatic one into a more covalent one. According to the considered theoretical indicators, the transformation from well separated molecules into a polymer happens gradually, and it is therefore not possible to highlight one specific pressure or distance at which a bond can be clearly defined. This is true for both, I_5^- and the TEAI crystal.

For chapter 8, however, the definition of molecules is essential. The X-ray constrained wavefunction (XCW) approach is based on the assumption that molecules in crystals are well defined. For molecular crystals this assumption is justified, but also in that case, a crystal is more than the sum of its constituting molecules, because they interact with each others. In an XCW calculation this is taken into account via the structure factors of the crystal. We examined how much an electric or crystal field perturbs the electron density of a molecule under these conditions. Furthermore, we calculated the self-electrostatic energy of a molecule in a crystal based on the structure factors and used this expression to modify the standard XCW equations.

Chapter 9 expanded the analysis of field effects in XCW calculations to molecular orbitals. The idea was that the polarization of electrons due to a field can not only be studied for the total molecular electron density but also for each individual orbital, especially when making use of the concept of ELMOs. As expected, the ELMOs of a given atom or bond are quite constant in various chemical environments, and they are perturbed little by the structure factors from the crystal. Nonetheless, the influence of the crystal does change the ELMOs, and we could analyse the small differences to rationalize effects of intermolecular interactions and the crystal field on molecular orbitals as well as on the total electron density.

X-ray diffraction is one of the most widely adopted techniques in crystallography because the structure and, with good quality data, also the electron density can be obtained with it. But in order to really exploit this information, it has to be further processed, for example by characterizing the atoms and functional groups, or by studying the bonds and interactions between them. Such studies provide information about the link between the microscopic building blocks and the macroscopic properties of the crystal. Numerous different tools can be used for this analysis and in the previous chapters some of them have been exploited to gain information about various different compounds. Going beyond the standard crystallographic analysis is the cornerstone of quantum crystallography and this dissertation has hopefully illustrated some of its facets.

List of Acronyms

ADP	atomic displacement parameter
AO	atomic orbital
BCP	bond critical point
CCP	cage critical point
CP	critical point
CSD	Cambridge Structural Database
DFT	Density Functional Theory
DI	delocalization index
EDA	energy decomposition analysis
ELMO	extremely localized molecular orbital
FDET	Frozen-density embedding theory
FT	Fourier transformation
HAR	Hirshfeld atom refinement
HF	Hartree-Fock
HOMO	highest occupied molecular orbital
HP	high pressure
IAM	independent atom model
IMM	independent molecule model
IQA	interacting quantum atoms
LCAO	linear combination of atomic orbitals
MM	multipolar model
MO	molecular orbital
MOF	metal organic framework
NCP	nuclear critical point
NTA-B	Boron-nitrilotriacetate
PI	Polyiodide
QTAIM	Quantum Theory of Atoms in Molecules
RCP	ring critical point
RDM	reduced density matrix
RSR	real space R
SF	structure factors
TEAI	tetraethylammonium diiodine triiodide
UrPrI$_7$	N-propylurotropinium heptaiodide
WFN	wavefunction
XC-ELMO	X-ray constrained extremely localized molecular orbital
XCW	X-ray constrained wavefunction
XR-ELMO	X-ray restrained extremely localized molecular orbital
XRW	X-ray restrained wavefunction
XWR	X-ray wavefunction refinement

A Appendix to Chapter 7

A.1 Powder diffraction data of TEAI collected in epoxy-nano-alumina 4:1

Powder diffraction measurements: A crystal of TEAI with 80 x 60 x 80 μm size was loaded in a Merrill-Bassett DAC in quasi-hydrostatic medium - a mixture of α-nanoalumina and epoxy glue 4:1. Upon closing the cell, the single crystal was immediately crushed into powder by diamond anvils and the surrounding pressure-transmitting medium. Gold was used as a pressure calibrant [307]. High-pressure powder diffraction data was collected using synchrotron radiation (λ = 0.49407 Å, as calibrated by diffraction of the LaB_6 NIST standard [308]) at the Material Science Beamline at Swiss Light Source [240]. Diffraction patterns were indexed and refined using the Le Bail method [309] implemented in Topas 5 software [310].

A.2 Determination of the correct space group of TEAI

Crystal structures of tetraetylammonium di-iodine triiodide (TEAI) were previously investigated at ambient pressure (entries HALTAP and HALTAP01 in Cambridge Structural Database [216]). HALTAP [217] and HALTAP01 [218] were synthesized in the same manner as it was done in this study, however HALTAP01 was then recrystallized from EtOH/MeCN 1:1. HALTAP is described in Aba2 (non-centrosymmetric) group, while HALTAP01 is claimed to be in Cmca (centrosymmetric) group with the quarternary ammonium nitrogen lying on an inversion center and carbon atoms disordered with occupancies 0.5 in two positions. However, our findings do not support the latter. Analysis of the absolute structure of TEAI recrystallized from EtOH/MeCN 1:1 using likelihood methods [311] was performed using PLATON [312]. The results indicated that the structure is a racemic twin (calculated probability that it is a racemic twin equal to 1.000). The method calculated that the probability that the structure is inverted is smaller than 1×10^{-9}, and that the structure is not a twin smaller than 5×10^{-12}. The absolute structure parameter y [311] was calculated using PLATON. The resulting value was y=0.53(7) indicating that the absolute structure has probably been determined correctly

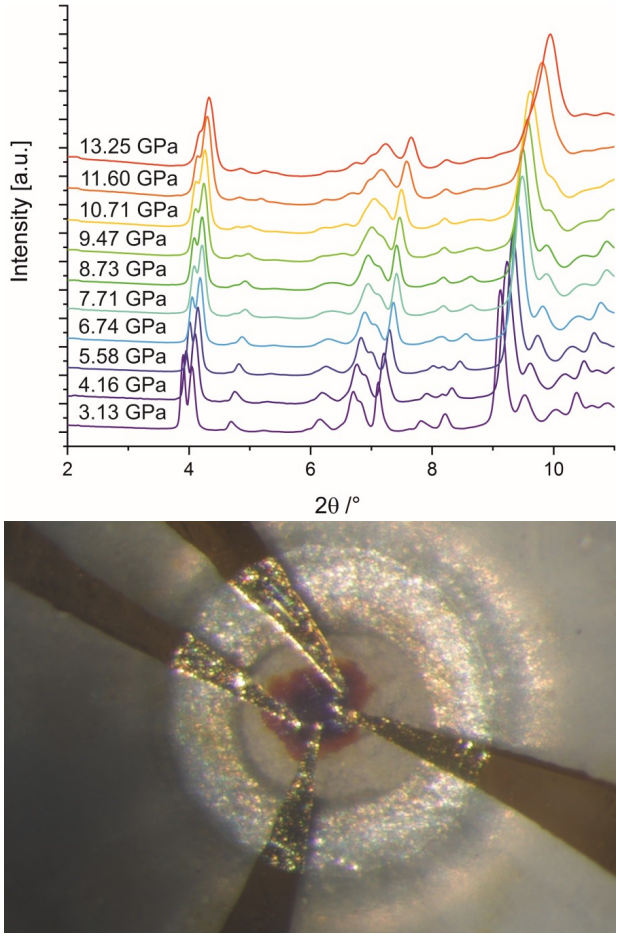


Figure A.1 – Top: Powder diffractograms of TEAI in 4:1 nanoalumina - epoxy. Bottom: photo of the measurement cell with four-probe setup (photo taken through the upper diamond).

as an inversion twin. Thus, TEAI crystallized from EtOH and the one recrystallized from EtOH/MeCN 1:1 are well-described by Aba2 symmetry as an inversion twin and this symmetry was further used for the high-pressure models. Nevertheless, the ratio between twins cannot be quantified by high-pressure measurement since it does not allow a reliable determination of the Friedel parameter.

A.3 Structural motifs of iodine chains in TEAI

Figure A.2 shows the structure of iodine chains in TEAI propagating through d_1 and d_2. The top left picture illustrates a zig-zag chain along the *c* axis, which, due to strong I···I interaction, hinders the axis compressibility. The top right picture shows the corrugated chain along *a* – it extends through I_3^- along the *b* axis forming bi-layers. These iodine bilayers stack on top of each other, stabilized by weak Van der Waals interactions (bottom picture). For this reason, the *b* axis is the most compressible one. In terms of relative compressibility the axes can be

ordered as: $b > a > c$, which is a direct consequence of the iodine chain geometry in a given crystallographic direction.

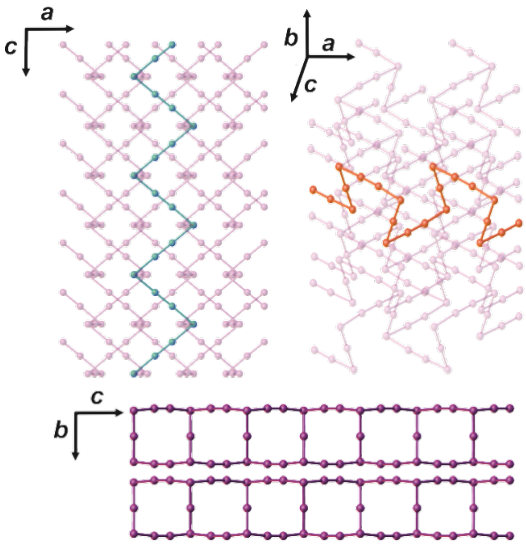


Figure A.2 – Structural motifs of iodine chains in TEAI

A.4 Bending of the structural motif in TEAI under high-pressure

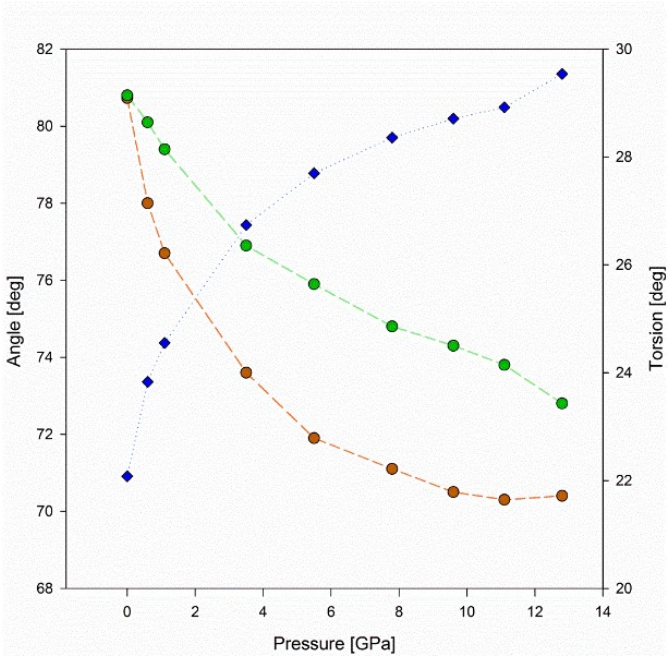


Figure A.3 – Decrease of α (orange) and β (green), with simultaneous increase in torsion between tri-iodides (blue).

A.5 Bonding analysis of I_3^-, I_5^- and I_7^- optimized in vacuo

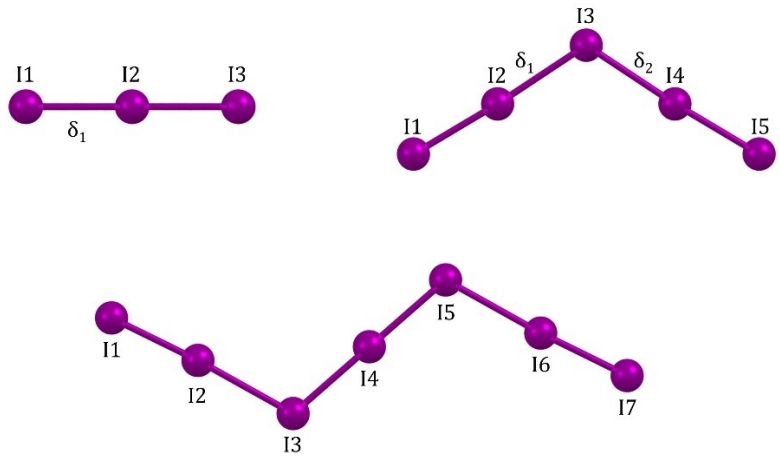


Figure A.4 – Optimized I_3^-, I_5^- and I_7^- with atom labels used throughout the text

I_3^- optimized *in vacuo* is linear and has two equal bond lengths of 3.04 Å. I_5^- has C_{2v} symmetry with bond lengths of 2.92 Å to the terminal iodine atoms (I1-I2 and I4-I5) and 3.16 Å to the central iodine atom (I2-I3 and I3-I4). The *in vacuo* optimized structure of I_7^- has C2 symmetry with I1-I2=2.87 Å, I2-I3= 3.25 Å, and I3-I4= 3.02 Å. The angles I2-I3-I4 and I4-I5-I6 are 109.84°.

Tables A.2 and A.3 list the bonding and interaction energies between the fragments in I_3^- (I_2 and I^- interacting), I_5^- (I_3^- and I_2 interacting) and I_7^- (I_5^- and I_2 interacting) calculated with EDA and IQA. For a detailed explanation of the terms refer to the recent article by Racioppi *et al.* [42]. The total bonding energy is decreasing from I_3^- to I_5^- to I_7^-. The net orbitalic energy is positive in all cases and in that sense the formation of I_3^- (as already discussed by Hoffmann [201]) as well as I_5^- and I_7^- is driven by electrostatics. With IQA a positive classic contribution to the interaction energy is obtained (Table A.2) which is however relativized by partitioning also the deformation energy (Table A.3) where a negative and dominantly classic contribution is calculated [42].

Abbreviations used in Tables A.1-A.3:
EDA
E_{elst}: Electrostatic term
E_{Pauli}: Pauli term
$E_{orbitalic}$: Orbitalic term
$E_{net\ orbitalic}$: Is the sum of the Pauli and the orbitalic term
For explication of these terms see [39]
E_{kin}: kinetic energy
$E_{Coulomb}$: Coulombic parts of E_{Pauli} and $E_{orbitalic}$
$E_{classic}$: The sum of E_{elst} and the coulombic parts of E_{Pauli} and $E_{orbitalic}$
E_{xc}: The exchange-correlation parts of E_{Pauli} and $E_{orbitalic}$
These terms are obtained by a partitioning of the Hamilton operator in the Hartree-Fock case

and the functional in the DFT case

IQA

$V_{int,cl}$: classic part of the interaction energy

$V_{int,xc}$: exchange-correlation part of the interaction energy

V_{int}: interaction energy

$V_{deformation}$: deformation energy

V_{bind}: binding energy

For further explication of these terms refer to [38]

$E_{bind,cl}$: $V_{int,cl}$ plus the classic part of the deformation energy

$E_{bind,xc}$: $V_{int,xc}$ plus the exchange-correlation part of the deformation energy

For further explication of these terms refer to [42].

Table A.1 - Total bonding energies

	ab-initio		
	I_3^- (I_2+I^-)	**I_5^- (I_3^-+I_2)**	**I_7^- (I_5^-+I_2)**
E_{bind}	***-171.4***	***-80.5***	***-52.6***

Table A.2 - Bonding energies for I_3^-, I_5^-, and I_7^- using a conventional decomposition scheme

	EDA				**IQA**		
	I_3^- (I_2+I^-)	**I_5^- (I_3^-+I_2)**	**I_7^- (I_5^-+I_2)**		**I_3^- (I_2+I^-)**	**I_5^- (I_3^-+I_2)**	**I_7^- (I_5^-+I_2)**
E_{elst}	-203.7	-112.3	-83.5	$V_{int,cl}$	57.45	43.06	43.00
E_{Pauli}	273.8	160.9	122.1	$V_{int,xc}$	-348.0	-269.0	-235.2
$E_{orbitalic}$	-237.7	-131.9	-95.0	V_{int}	-290.5	-225.9	-192.2
$E_{net\ Orbitalic}$	36.13	28.96	27.19	$V_{deformation}$	128.67	147.0	142.1
E_{bind}	***-167.5***	***-83.3***	***-56.3***	V_{bind}	***-161.85***	***-78.9***	***-50.1***

Table A.3 - Bonding energies for I_3^-, I_5^-, and I_7^- using Racioppi's [42] decomposition scheme for IQA

	EDA				***IQA***		
	I_3^- (I_2+I^-)	**I_5^- (I_3^-+I_2)**	**I_7^- (I_5^-+I_2)**		**I_3^- (I_2+I^-)**	**I_5^- (I_3^-+I_2)**	**I_7^- (I_5^-+I_2)**
E_{kin}	715.83	440.17	345.61	$\Delta E_{self,T}$	619.75	386.65	294.72
E_{elst}	-203.65	-112.30	-83.53				
$E_{Coulomb\ (Steric\ +OrbInt)\ Energy}$	-515.29	-323.06	-255.89				
$E_{classic}$	-718.94	-435.36	-339.42	$E_{bind,cl}$	-581.81	-330.73	-237.99
E_{xc}	-164.41	-88.14	-62.53	$E_{bind,xc}$	-199.79	-134.81	-106.99
% classic	81.39	83.16	84.44	% classic	80.95	80.41	79.07
% XC	18.61	16.84	15.56	% XC	25.56	28.96	31.01

A.6 Further IQA interaction terms for I_5^-

Figure A.5 shows additional aspects of the bonding energy partitioning for I_5^- obtained with IQA calculations.

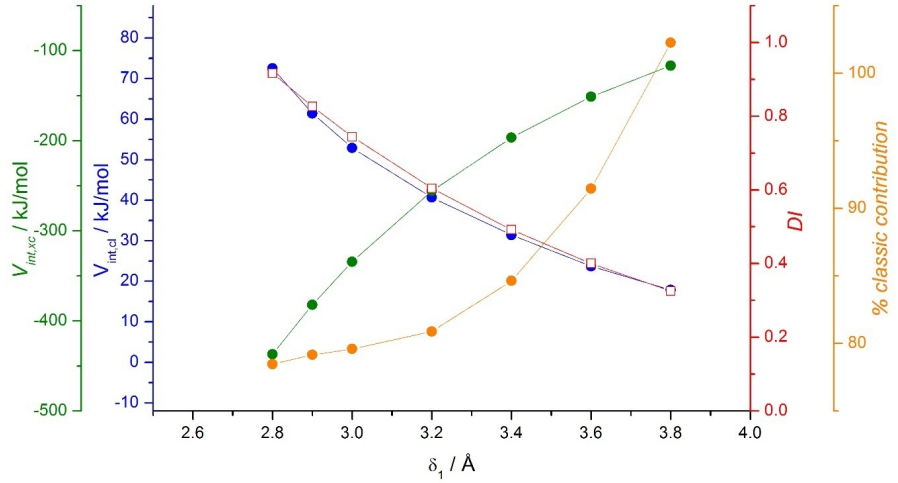


Figure A.5 – Exchange-correlation and classic part of the interaction energy $V_{int,xc}$ and $V_{int,cl}$, delocalization index DI, and percentage classic contribution for I_5^- with constrained distance I_3^--I_2.

$V_{int,cl}$ and $V_{int,xc}$ correspond to the classic and the exchange-correlation part of the interaction energy V_{int}. With increasing distance, both the classic and the exchange correlation term are decreasing in absolute value. It is important to keep in mind that these two terms sum up to V_{int} and not to the total binding energy E_{bind}. The percentage classic contribution shown in this figure as well as the The classic part of the interaction energy $V_{int,cl}$ and the delocalization index (DI) show the same decrease. Their correlation was described in a paper by Francisco *et al.* [44].

A.7 IQA and EDA terms for tri-iodide I_3^-

The behaviour of the relative non-classic (exchange-correlation) contribution to the binding energy is basically the same as observed for I_5^-.

A.8 Temperature effect on the lattice parameters and interatomic distances in TEAI under ambient pressure

In order to check the effect of low temperature on the unit cell parameters and molecular structure of TEAI, a diffraction experiment at 110 K was performed. Table A.4 shows the comparison of selected structural parameters between structures collected at 110 and 29 8K [217]. One can see that the *c* axis shortening at low temperature is greatly hindered by the iodine motif running along this axis. It is in accordance with the compressibility trend under high-pressure. The same behaviour is presented by the shortening of d_1 – low temperature affects stronger compression of this contact (with respect to d_2), just like in the high-pressure structures.

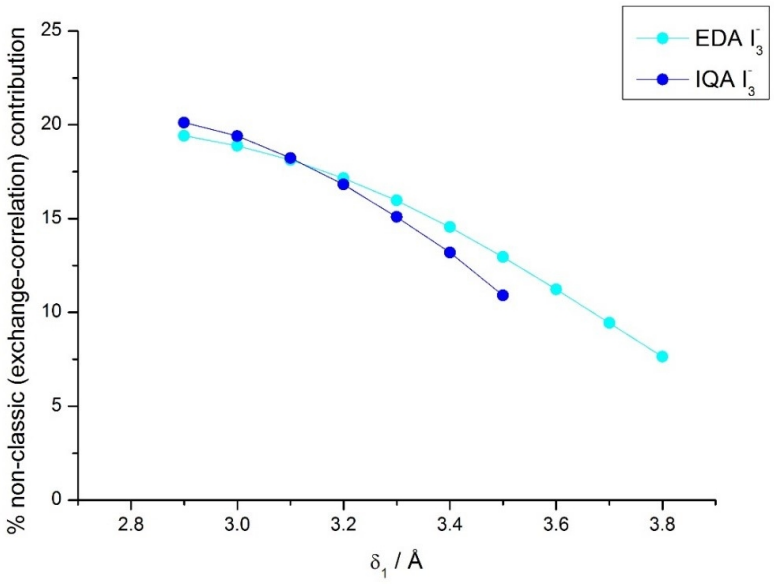


Figure A.6 – Relative non-classic contribution to the binding energy upon elongation of δ_1, which is the distance between I^- and I_2.

Table A.4 – Unit cell axes lengths, volume and values of d_1 and d_2 (see Figure 7.1) in the low- and ambient temperature structure of TEAI

	110 K	298 K
a [Å]	11.2757(1)	11.511(3)
b [Å]	15.3799(2)	15.654(5)
c [Å]	12.3216(1)	12.358(5)
V [Å^3]	2136.80(4)	2226.83(8)
d_1 [Å]	3.4197(4)	3.497(4)
d_2 [Å]	3.4846(4)	3.494(4)

A.9 Electrical resistivity measurements under high pressure and variable temperature

A DAC pressurized to circa 10 GPa was prepared for conductivity measurements as described in the Experimental section. The sample thickness was determined by the external measurement of the distance between the diamonds before and after pressurization (the gasket thickness was measured beforehand) using a micrometer. Subsequently the Pt100 thermocouple was placed inside a 2 mm long on-purpose drilled hole within the DAC body, in order to measure its temperature. The entire system was placed inside a Büchi B-585 glass oven. I-V scans between 0.5 and 3 V were performed at 26.0; 31.5; 35.8, 42.7; 49.7 and 54.9 °C (stable readings from the digital thermometer connected to the thermocouple). Firstly, bulk resistivities at each temperature point were calculated using the Van der Pauw method as: $\rho = \pi R \cdot t / \ln 2$, where R is the measured resistance (Table A.5) and t is the constant sample thickness (42μm).

Table A.5 – Recorded resistances of TEAI sample under 10 GPa at variable temperatures.

T [°C]	R [MΩ]
26.0	7.17(6)
31.5	5.46(20)
35.8	2.44(47)
42.7	1.35(12)
49.7	0.48(5)
54.9	0.46(2)

We then calculated the band gap using Arrhenius equation ($E_g = 2k \ln \rho / \frac{1}{T}$) and linearly fitting the function slope (Figure A.7). The obtained value of the bad gap is to 1.60 ± 0.03 eV.

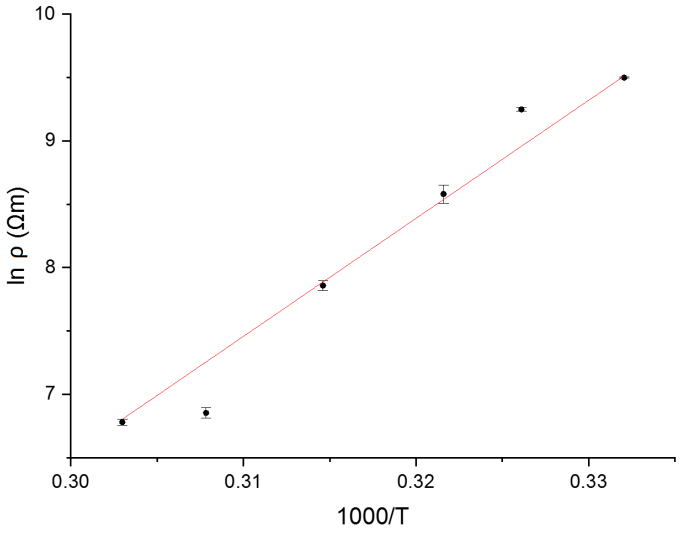


Figure A.7 – Resistivity versus temperature plot for TEAI under circa 10 GPa. The red line indicates the linear trend model fitted to the data.

B Appendix to Chapter 8

B.1 Implementation of new formula in *Tonto*

The implementation of the new formula where χ^2 has units of energy into *Tonto* [126] is not trivial due to the complexity of this software code. After a lot of trial and error, it was found that the formula:

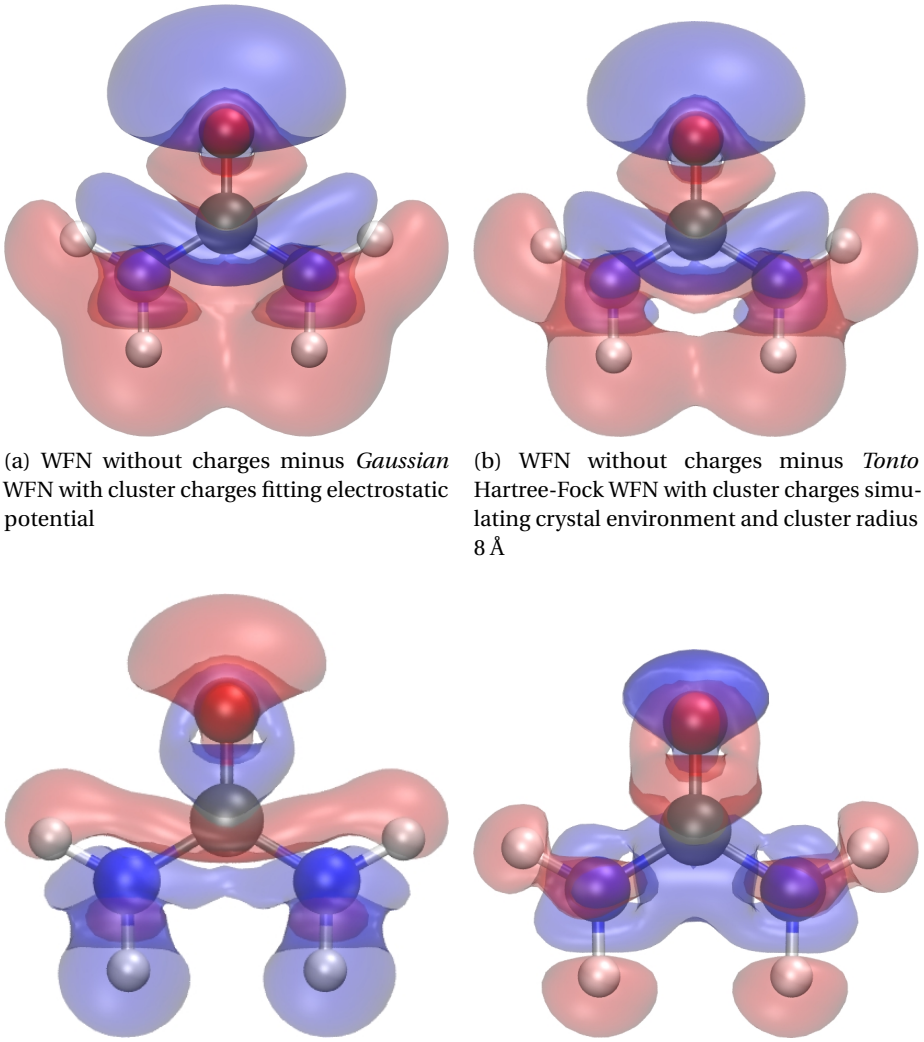
*fac_times_ alpha=fac*alpha*(alpha*Fm-Fexp)/(Fsig*Fsig*Fm)*

(reproduced here in the way it is written in the code) has to be modified in the following way to include **H**:

*fac_times_ alpha=fac*alpha*(alpha*Fm-Fexp)/(Fsig*stl*2*Fsig*stl*2*Fm*2*3.141592654*volume)*

B.2 Point charges from *Gaussian* and *Tonto*

Point charges can simulate the electric field around a molecule. In *Gaussian* charges fitting the electrostatic potential of the first coordination sphere were put. In *Tonto* one can choose to put charges that are within a given radius around the molecule. A cluster radius of 8 Å was chosen.
The effect of point charges manually put in *Gaussian* and the point charges automatically assigned in *Tonto* for a cluster radius of 8 Å is very similar as seen in Figure B.1a and b. Figure B.1c shows that with the way the charges were put in *Gaussian*, the effect is a bit stronger. Figure B.1d shows the difference between the wavefunction calculated without field and an X-ray constrained density fitted to structure factors from a wavefunction with the field. The effect is the same as seen in (a) and (b), just less strong.



(a) WFN without charges minus *Gaussian* WFN with cluster charges fitting electrostatic potential

(b) WFN without charges minus *Tonto* Hartree-Fock WFN with cluster charges simulating crystal environment and cluster radius 8 Å

(c) *Gaussian* WFN with cluster charges fitting electrostatic potential minus *Tonto* Hartree-Fock WFN with cluster charges simulating crystal environment and cluster radius 8 Å

(d) WFN without charges minus *Tonto* XCW with $\lambda = 10$ and resolution = 0.7 $Å^{-1}$

Figure B.1 – Density difference maps with an isovalue of 0.0004 e/Bohr^3. Red regions indicate an accumulation of electron density for the first term and blue for the second (subtracted) term.

C Appendix to Chapter 9

C.1 Dipole moment of HCN

HCN has a relatively strong dipole moment that is influenced by the electric field as shown on the very left of Figure C.1 (direct space). The effect is damped a lot when going via simulated structure factors and XRW fitting (right side of Figure C.1). Again the effect is stronger at lower resolution for the same weighting parameter (λ=5), and the typical sampling enabled by a medium-small unit cell is insufficient for a proper description.

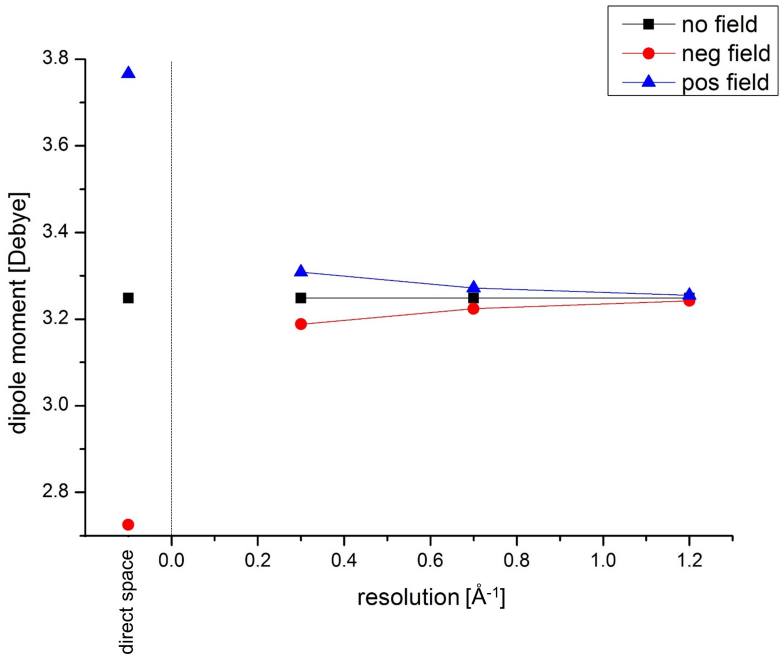


Figure C.1 – Magnitude of dipole moment for the three electric fields obtained from the primary wavefunctions (direct space) and from the X-ray restrained wavefunctions (fitting in reciprocal space).

C.2 Electric field of 0.1 a.u. along HCN

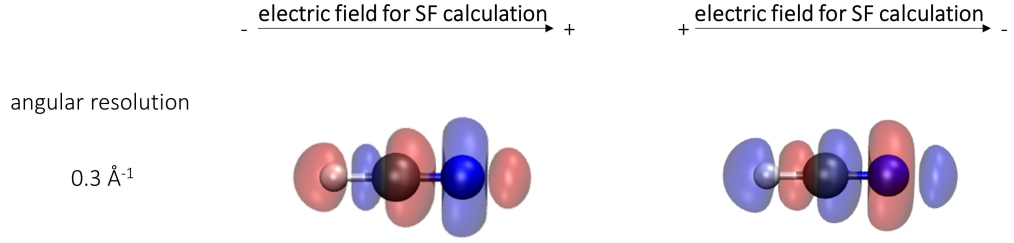


Figure C.2 – Density differences obtained by comparing XR-RHF calculations with SF including and excluding the electric field of 0.1 a.u., which is ten times stronger than the one used for all other results. $\rho^{XR}_{no\ field} - \rho^{XR}_{field}$; blue regions indicate an accumulation of electron density for XR-RHF calculations with SF including the electric field effects. For all the plots, the isovalue is 0.002 e/bohr^3; which is ten times higher than the one used for the density difference plots in Figure 9.3.

C.3 Electron density difference for oxalic acid with and without water

Figure C.3 shows electron density differences obtained by subtracting the density of a direct-space calculation on the oxalic acid with the water molecules (at the position shown, which are taken from the experimental structure) from the density resulting from a direct-space computation on the oxalic acid with the water molecules 5 Å apart. Therefore this figure shows the effect of directly including or excluding the interaction of the oxalic acid with the water molecules in the primary wavefunction calculation. Figure 9.9, instead, shows electron density differences from XR-RHF calculations where the interaction of the oxalic acid molecule with the water molecules is only considered with the structure factors but not directly in the pure wavefunction calculation. Nevertheless, the density difference features in these two figures are very similar. The picture on the left of Figure C.3, for example, shows a red lobe on the carboxylic hydrogen atom and blue lobes around the oxygen atom, which are features also observed in Figure 9.9. The picture on the right indicates the shift of electrons in the carbonyl group form the carbon to the oxygen atom, which is again also observed in Figure 9.9. This highlights the similarity between directly including and excluding water molecules in the calculation of the primary wavefunction and only including their effect via structure factors from the crystal.

C.4 Periodic DFT calculations for NTA-B

Periodic DFT calculations were performed with the software *CRYSTAL17* [23] using the hybrid functional B3LYP. The basis set pob-TZVP [245] from the *Crystal* library was employed for all the atoms. London-type pairwise empirical correction to the energy was used to account for dispersion effects [243]. The experimental XRD structure at ambient pressure was used as

Figure C.3 – Electron density differences obtained by comparing direct-space calculations on oxalic acid dihydrate: $\rho_{water\ away} - \rho_{water\ close}$. The plotted isosurface corresponds to 0.004 e/bohr^3. Blue regions indicate an accumulation of electron density when the water molecules are close (as they are shown in this Figure).

starting guess for geometry optimization at ambient pressure. Calculations were performed at 0, 1, 2, 3, 4, 5, 10, 15, 20, 30, and 40 GPa in $Pna2_1$ space-group. After 40 GPa, convergence could not be obtained with the pob-TZVP basis set. Two additional pressure points at 50 and 60 GPa could be obtained by changing the scaling factor of the most diffuse functions of the basis set from 1.00 to 1.05. This is not the cause for the increase of several intermolecular distances above 40 GPa in Figure 9.12a, since calculations below 40 GPa with the scaled basis set led to very similar results as with the scaled basis set. In addition, calculations in the space group P1 were performed for different pressure points to check for the possible existence of alternative phases, but the geometry remained the same as in the $Pna2_1$ optimization.

C.5 Electron density differences of NTA-B for structure factors simulated at different pressures

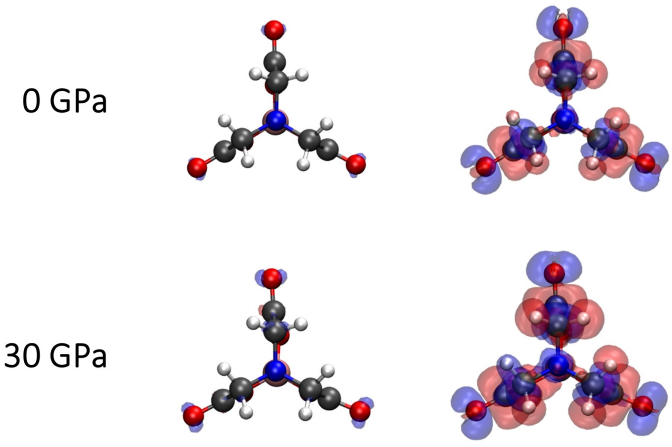


Figure C.4 – Density differences for the 0 GPa and 30 GPa structures resulting from the comparison of the XRW calculation using IMM structure factors with the XRW calculation using crystal structure factors. The iso-contour level on the left is 0.01 e/Bohr^3 and on the right 0.001 e/Bohr^3, and blue signifies negative and red positive iso-contour values.

C.6 Change of carbonyl ELMOs with point charge

Two point charges of +1 were positioned around the carbonyl group of an acetone molecule as shown in Figure C.5 to mimic the electrostatic part of a carbonyl-carbonyl interaction in the crystal and observe its influence on the orbitals.
Structure factors were simulated for the wavefunctions obtained with and without point charges. Both structure factor sets were used for XR-ELMO calculations. The resulting differences for the s, p, σ, and π orbital are shown in Figure C.6 from three different perspectives.
Comparison between these simplified situation with the ELMO differences of the carbonyl groups of NTA-B in the crystal shows many similarities. Therefore, the behaviour of the ELMOs shown in Figure C.6 can be taken as exemplary for how the molecular orbitals of a carbonyl functional group change when influenced by charges.

Figure C.5 – The positions of the two point charges of +1 are shown in light blue from two different perspectives.

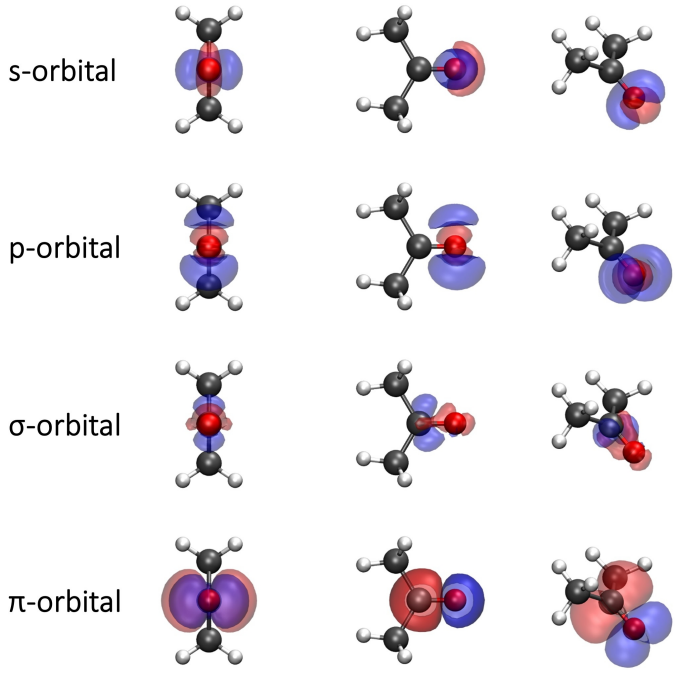


Figure C.6 – Differences for the XR-ELMOs associated with the carbonyl group shown from three different perspectives ($|\varphi^{XR}_{no\ charges}|^2 - |\varphi^{XR}_{charges}|^2$; iso-contour level set to $4*10^{-5}$ e/bohr^3, blue indicates negative and red positive regions).

Bibliography

[1] N. S. Ostlund and A. Szabo, *Modern Quantum Chemistry: Introduction to advanced electronic structure theory*. Dover Publications Inc: New York, 1996.

[2] S. Leutwyler, "Lecture notes from the course ab-initio computational chemistry." Bern, 2014.

[3] R. McWeeny, "Some Recent Advances in Density Matrix Theory," *Reviews of Modern Physics*, vol. 32, no. 2, pp. 335–369, 1960.

[4] P.-O. Löwdin, "Quantum Theory of Many-Particle Systems. I. Physical Interpretations by Means of Density Matrices, Natural Spin-Orbitals, and Convergence Problems in the Method of Configurational Interaction," *Physical Review*, vol. 97, no. 6, pp. 1474–1489, 1955.

[5] A. Genoni, "Searching for "Experimental" Wave Functions: Past, Present and Future," Lecture at the 1st Robert F. Stewart School on Electron Density and Related Properties, (Nancy, France), 2016.

[6] J.-M. Gillet and T. Koritsanszky, "Past, present and future of charge density and density matrix refinements," in *Modern Charge-Density Analysis* (C. Gatti and P. Macchi, eds.), pp. 181–211, Springer Netherlands, 2012.

[7] V. Tsirelson, "Early days of quantum crystallography: A personal account," *Journal of Computational Chemistry*, vol. 39, no. 17, pp. 1029–1037, 2018.

[8] A. J. Coleman, "Structure of Fermion Density Matrices," *Reviews of Modern Physics*, vol. 35, no. 3, pp. 668–686, 1963.

[9] J. Coleman, "N-representability," in *Reduced-Density-Matrix Mechanics: With Application to Many-Electron Atoms and Molecules* (D. A. Mazziotti, ed.), pp. 3–9, John Wiley & Sons, Inc., 2007.

[10] L. H. Thomas, "The calculation of atomic fields," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 23, no. 5, pp. 542–548, 1927.

[11] E. Fermi, "Eine statistische Methode zur Bestimmung einiger Eigenschaften des Atoms und ihre Anwendung auf die Theorie des periodischen Systems der Elemente," *Zeitschrift für Physik*, vol. 48, no. 1-2, pp. 73–79, 1928.

[12] P. Hohenberg and W. Kohn, "Inhomogeneous electron gas," *Physical review*, vol. 136, no. 3B, p. B864, 1964.

[13] W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects," *Physical review*, vol. 140, no. 4A, p. A1133, 1965.

[14] F. M. Bickelhaupt and E. J. Baerends, "Kohn-Sham Density Functional Theory: Predicting and Understanding Chemistry," in *Reviews in Computational Chemistry*, vol. 15, pp. 1–86, 2000.

[15] J. P. Perdew, "Density-functional approximation for the correlation energy of the inhomogeneous electron gas," *Phys. Rev. B*, vol. 33, pp. 8822–8824, 1986.

[16] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized Gradient Approximation Made Simple," *Phys. Rev. Lett.*, vol. 77, pp. 3865–3868, 1996.

[17] R. Dovesi, B. Civalleri, R. Orlando, C. Roetti, and V. R. Saunders, "Ab Initio Quantum Simulation in Solid State Chemistry," *ChemInform*, vol. 36, no. 48, 2005.

[18] F. Bloch, "Über die quantenmechanik der elektronen in kristallgittern," *Zeitschrift für physik*, vol. 52, no. 7-8, pp. 555–600, 1929.

[19] P. Giannozzi, O. Andreussi, T. Brumme, O. Bunau, M. B. Nardelli, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, M. Cococcioni, *et al.*, "Advanced capabilities for materials modelling with Quantum ESPRESSO," *Journal of Physics: Condensed Matter*, vol. 29, no. 46, p. 465901, 2017.

[20] J. Hafner and G. Kresse, "The Vienna ab-Initio Simulation Program VASP: An Efficient and Versatile Tool for Studying the Structural, Dynamic, and Electronic Properties of Materials," in *Properties of Complex Inorganic Solids*, pp. 69–82, Springer, 1997.

[21] P. Blaha, K. Schwarz, G. K. Madsen, D. Kvasnicka, and J. Luitz, "wien2k - An augmented plane wave+ local orbitals program for calculating crystal properties," 2001.

[22] F. Furche, R. Ahlrichs, C. Hättig, W. Klopper, M. Sierka, and F. Weigend, "Turbomole," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 4, no. 2, pp. 91–100, 2014.

[23] R. Dovesi, A. Erba, R. Orlando, C. M. Zicovich-Wilson, B. Civalleri, L. Maschio, M. Rérat, S. Casassa, J. Baima, S. Salustro, *et al.*, "Quantum-mechanical condensed matter simulations with CRYSTAL," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 8, no. 4, p. e1360, 2018.

[24] F. L. Hirshfeld, "Bonded-atom fragments for describing molecular charge densities," *Theoretica chimica acta*, vol. 44, no. 2, pp. 129–138, 1977.

[25] R. F. W. Bader, "Atoms in molecules," *Accounts of Chemical Research*, vol. 18, no. 1, pp. 9–15, 1985.

[26] R. F. W. Bader, "Properties of atoms and bonds in carbocations," *Canadian Journal of Chemistry*, vol. 64, no. 6, pp. 1036–1045, 1986.

[27] C. F. Matta and R. J. Boyd, "An introduction to the quantum theory of atoms in molecules," *The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design*, pp. 1–30, 2007.

[28] B. B. Iversen, F. K. Larsen, M. Souhassou, and M. Takata, "Experimental evidence for the existence of non-nuclear maxima in the electron-density distribution of metallic beryllium. A comparative study of the maximum entropy method and the multipole refinement method," *Acta Crystallographica Section B*, vol. 51, no. 4, pp. 580–591, 1995.

[29] D. Cremer and E. Kraka, "A description of the chemical bond in terms of local properties of electron density and energy," *Croatica Chemica Acta*, vol. 57, no. 6, pp. 1259–1281, 1984.

[30] D. Cremer and E. Kraka, "Chemical bonds without bonding electron density—does the difference electron-density analysis suffice for a description of the chemical bond?," *Angewandte Chemie International Edition in English*, vol. 23, no. 8, pp. 627–628, 1984.

[31] X. Fradera, M. A. Austen, and R. F. W. Bader, "The Lewis Model and Beyond," *The Journal of Physical Chemistry A*, vol. 103, no. 2, pp. 304–314, 1999.

[32] T. A. Keith, "AIMAll (Version 14.23.11)." TK Gristmill Software, Overland Park KS, USA (aim.tkgristmill.com).

[33] E. R. Johnson, S. Keinan, P. Mori-Sanchez, J. Contreras-García, A. J. Cohen, and W. Yang, "Revealing noncovalent interactions," *Journal of the American Chemical Society*, vol. 132, no. 18, pp. 6498–6506, 2010.

[34] A. D. Becke and K. E. Edgecombe, "A simple measure of electron localization in atomic and molecular systems," *The Journal of chemical physics*, vol. 92, no. 9, pp. 5397–5403, 1990.

[35] M. D. Gould, C. Taylor, S. K. Wolff, G. S. Chandler, and D. Jayatilaka, "A definition for the covalent and ionic bond index in a molecule," *Theoretical Chemistry Accounts*, vol. 119, no. 1-3, pp. 275–290, 2008.

[36] E. D. Glendening, C. R. Landis, and F. Weinhold, "Natural bond orbital methods," *Wiley interdisciplinary reviews: computational molecular science*, vol. 2, no. 1, pp. 1–42, 2012.

[37] M. Sironi, A. Genoni, M. Civera, S. Pieraccini, and M. Ghitti, "Extremely localized molecular orbitals: theory and applications," *Theoretical Chemistry Accounts*, vol. 117, no. 5-6, pp. 685–698, 2007.

[38] M. A. Blanco, Á. M. Pendás, and E. Francisco, "Interacting quantum atoms: A correlated energy decomposition scheme based on the quantum theory of atoms in molecules," *Journal of Chemical Theory and Computation*, vol. 1, no. 6, pp. 1096–1109, 2005.

[39] M. von Hopffgarten and G. Frenking, "Energy decomposition analysis," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 2, no. 1, pp. 43–62, 2012.

[40] E. J. Baerends and T. Ziegler, "ADF." SCM, Theoretical Chemistry, Vrije Universiteit, Amsterdam, TheNetherlands, 2017.

[41] Á. M. Pendás, M. A. Blanco, and E. Francisco, "The nature of the hydrogen bond: A synthesis from the interacting quantum atoms picture," *Journal of Chemical Physics*, vol. 125, no. 18, 2006.

[42] S. Racioppi, R. Della Pergola, V. Colombo, A. Sironi, and P. Macchi, "Electron Density Analysis of Metal Clusters with Semi-Interstitial Main Group Atoms. Chemical Bonding in [Co6X(CO)16]- Species," *The Journal of Physical Chemistry A*, vol. 122, no. 22, pp. 5004–5015, 2018.

[43] P. Maxwell, Á. M. Pendás, and P. L. A. Popelier, "Extension of the interacting quantum atoms (IQA) approach to B3LYP level density functional theory (DFT)," *Phys. Chem. Chem. Phys.*, vol. 18, no. 31, pp. 20986–21000, 2016.

[44] E. Francisco, M. A. Blanco, and Á. M. Pendás, "Chemical Fragments in Real Space: Definitions, Properties, and Energetic Decoompositions," *Journal of computational chemistry*, vol. 28, no. 1, p. 161, 2006.

[45] I. Cukrowski, "IQA-embedded fragment attributed molecular system energy change in exploring intramolecular interactions," *Computational and Theoretical Chemistry*, vol. 1066, pp. 62–75, 2015.

[46] I. Cukrowski, F. Sagan, and M. P. Mitoraj, "On the Stability of Cis- and Trans-2-Butene Isomers. An Insight Based on the FAMSEC, IQA, and ETS-NOCV Schemes," *Journal of Computational Chemistry*, vol. 37, no. 32, pp. 2783–2798, 2016.

[47] K. Kitaura and K. Morokuma, "A new energy decomposition scheme for molecular interactions within the Hartree-Fock approximation," *International Journal of Quantum Chemistry*, vol. 10, no. 2, pp. 325–340, 1976.

[48] M. F. Guest, I. J. Bush, H. J. Van Dam, P. Sherwood, J. M. Thomas, J. H. Van Lenthe, R. W. Havenith, and J. Kendrick, "The GAMESS-UK electronic structure package: algorithms, developments and applications," *Molecular physics*, vol. 103, no. 6-8, pp. 719–747, 2005.

[49] D. Tiana, E. Francisco, M. A. Blanco, P. Macchi, A. Sironi, and Á. M. Pendás, "Bonding in Classical and Nonclassical Transition Metal Carbonyls: The Interacting Quantum Atoms Perspective," *Journal of Chemical Theory and Computation*, pp. 1064–1074, 2009.

[50] R. P. Feynman, R. B. Leighton, and M. Sands, *11 - Inside Dielectrics*, vol. 2 of *The Feynman Lectures on Physics*. Basic books, 2011.

[51] W. Volksen, R. D. Miller, and G. Dubois, "Low dielectric constant materials," *Chemical reviews*, vol. 110, no. 1, pp. 56–110, 2009.

[52] T. Higashihara and M. Ueda, "Recent Progress in High Refractive Index Polymers," *Macromolecules*, vol. 48, no. 7, pp. 1915–1929, 2015.

[53] M. Ernst, L. H. Dos Santos, and P. Macchi, "Optical properties of metal-organic networks from distributed atomic polarizabilities," *CrystEngComm*, vol. 18, no. 38, pp. 7339–7346, 2016.

[54] M. v. Laue, "In welchem Sinne kann man von einem Mikroskopieren des Feinbaus der Kristalle mittels der Röntgenstrahlen reden?," *Naturwissenschaften*, vol. 8, pp. 968–971, 1920.

[55] W. H. Bragg and W. L. Bragg, "The reflection of X-rays by crystals," *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, vol. 88, no. 605, pp. 428–438, 1913.

[56] W. L. Bragg, "The structure of some crystals as indicated by their diffraction of X-rays," *Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character*, vol. 89, no. 610, pp. 248–277, 1913.

[57] C. Hammond and C. Hammond, *The Basics of Cristallography and Diffraction*, vol. 214. Oxford, 2001.

[58] K. Trueblood, H.-B. Bürgi, H. Burzlaff, J. Dunitz, C. Gramaccioli, H. Schulz, U. Shmueli, and S. Abrahams, "Atomic dispacement parameter nomenclature. report of a subcommittee on atomic displacement parameter nomenclature," *Acta Crystallographica Section A: Foundations of Crystallography*, vol. 52, no. 5, pp. 770–781, 1996.

[59] P. Macchi, "Modern charge density studies : the entanglement of experiment and theory," *Crystallography Reviews*, vol. 19, no. 2, pp. 58–101, 2013.

[60] J. Als-Nielsen and D. McMorrow, *Elements of modern X-ray physics*. John Wiley & Sons, 2011.

[61] D. M. Collins, "Electron density images from imperfect data by iterative entropy maximization," *Nature*, vol. 298, no. 5869, p. 49, 1982.

[62] N. K. Hansen and P. Coppens, "Testing aspherical atom refinements on small-molecule data sets," *Acta Crystallographica Section A*, vol. 34, pp. 909–921, 1978.

[63] D. Jayatilaka, *Using Wavefunctions to Get More Information Out of Diffraction Experiments*. Springer Netherlands, 2012.

[64] A. Genoni, L. H. R. Dos Santos, B. Meyer, and P. Macchi, "Can X-ray constrained Hartree–Fock wavefunctions retrieve electron correlation?," *IUCrJ*, vol. 4, no. 2, pp. 136–146, 2017.

[65] T. Koritsanszky, P. Macchi, C. Gatti, L. Farrugia, P. Mallinson, A. Volkov, and T. Richter, "XD-2016 -A computer program package for multipole refinement and topological analysis of charge densities and evaluation of intermolecular energies from experimental or theoretical structure factors."

[66] W. Friedrich, P. Knipping, and M. Laue, "Interferenzerscheinungen bei Röntgenstrahlen," *Annalen der Physik*, vol. 346, no. 10, pp. 971–988, 1913.

[67] W. L. Bragg, "The Diffraction of Short Electromagnetic Waves by a Crystal.," *Proceedings of the Cambridge Philosophical Society*, vol. 17, pp. 43–57, 1912.

[68] S. W. H. Bragg, "The Intensity of X-Ray Reflection by Diamond," *Proceedings of the Physical Society of London*, vol. 33, no. 1, pp. 304–311, 1920.

[69] W. L. Clinton and L. J. Massa, "Determination of the Electron Density Matrix from X-Ray Diffraction Data," *Phys. Rev. Lett.*, vol. 29, no. 20, pp. 1363–1366, 1972.

[70] A. Genoni, L. Bučinský, N. Claiser, J. Contreras-García, B. Dittrich, P. M. Dominiak, E. Espinosa, C. Gatti, P. Giannozzi, J.-M. Gillet, D. Jayatilaka, P. Macchi, A. Ø. Madsen, L. Massa, C. F. Matta, K. M. Merz, P. N. H. Nakashima, H. Ott, U. Ryde, K. Schwarz, M. Sierka, and S. Grabowsky, "Quantum Crystallography: Current Developments and Future Perspectives," *Chemistry - A European Journal*, vol. 24, no. 43, pp. 10881–10905, 2018.

[71] S. Grabowsky, A. Genoni, and H.-B. Bürgi, "Quantum crystallography," *Chemical Science*, vol. 8, no. 6, pp. 4159–4176, 2017.

[72] L. Massa and C. F. Matta, "Exploiting the full quantum crystallography," *Canadian Journal of Chemistry*, vol. 96, no. 7, pp. 599–605, 2018.

[73] R. James and G. Brindley, "VI. Some numerical calculations of atomic scattering factors," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 12, no. 75, pp. 81–112, 1931.

[74] McWeeny, "X-ray scattering by aggregates of bonded atoms. I. Analytical approximations in single-atom scattering," *Acta Crystallographica*, vol. 4, no. 6, pp. 513–519, 1951.

[75] R. McWeeny, "X-ray scattering by aggregates of bonded atoms. II. The effect of the bonds: with an application to H2," *Acta Crystallographica*, vol. 5, no. 4, pp. 463–468, 1952.

[76] R. McWeeny, "X-ray scattering by aggregates of bonded atoms. III. The bond scattering factor: simple methods of approximation in the general case," *Acta Crystallographica*, vol. 6, no. 7, pp. 631–637, 1953.

[77] R. McWeeny, "X-ray scattering by aggregates of bonded atoms. IV. Applications to the carbon atom," *Acta Crystallographica*, vol. 7, no. 2, pp. 180–186, 1954.

[78] L. Kleinman and J. C. Phillips, "Covalent Bonding and Charge Density in Diamond," *Physical Review*, vol. 125, no. 3, pp. 819–824, 1962.

[79] R. J. Weiss, "Charge density in diamond," *Phys. Lett.*, vol. 12, no. 4, pp. 293–294, 1964.

[80] B. Dawson, "The covalent bond in diamond," *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 298, no. 1454, pp. 264–288, 1967.

[81] J. J. DeMarco and R. J. Weiss, "Absolute X-Ray Scattering Factors of Silicon and Germanium," *Physical Review*, vol. 137, no. 6A, pp. A1869–A1871, 1965.

[82] R. F. Stewart, E. R. Davidson, and W. T. Simpson, "Coherent X-ray scattering for the hydrogen atom in the hydrogen molecule," *The Journal of Chemical Physics*, vol. 42, no. 9, pp. 3175–3187, 1965.

[83] R. F. Stewart, "Generalized X-Ray Scattering Factors," *The Journal of Chemical Physics*, vol. 51, no. 10, pp. 4569–4577, 1969.

[84] R. F. Stewart, "Electron population analysis with generalized x-ray scattering factors: Higher multipoles," *The Journal of Chemical Physics*, vol. 58, no. 4, pp. 1668–1676, 1973.

[85] R. F. Stewart, "Valence Structure from X-Ray Diffraction Data: Physical Properties," *The Journal of Chemical Physics*, vol. 57, no. 4, pp. 1664–1668, 1972.

[86] P. Coppens, L. Csonka, and T. V. Willoughby, "Electron Population Parameters from Least-Squares Refinement of X-ray Diffraction Data," *Science*, vol. 167, no. 3921, pp. 1126–1128, 1970.

[87] P. Coppens, T. V. Willoughby, and L. N. Csonka, "Electron population analysis of accurate diffraction data. I. Formalisms and restrictions," *Acta Crystallographica Section A*, vol. 27, no. 3, pp. 248–256, 1971.

[88] P. Coppens, D. Pautler, and J. F. Griffin, "Electron population analysis of accurate diffraction data. II. Application of one-center formalisms to some organic and inorganic molecules," *Journal of the American Chemical Society*, vol. 93, no. 5, pp. 1051–1058, 1971.

[89] K. Kurki-Suonio, "On the information about deformations of the atoms in X-ray diffraction data," *Acta Crystallographica Section A*, vol. 24, no. 3, pp. 379–390, 1968.

[90] F. L. Hirshfeld, "Difference densities by least-squares refinement: fumaramic acid," *Acta Crystallographica Section B Structural Crystallography and Crystal Chemistry*, vol. 27, no. 4, pp. 769–781, 1971.

[91] R. F. Stewart, J. Bentley, and B. Goodman, "Generalized x-ray scattering factors in diatomic molecules," *The Journal of Chemical Physics*, vol. 63, no. 9, p. 3786, 1975.

[92] R. F. Stewart, "Electron population analysis with rigid pseudoatoms," *Acta Crystallographica Section A*, vol. 32, no. 4, pp. 565–574, 1976.

[93] A. Mukherji and M. Karplus, "Constrained Molecular Wavefunctions: HF Molecule," *The Journal of Chemical Physics*, vol. 38, no. 1, pp. 44–48, 1963.

[94] Y. Rasiel and D. R. Whitman, "Constrained Variation Method in Molecular Quantum Mechanics. Application to Lithium Hydride," *The Journal of Chemical Physics*, vol. 42, no. 6, pp. 2124–2131, 1965.

[95] W. B. Brown, "Perturbation theory of the constrained variational method in molecular quantum mechanics," *The Journal of Chemical Physics*, vol. 44, no. 2, pp. 567–576, 1966.

[96] W. L. Clinton, J. Nakhlehf, and F. Wunderlich, "Direct Determination of Pure-State Density Matrices. I. Some Simple Introductory Calculations," *The Physical Review*, vol. 177, no. 1, pp. 1–6, 1969.

[97] W. L. Clinton, A. J. Galli, and L. J. Massa, "Direct Determination of Pure-State Density Matrices. II. Construction of Constrained Idempotent One-Body Densities," *Physical Review*, vol. 177, no. 1, pp. 7–13, 1969.

[98] W. L. Clinton, G. A. Henderson, and J. V. Prestia, "Direct Determination of Pure-State Density Matrices. III. Purely Theoretical Densities Via an Electrostatic-Virial Theorem," *Physical Review*, vol. 177, no. 1, pp. 13–18, 1969.

[99] W. L. Clinton and G. B. Lamers, "Direct Determination of Pure-State Density Matrices. IV. Investigation of Another Constraint and Another Application of the P Equations," *Physical Review*, vol. 177, no. 1, pp. 19–27, 1969.

[100] W. L. Clinton, A. J. Galli, G. A. Henderson, G. B. Lamers, L. J. Massa, and J. Zarur, "Direct determination of pure-state density matrices. V. Constrained eigenvalue problems," *Physical Review*, vol. 177, no. 1, pp. 27–33, 1969.

[101] W. L. Clinton and L. J. Massa, "The cusp condition: Constraint on the electron density matrix," *International Journal of Quantum Chemistry*, vol. 6, no. 3, pp. 519–523, 1972.

[102] G. A. Henderson and R. K. Zimmerman, "One-electron properties as variational parameters," *The Journal of Chemical Physics*, vol. 65, no. 2, pp. 619–622, 1976.

[103] C. A. Frishberg and L. J. Massa, "Notes on density matrix model for coherent x-ray diffraction," *International Journal of Quantum Chemistry*, vol. 13, no. 6, pp. 801–810, 1978.

[104] L. Massa, L. Huang, and J. Karle, "Quantum crystallography and the use of kernel projector matrices," *International Journal of Quantum Chemistry*, vol. 56, no. S29, pp. 371–384, 1995.

[105] L. Huang, L. Massa, and J. Karle, "Quantum crystallography applied to crystalline maleic anhydride," *International journal of quantum chemistry*, vol. 73, no. 5, pp. 439–450, 1999.

[106] D. Jayatilaka, "Wave Function for Beryllium from X-Ray Diffraction Data," *Physical Review Letters*, vol. 80, no. 4, pp. 798–801, 1998.

[107] R. Stewart, M. Spackman, and C. Flensburg, "VALRAY User's Manual." Chemistry Department, Carnegie-Mellon University, USA, 1983.

[108] M. Deutsch, N. Claiser, S. Pillet, Y. Chumakov, P. Becker, J.-M. Gillet, B. Gillon, C. Lecomte, and M. Souhassou, "Experimental determination of spin-dependent electron density by joint refinement of X-ray and polarized neutron diffraction data," *Acta Crystallographica Section A Foundations of Crystallography*, vol. 68, no. 6, pp. 675–686, 2012.

[109] C. Jelsch, B. Guillot, A. Lagoutte, and C. Lecomte, "Advances in protein and small-molecule charge-density refinement methods using MoPro," *Journal of Applied Crystallography*, vol. 38, no. 1, pp. 38–54, 2005.

[110] V. Petříček, M. Dušek, and L. Palatinus, "Crystallographic computing system JANA2006: general features," *Zeitschrift für Kristallographie-Crystalline Materials*, vol. 229, no. 5, pp. 345–352, 2014.

[111] A. Fischer, D. Tiana, W. Scherer, K. Batke, G. Eickerling, H. Svendsen, N. Bindzus, and B. B. Iversen, "Experimental and theoretical charge density studies at subatomic resolution," *The Journal of Physical Chemistry A*, vol. 115, no. 45, pp. 13061–13071, 2011.

[112] E. Prince and C. Spiegelman, "Statistical significance tests," *International Tables for Crystallography Vol.C, Chapter 8.4*, pp. 702–706, 2006.

[113] A. Ø. Madsen, H. O. Sørensen, C. Flensburg, R. F. Stewart, and S. Larsen, "Modeling of the nuclear parameters for H atoms in X-ray charge-density studies," *Acta Crystallographica Section A: Foundations of Crystallography*, vol. 60, no. 6, pp. 550–561, 2004.

[114] A. Ø. Madsen, "SHADE web server for estimation of hydrogen anisotropic displacement parameters," *Journal of applied crystallography*, vol. 39, no. 5, pp. 757–758, 2006.

[115] V. Schomaker and K. Trueblood, "On the rigid-body motion of molecules in crystals," *Acta Crystallographica Section B: Structural Crystallography and Crystal Chemistry*, vol. 24, no. 1, pp. 63–76, 1968.

[116] K. N. Jarzembska, K. Ślepokura, R. Kamiński, M. J. Gutmann, P. M. Dominiak, and K. Woźniak, "Multi-temperature study of potassium uridine-5′-monophosphate: electron density distribution and anharmonic motion modelling," *Acta Crystallographica Section B*, vol. 73, no. 4, pp. 550–564, 2017.

[117] N. Casati, A. Kleppe, A. P. Jephcoat, and P. Macchi, "Putting pressure on aromaticity along with in situ experimental electron density of a molecular crystal," *Nature Communications*, vol. 7, no. 1, p. 10901, 2016.

[118] V. Pichon-Pesme, C. Lecomte, and H. Lachekar, "On Building a Data Bank of Transferable Experimental Electron Density Parameters Applicable to Polypeptides," *The Journal of Physical Chemistry*, vol. 99, no. 16, pp. 6242–6250, 1995.

[119] B. Zarychta, V. Pichon-Pesme, B. Guillot, C. Lecomte, and C. Jelsch, "On the application of an experimental multipolar pseudo-atom library for accurate refinement of small-molecule and protein crystal structures," *Acta Crystallographica Section A: Foundations of Crystallography*, vol. 63, no. 2, pp. 108–125, 2007.

[120] S. Domagała, B. Fournier, D. Liebschner, B. Guillot, and C. Jelsch, "An improved experimental databank of transferable multipolar atom models–ELMAM2. Construction details and applications," *Acta Crystallographica Section A: Foundations of Crystallography*, vol. 68, no. 3, pp. 337–351, 2012.

[121] S. Domagała and C. Jelsch, "Optimal local axes and symmetry assignment for charge-density refinement," *Journal of Applied Crystallography*, vol. 41, no. 6, pp. 1140–1149, 2008.

[122] P. M. Dominiak, A. Volkov, X. Li, M. Messerschmidt, and P. Coppens, "A theoretical databank of transferable aspherical atoms and its application to electrostatic interaction energy calculations of macromolecules," *Journal of chemical theory and computation*, vol. 3, no. 1, pp. 232–247, 2007.

[123] K. N. Jarzembska and P. M. Dominiak, "New version of the theoretical databank of transferable aspherical pseudoatoms, UBDB2011–towards nucleic acid modelling," *Acta Crystallographica Section A: Foundations of Crystallography*, vol. 68, no. 1, pp. 139–147, 2012.

[124] B. Dittrich, C. B. Hübschle, K. Pröpper, F. Dietrich, T. Stolper, and J. J. Holstein, "The generalized invariom database (GID)," *Acta Crystallographica Section B Structural Science, Crystal Engineering and Materials*, vol. 69, no. 2, pp. 91–104, 2013.

[125] J. M. Bąk, S. Domagała, C. Hübschle, C. Jelsch, B. Dittrich, and P. M. Dominiak, "Verification of structural and electrostatic properties obtained by the use of different pseudoatom databases," *Acta Crystallographica Section A: Foundations of Crystallography*, vol. 67, no. 2, pp. 141–153, 2011.

[126] D. Jayatilaka and D. J. Grimwood, "Tonto: A Fortran Based Object-Oriented System for Quantum Chemistry and Crystallography," in *Computational Science - Iccs 2003, Pt Iv, Proceedings*, vol. 2660, pp. 142–151, 2003.

[127] M. Hudák, D. Jayatilaka, L. Perašínová, S. Biskupič, J. Koíšek, and L. Bučinsk, "X-ray constrained unrestricted Hartree-Fock and Douglas-Kroll-Hess wavefunctions," *Acta Crystallographica Section A: Foundations of Crystallography*, vol. 66, no. 1, pp. 78–92, 2010.

[128] D. J. Grimwood and D. Jayatilaka, "Wavefunctions derived from experiment. II. A wavefunction for oxalic acid dihydrate," *Acta Crystallographica Section A Foundations of Crystallography*, vol. 57, no. 1, pp. 87–100, 2001.

[129] A. E. Whitten, D. Jayatilaka, and M. A. Spackman, "Effective molecular polarizabilities and crystal refractive indices estimated from x-ray diffraction data," *The Journal of Chemical Physics*, vol. 125, no. 17, p. 174505, 2006.

[130] L. H. R. Dos Santos, A. Genoni, and P. Macchi, "Unconstrained and X-ray constrained extremely localized molecular orbitals: analysis of the reconstructed electron density," *Acta Crystallographica Section A Foundations and Advances*, vol. 70, no. 6, pp. 532–551, 2014.

[131] S. C. Capelli, H.-B. Bürgi, B. Dittrich, S. Grabowsky, and D. Jayatilaka, "Hirshfeld atom refinement," *IUCrJ*, vol. 1, no. 5, pp. 361–379, 2014.

[132] F. H. Allen and I. J. Bruno, "Bond lengths in organic and metal-organic compounds revisited: X—H bond lengths from neutron diffraction data," *Acta Crystallographica Section B: Structural Science*, vol. 66, no. 3, pp. 380–386, 2010.

[133] F. Hirshfeld, "XVII. Spatial Partitioning of Charge Density," *Israel Journal of Chemistry*, vol. 16, no. 2-3, pp. 198–201, 1977.

[134] D. Jayatilaka and B. Dittrich, "X-ray structure refinement using aspherical atomic density functions obtained from quantum-mechanical calculations," *Acta Crystallographica Section A: Foundations of Crystallography*, vol. 64, no. 3, pp. 383–393, 2008.

[135] M. Woińska, D. Jayatilaka, M. A. Spackman, A. J. Edwards, P. M. Dominiak, K. Woźniak, E. Nishibori, K. Sugimoto, and S. Grabowsky, "Hirshfeld atom refinement for modelling strong hydrogen bonds," *Acta Crystallographica Section A Foundations and Advances*, vol. 70, no. 5, pp. 483–498, 2014.

[136] S. Grabowsky, P. Luger, J. Buschmann, T. Schneider, T. Schirmeister, A. N. Sobolev, and D. Jayatilaka, "The Significance of Ionic Bonding in Sulfur Dioxide: Bond Orders from X-ray Diffraction Data," *Angewandte Chemie International Edition*, vol. 51, no. 27, pp. 6776–6779, 2012.

[137] L. Chęcińska, W. Morgenroth, C. Paulmann, D. Jayatilaka, and B. Dittrich, "A comparison of electron density from Hirshfeld-atom refinement, X-ray wavefunction refinement and multipole refinement on three urea derivatives," *CrystEngComm*, vol. 15, no. 11, p. 2084, 2013.

[138] C. Pisani, R. Dovesi, A. Erba, and P. Giannozzi, "Electron densities and related properties from the ab-initio simulation of crystalline solids," in *Modern Charge-Density Analysis*, pp. 79–132, Springer, 2011.

[139] J. A. Duffy and M. J. Cooper, "Spin and the Complementary Worlds of Electron Position and Momentum Densities," in *Modern Charge-Density Analysis*, pp. 165–180, Springer, 2011.

[140] P. Coppens, R. Boehme, P. F. Price, and E. D. Stevens, "Joint X-ray and neutron data refinement of structural and charge density parameters," *Acta Crystallographica Section A*, vol. 37, no. 6, pp. 857–863, 1981.

[141] P. M. B. Piccoli, T. F. Koetzle, and A. J. Schultz, "Singe crystal neutron diffraction for the inorganic chemist - a practical guide," *Comments on Inorganic Chemistry*, vol. 28, no. 1-2, pp. 3–38, 2007.

[142] T. Poręba, M. Ernst, D. Zimmer, P. Macchi, and N. Casati, "Pressure-Induced Polymerization and Electrical Conductivity of a Polyiodide," *Angewandte Chemie International Edition*, vol. 58, no. 20, pp. 6625–6629, 2019.

[143] P. Deplano, J. R. Ferraro, M. L. Mercuri, and E. F. Trogu, "Structural and Raman spectroscopic studies as complementary tools in elucidating the nature of the bonding in polyiodides and in donor-I2 adducts," *Coordination Chemistry Reviews*, vol. 188, no. 1, pp. 71–95, 1999.

[144] R. Scatena, Y. T. Guntern, and P. Macchi, "Electron Density and Dielectric Properties of Highly Porous MOFs: Binding and Mobility of Guest Molecules in Cu 3 (BTC) 2 and Zn 3 (BTC) 2," *Journal of the American Chemical Society*, vol. 141, no. 23, pp. 9382–9390, 2019.

[145] T. L. Allen and H. Shull, "The Chemical Bond in Molecular Quantum Mechanics," *The Journal of Chemical Physics*, vol. 35, no. 5, pp. 1644–1651, 1961.

[146] E. Prodan and W. Kohn, "Nearsightedness of electronic matter in one dimension," *Physical Review B*, vol. 73, no. 8, p. 085108, 2006.

[147] M. Ernst, L. H. Dos Santos, A. Krawczuk, P. Macchi, *et al.*, "Towards a Generalized Database of Atomic Polarizabilities," in *Understanding Intermolecular Interactions in the Solid State*, pp. 211–242, 2018.

[148] L. P. Hammett, "The effect of structure upon the reactions of organic compounds. Benzene derivatives," *Journal of the American Chemical Society*, vol. 59, no. 1, pp. 96–103, 1937.

[149] R. Bader, T. Keith, K. Gough, and K. Laidig, "Properties of atoms in molecules: additivity and transferability of group polarizabilities," *Molecular Physics*, vol. 75, no. 5, pp. 1167–1189, 1992.

[150] R. F. W. Bader, A. Larouche, C. Gatti, M. T. Carroll, P. J. MacDougall, and K. B. Wiberg, "Properties of atoms in molecules: Dipole moments and transferability of properties," *The Journal of Chemical Physics*, vol. 87, no. 2, pp. 1142–1152, 1987.

[151] T. Matsuda, Y. Funae, M. Yoshida, T. Yamamoto, and T. Takaya, "Optical material of high refractive index resin composed of sulfur-containing aliphatic and alicyclic methacrylates," *Journal of Applied Polymer Science*, vol. 76, no. 1, pp. 45–49, 2000.

[152] Y.-G. Ju, G. Almuneau, T.-H. Kim, and B.-W. Lee, "Numerical Analysis of High-Index Nano-Composite Encapsulant for Light-Emitting Diodes," *Japanese Journal of Applied Physics*, vol. 45, no. 4A, pp. 2546–2549, 2006.

[153] K. C. Krogman, T. Druffel, and M. K. Sunkara, "Anti-reflective optical coatings incorporating nanoparticles," *Nanotechnology*, vol. 16, no. 7, pp. S338–S343, 2005.

[154] B. D. Hatton, K. Landskron, W. J. Hunks, M. R. Bennett, D. Shukaris, D. D. Perovic, and G. A. Ozin, "Materials chemistry for low-k materials," *Materials Today*, vol. 9, no. 3, pp. 22–31, 2006.

[155] J.-g. Liu and M. Ueda, "High refractive index polymers: fundamental research and practical applications," *Journal of Materials Chemistry*, vol. 19, no. 47, p. 8907, 2009.

[156] J. Wang, M. Hutchins, K. Woo, T. Konish, and M. J. Idacavage, "Halogen-Free, Radiation-Curable, High-Refractive Index Materials," *RadTech Report*, vol. 24, no. 4, p. 11, 2010.

[157] C. Wang, D. Danovich, S. Shaik, and Y. Mo, "Halogen Bonds in Novel Polyhalogen Monoanions," *Chemistry - A European Journal*, vol. 23, no. 36, pp. 8719–8728, 2017.

[158] E. K. Macdonald and M. P. Shaver, "Intrinsic high refractive index polymers," *Polymer International*, vol. 64, no. 1, pp. 6–14, 2015.

[159] R. A. Gaudiana, R. A. Minns, and H. G. Rogers, "High refractive index polymers," 1992. US Patent 5,132,430.

[160] A. Stiegman, "High refractive index polymers," 2013. US Patent 8,470,948.

[161] C. L. Jones, B. U. Kolb, D. B. Olson, E. S. Goenner, S. D. Hart, and J. T. Cowher, "Optical film assembly and display device," 2009. US Patent 7,622,164.

[162] W. Caseri, "Nanocomposites of polymers and metals or semiconductors: historical background and optical properties," *Macromolecular Rapid Communications*, vol. 21, no. 11, pp. 705–722, 2000.

[163] D. J. Griffiths, *Introduction to Electrodynamics*. Pearson Benjamin Cummings, San Francisco, 3rd edn, 2008.

[164] A. I. Vogel, "Physical Properties and Chemical Constitution. Part XXIII," *Journal of the Chemical Society*, pp. 1833–1855, 1948.

[165] J. Applequist, "An atom dipole interaction model for molecular optical properties," *Accounts of Chemical Research*, vol. 10, no. 3, pp. 79–85, 1977.

[166] K. J. Miller, "Additivity methods in molecular polarizability," *Journal of the American Chemical Society*, vol. 112, no. 23, pp. 8533–8542, 1990.

[167] N. E.-B. Kassimi and A. J. Thakkar, "A simple additive model for polarizabilities: Application to amino acids," *Chemical Physics Letters*, vol. 472, no. 4-6, pp. 232–236, 2009.

[168] J. M. Stout and C. E. Dykstra, "Static Dipole Polarizabilities of Organic Molecules. Ab Initio Calculations and a Predictive Model," *Journal of the American Chemical Society*, vol. 117, no. 18, pp. 5127–5132, 1995.

[169] A. Stone, "Distributed polarizabilities," *Molecular Physics*, vol. 56, no. 5, pp. 1065–1082, 1985.

[170] A. Stone, "Distributed multipole analysis, or how to describe a molecular charge distribution," *Chemical Physics Letters*, vol. 83, no. 2, pp. 233–239, 1981.

[171] A. Stone and M. Alderton, "Distributed multipole analysis: methods and applications," *Molecular Physics*, vol. 56, no. 5, pp. 1047–1064, 1985.

[172] C. Ruth Le Sueur and A. J. Stone, "Localization methods for distributed polarizabilities," *Molecular Physics*, vol. 83, no. 2, pp. 293–307, 1994.

[173] G. Ivanov-Smolenskii, V. Tsirel'son, and R. Ozerov, "Determination of the diamagnetic susceptibility and the electron static polarizability of crystals from X-ray diffraction data," *Acta Crystallographica Section A: Foundations of Crystallography*, vol. 39, no. 3, pp. 411–415, 1983.

[174] K. Y. Suponitsky, V. G. Tsirelson, and D. Feil, "Electron-density-based calculations of intermolecular energy: case of urea," *Acta Crystallographica Section A Foundations of Crystallography*, vol. 55, no. 5, pp. 821–827, 1999.

[175] A. Krawczuk, D. Pérez, and P. Macchi, "PolaBer : a program to calculate and visualize distributed atomic polarizabilities based on electron density partitioning," *Journal of Applied Crystallography*, vol. 47, no. 4, pp. 1452–1458, 2014.

[176] L. H. R. Dos Santos, A. Krawczuk, and P. Macchi, "Distributed Atomic Polarizabilities of Amino Acids and their Hydrogen-Bonded Aggregates," *The Journal of Physical Chemistry A*, vol. 119, no. 13, pp. 3285–3298, 2015.

[177] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, 2010.

[178] T. Seidler, A. Krawczuk, B. Champagne, and K. Stadnicka, "QTAIM-Based Scheme for Describing the Linear and Nonlinear Optical Susceptibilities of Molecular Crystals Composed of Molecules with Complex Shapes," *The Journal of Physical Chemistry C*, vol. 120, no. 8, pp. 4481–4494, 2016.

[179] A. Volkov, X. Li, T. Koritsanszky, and P. Coppens, "Ab Initio Quality Electrostatic Atomic and Molecular Properties Including Intermolecular Energies from a Transferable Theoretical Pseudoatom Databank," *The Journal of Physical Chemistry A*, vol. 108, no. 19, pp. 4283–4300, 2004.

[180] J. F. Nye *et al.*, *Physical properties of crystals: their representation by tensors and matrices*. Oxford university press, 1985.

[181] N. Otero, C. Van Alsenoy, C. Pouchan, and P. Karamanis, "Hirshfeld-based intrinsic polarizability density representations as a tool to analyze molecular polarizability," *Journal of computational chemistry*, vol. 36, no. 24, pp. 1831–1843, 2015.

[182] A. Krawczuk-Pantula, D. Pérez, and P. Macchi, "Distributed atomic polarizabilities from electron density. 1. Motivations and Theory," *Trans Amer Cryst Ass*, vol. 42, pp. 1–25, 2012.

[183] F. G.A., M. J. Truck, G. W. Schlegel, H. B. Scuseria, G. E. Robb, M. A. Cheeseman, J. R. Scalmani, G. Barone, and V. Petersson, "Gaussian, Inc.." Wallingford, CT, 2016.

[184] N. M. O'Boyle, T. Vandermeersch, C. J. Flynn, A. R. Maguire, and G. R. Hutchison, "Confab - Systematic generation of diverse low-energy conformers," *Journal of Cheminformatics*, vol. 3, no. 1, p. 8, 2011.

[185] G. Zhou, S. Zhang, and C. Guan *Prog. Electromagn. Res. Symp.*, pp. 997–1002, 2009.

[186] P. Murray-Rust and S. Motherwell, "Computer retrieval and analysis of molecular geometry. I. General principles and methods," *Acta Crystallographica Section B Structural Crystallography and Crystal Chemistry*, vol. 34, no. 8, pp. 2518–2526, 1978.

[187] U. Maulik and S. Bandyopadhyay, "Genetic algorithm-based clustering technique," *Pattern Recognition*, vol. 33, pp. 1455–1465, 2000.

[188] P. L. A. Popelier and F. M. Aicken, "Atomic Properties of Selected Biomolecules: Quantum Topological Atom Types of Carbon Occurring in Natural Amino Acids and Derived Molecules," *Journal of the American Chemical Society*, vol. 125, no. 5, pp. 1284–1292, 2003.

[189] W. Hummel, J. Hauser, and H.-B. Bürgi, "PEANUT: Computer graphics program to represent atomic displacement parameters," *Journal of molecular graphics*, vol. 8, no. 4, pp. 214–220, 1990.

[190] M. Ernst, L. H. R. Dos Santos, and P. Macchi, "Optical properties of metal organic networks from distributed atomic polarizabilities," *CrystEngComm*, vol. 18, pp. 7339–7346, 2016.

[191] A. Krawczuk, "Influence of a hydrogen bond on optical properties of materials." Presentation at the European Crystallographic Meeting 32, Vienna, 2019.

[192] A. Krawczuk, H. L. Dos Santos, and P. Macchi, "to be published."

[193] D. M. Wilkins, A. Grisafi, Y. Yang, K. U. Lao, R. A. DiStasio, and M. Ceriotti, "Accurate molecular polarizabilities with coupled cluster theory and machine learning," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 116, no. 9, pp. 3401–3406, 2019. online: http://alphaml.org, accessed 1-10-2019.

[194] T. Leduc, E. Aubert, E. Espinosa, C. Jelsch, C. Iordache, and B. Guillot, "Polarization of Electron Density Databases of Transferable Multipolar Atoms," *The Journal of Physical Chemistry A*, vol. 123, no. 32, pp. 7156–7170, 2019.

[195] T. Poręba, M. Ernst, D. Zimmer, P. Macchi, and N. Casati, "Pressure-induced Polymerization and Electrical Conductivity of a Polyiodide," *Angewandte Chemie*, vol. 131, no. 20, pp. 6697–6701, 2019.

[196] P. Pelletier and J. B. Caventou, "Sur un nouvel Alcali vegetal (la Strychine) trouve dans la feve se Saint-Ignac, la noix vomique, etc.," *Ann. Chim Phys.*, vol. 10, p. 142, 1819.

[197] P. H. Svensson and L. Kloo, "Synthesis, structure, and bonding in polyiodide and metal iodide-iodine systems," *Chemical Reviews*, vol. 103, no. 5, pp. 1649–1684, 2003.

[198] R. A. Hoyt, E. M. Remillard, E. D. Cubuk, C. D. Vecitis, and E. Kaxiras, "Polyiodide-Doped Graphene," *The Journal of Physical Chemistry C*, vol. 121, no. 1, pp. 609–615, 2017.

[199] H. Haller and S. Riedel, "Recent discoveries of polyhalogen anions - From bromine to fluorine," *Zeitschrift fur Anorganische und Allgemeine Chemie*, vol. 640, no. 7, pp. 1281–1291, 2014.

[200] L. Kloo, J. Rosdahl, and P. H. Svensson, "On the Intra-and Intermolecular Bonding in Polyiodides," *European Journal of Inorganic Chemistry*, vol. 2002, no. 5, pp. 1203–1209, 2002.

[201] G. A. Landrum, N. Goldberg, and R. Hoffmann, "Bonding in the trihalides (X3-), mixed trihalides (X2Y-) and hydrogen bihalided (X2H-). The connection between hypervalent, electron-rich three-center, donor-acceptor and strong hydrogen bonding," *Journal of the Chemical Society, Dalton Transactions*, no. 19, pp. 3605–3613, 1997.

[202] K. Lamberts, P. Handels, U. Englert, E. Aubert, and E. Espinosa, "Stabilization of polyiodide chains via anion-anion interactions: experiment and theory," *CrystEngComm*, vol. 18, no. 21, pp. 3832–3841, 2016.

[203] E. Bartashevich, I. Yushina, K. Kropotina, S. Muhitdinova, and V. Tsirelson, "Testing the tools for revealing and characterizing the iodine-iodine halogen bond in crystals," *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials*, vol. 73, no. 2, pp. 217–226, 2017.

[204] D. Cremer and E. Kraka, "Chemical Bonds without Bonding Electron Density - Does the Difference Electron-Density Analysis Suffice for a Description of the Chemical Bond?," *Angewandte Chemie International Edition in English*, vol. 23, no. 8, pp. 627–628, 1984.

[205] O. A. Syzgantseva, V. Tognetti, and L. Joubert, "On the Physical Nature of Halogen Bonds: A QTAIM Study," *The Journal of Physical Chemistry A*, vol. 117, no. 36, pp. 8969–8980, 2013.

[206] V. Oliveira, E. Kraka, and D. Cremer, "The intrinsic strength of the halogen bond: Electrostatic and covalent contributions described by coupled cluster theory," *Physical Chemistry Chemical Physics*, vol. 18, no. 48, pp. 33031–33046, 2016.

[207] B. Li, Z. Nie, M. Vijayakumar, G. Li, J. Liu, V. Sprenkle, and W. Wang, "Ambipolar zinc-polyiodide electrolyte for a high-energy density aqueous redox flow battery," *Nature communications*, vol. 6, p. 6303, 2015.

[208] H. Abe, H. Kishimura, M. Takaku, M. Watanabe, and N. Hamaya, "Low-temperature and high-pressure phases of a room-temperature ionic liquid and polyiodides: 1-methyl-3-propylimidazolium iodide," *Faraday Discuss.*, vol. 206, pp. 49–60, 2018.

[209] J. S. Miller, *Extended linear chain compounds*, vol. 3. Springer Science & Business Media, 2012.

[210] A. Oza, "Electrical conduction in organic polyiodide chain complexes," *Crystal Research and Technology*, vol. 19, no. 5, pp. 697–707, 1984.

[211] Y. X. Ren, M. Liu, T. S. Zhao, L. Zeng, and M. Wu, "An aprotic lithium/polyiodide semi-liquid battery with an ionic shield," *Journal of Power Sources*, vol. 342, pp. 9–16, 2017.

[212] H. Yu, L. Yan, Y. He, H. Meng, and W. Huang, "An unusual photoconductive property of polyiodide and enhancement by catenating with 3-thiophenemethylamine salt," *Chemical Communications*, vol. 53, no. 2, pp. 432–435, 2017.

[213] A. Varvoglis, *Hypervalent iodine in organic synthesis*. Academic Press, 1997.

[214] A. Sengupta, E. L. Quitevis, and M. W. Holtz, "Effect of high pressure on vibrational modes of polyiodides in poly (vinyl alcohol) films," *The Journal of Physical Chemistry B*, vol. 101, no. 51, pp. 11092–11098, 1997.

[215] L. Alvarez, J.-L. Bantignies, R. Le Parc, R. Aznar, J.-L. Sauvajol, A. Merlen, D. Machon, and A. San Miguel, "High-pressure behavior of polyiodides confined into single-walled carbon nanotubes: a Raman study," *Physical Review B*, vol. 82, no. 20, p. 205403, 2010.

[216] C. R. Groom, I. J. Bruno, M. P. Lightfoot, and S. C. Ward, "The Cambridge Structural Database," *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials*, vol. 72, no. 2, pp. 171–179, 2016.

[217] K.-F. Tebbe and T. Gilles, "Verfeinerung der Kristallstruktur des Tetraethylammoniumheptaiodids," *Acta Crystallographica Section C: Crystal Structure Communications*, vol. 49, no. 11, pp. 2042–2045, 1993.

[218] Y. Wang, Y. Xue, X. Wang, Z. Cui, and L. Wang, "The stable polyiodides: Experimental and theoretical studies of formation mechanism," *Journal of Molecular Structure*, vol. 1074, pp. 231–239, 2014.

[219] P. H. Svensson, M. Gorlov, and L. Kloo, "Dimensional caging of polyiodides," *Inorganic chemistry*, vol. 47, no. 24, pp. 11464–11466, 2008.

[220] M. D. García, J. Martí-Rujas, P. Metrangolo, C. Peinador, T. Pilati, G. Resnati, G. Terraneo, and M. Ursini, "Dimensional caging of polyiodides: cation-templated synthesis using bipyridinium salts," *CrystEngComm*, vol. 13, no. 13, pp. 4411–4416, 2011.

[221] C. He, D. A. Parrish, and J. M. Shreeve, "Alkyl ammonium cation stabilized biocidal polyiodides with adaptable high density and low pressure," *Chemistry–A European Journal*, vol. 20, no. 22, pp. 6699–6706, 2014.

[222] A. Gräfe-Kavoosian, S. Nafepour, K. Nagel, and K.-F. Tebbe, "Untersuchungen an Polyhalogeniden, XXXVI. Zur Kenntnis des Oktaiodid-Ions I_8^{2-}," *Zeitschrift für Naturforschung B*, vol. 53, no. 7, pp. 641–652, 1998.

[223] M. Mizuno, J. Tanaka, and I. Harada, "Electronic spectra and structures of polyiodide chain complexes," *The Journal of Physical Chemistry*, vol. 85, no. 13, pp. 1789–1794, 1981.

[224] X. An, J. Sun, Z. Lu, F. Ma, and G. Zhang, "Pressure-induced insulator-semiconductor transition in bilayer hexagonal boron nitride," *Ceramics International*, vol. 43, no. 8, pp. 6626–6630, 2017.

[225] N. García, P. Esquinazi, J. Barzola-Quiquia, and S. Dusari, "Evidence for semiconducting behavior with a narrow band gap of Bernal graphite," *New Journal of Physics*, vol. 14, no. 5, p. 053015, 2012.

[226] A. Möller, T. Taetz, N. Hollmann, J. Mydosh, V. Kataev, M. Yehia, E. Vavilova, and B. Büchner, "Insulator to semiconductor transition and magnetic properties of the one-dimensional S=1/2 system In_2VO_5," *Physical Review B*, vol. 76, no. 13, p. 134411, 2007.

[227] S. Klotz, J. Chervin, P. Munsch, and G. Le Marchand, "Hydrostatic limits of 11 pressure transmitting media," *Journal of Physics D: Applied Physics*, vol. 42, no. 7, p. 075413, 2009.

[228] P. Vinet, J. Ferrante, J. Smith, and J. Rose, "A universal equation of state for solids," *Journal of Physics C: Solid State Physics*, vol. 19, no. 20, p. L467, 1986.

[229] R. J. Angel, M. Alvaro, and J. Gonzalez-Platas, "EosFit7c and a Fortran module (library) for equation of state calculations," *Zeitschrift für Kristallographie-Crystalline Materials*, vol. 229, no. 5, pp. 405–419, 2014.

[230] A. M. James and M. P. Lord, *Macmillan's chemical and physical data*. Macmillan, 1992.

[231] H.-B. Bürgi and J. D. Dunitz, *Structure correlation*. John Wiley & Sons, 2008.

[232] O. V. Dolomanov, L. J. Bourhis, R. J. Gildea, J. A. Howard, and H. Puschmann, "OLEX2: a complete structure solution, refinement and analysis program," *Journal of Applied Crystallography*, vol. 42, no. 2, pp. 339–341, 2009.

[233] A. M. Pendás, E. Francisco, and M. Blanco, "An electron number distribution view of chemical bonds in real space," *Physical Chemistry Chemical Physics*, vol. 9, no. 9, pp. 1087–1092, 2007.

[234] D. Cremer and E. Kraka, "A description of the chemical bond in terms of local properties of electron density and energy," *Croatica Chemica Acta*, vol. 57, no. 6, pp. 1259–1281, 1984.

[235] R. F. W. Bader, *Atoms in Molecules: A Quantum Theory. International Series of Monographs on Chemistry*. Wiley, Hoboken, 1994.

[236] A. Kovács, C. Esterhuysen, and G. Frenking, "The Nature of the Chemical Bond Revisited: An Energy-Partitioning Analysis of Nonpolar Bonds," *Chemistry–A European Journal*, vol. 11, no. 6, pp. 1813–1825, 2005.

[237] L. J. van der Paul, "A method of measuring specific resistivity and Hall effect of discs of arbitrary shape," *Philips Res. Rep*, vol. 20, pp. 220–224, 1958.

[238] W. Kubo, K. Murakoshi, T. Kitamura, S. Yoshida, M. Haruki, K. Hanabusa, H. Shirai, Y. Wada, and S. Yanagida, "Quasi-solid-state dye-sensitized TiO2 solar cells: effective charge transport in mesoporous space filled with gel electrolytes containing iodide and iodine," *The Journal of Physical Chemistry B*, vol. 105, no. 51, pp. 12809–12815, 2001.

[239] H. Mao, J.-A. Xu, and P. Bell, "Calibration of the ruby pressure gauge to 800 kbar under quasi-hydrostatic conditions," *Journal of Geophysical Research: Solid Earth*, vol. 91, no. B5, pp. 4673–4676, 1986.

[240] P. Willmott, D. Meister, S. Leake, M. Lange, A. Bergamaschi, M. Böge, M. Calvi, C. Cancellieri, N. Casati, A. Cervellino, *et al.*, "The materials science beamline upgrade at the Swiss Light Source," *Journal of synchrotron radiation*, vol. 20, no. 5, pp. 667–682, 2013.

[241] G. M. Sheldrick, "Crystal structure refinement with SHELXL," *Acta Crystallographica Section C: Structural Chemistry*, vol. 71, no. 1, pp. 3–8, 2015.

[242] R. Dovesi, R. Orlando, A. Erba, C. M. Zicovich-Wilson, B. Civalleri, S. Casassa, L. Maschio, M. Ferrabone, M. De La Pierre, P. D'Arco, *et al.*, "CRYSTAL14: A program for the ab initio investigation of crystalline solids," *International Journal of Quantum Chemistry*, vol. 114, no. 19, pp. 1287–1317, 2014.

[243] S. Grimme, "Semiempirical GGA-type density functional constructed with a long-range dispersion correction," *Journal of computational chemistry*, vol. 27, no. 15, pp. 1787–1799, 2006.

[244] J. Laun, D. Vilela Oliveira, and T. Bredow, "Consistent gaussian basis sets of double-and triple-zeta valence with polarization quality of the fifth period for solid-state calculations," *Journal of computational chemistry*, vol. 39, no. 19, pp. 1285–1290, 2018.

[245] M. F. Peintinger, D. V. Oliveira, and T. Bredow, "Consistent Gaussian basis sets of triple-zeta valence with polarization quality for solid-state calculations," *Journal of Computational Chemistry*, vol. 34, no. 6, pp. 451–459, 2013.

[246] M. N. Glukhovtsev, A. Pross, M. P. McGrath, and L. Radom, "Extension of Gaussian-2 (G2) theory to bromine-and iodine-containing molecules: Use of effective core potentials," *The Journal of chemical physics*, vol. 103, no. 5, pp. 1878–1885, 1995.

[247] F. Bertolotti, A. V. Shishkina, A. Forni, G. Gervasio, A. I. Stash, and V. G. Tsirelson, "Intermolecular Bonding Features in Solid Iodine," *Crystal Growth & Design*, vol. 14, no. 7, pp. 3587–3595, 2014.

[248] N. Greenwood and A. Earnshaw, "The Halogens: Fluorine, Chlorine, Bromine, Iodine and Astatine," *Chemistry of the Elements*, vol. 2, p. 804, 1997.

[249] P. Coppens, Z. Su, and P. Becker, "Analysis of charge and spin densities," *International tables for Crystallography*, 2006.

[250] T. A. Jones, J. Y. Zou, S. W. Cowan, and M. Kjeldgaard, "Improved methods for building protein models in electron density maps and the location of errors in these models," *Acta Crystallographica Section A*, vol. 47, no. 2, pp. 110–119, 1991.

[251] I. Bytheway, G. Chandler, B. Figgis, and D. Jayatilaka, "Are intramolecular dynamic electron correlation effects detectable in X-ray diffraction experiments on molecular crystals?," *Acta Crystallographica Section A: Foundations of Crystallography*, vol. 63, no. 2, pp. 135–145, 2007.

[252] T. Lu and F. Chen, "Multiwfn: a multifunctional wavefunction analyzer," *Journal of computational chemistry*, vol. 33, no. 5, pp. 580–592, 2012.

[253] O. L. Santos, T. L. Fonseca, J. R. Sabino, H. C. Georg, and M. A. Castro, "Polarization effects on the electric properties of urea and thiourea molecules in solid phase," *The Journal of Chemical Physics*, vol. 143, no. 23, p. 234503, 2015.

[254] T. A. Wesołowski and A. Warshel, "Frozen density functional approach for ab initio calculations of solvated molecules," *Journal of Physical Chemistry*, vol. 97, no. 30, pp. 8050–8053, 1993.

[255] T. A. Wesołowski, "Embedding a multideterminantal wave function in an orbital-free environment," *Physical Review A*, vol. 77, no. 1, p. 012504, 2008.

[256] A. Krawczuk and P. Macchi, "Charge density analysis for crystal engineering," *Chemistry Central Journal*, vol. 8, no. 1, p. 68, 2014.

[257] A. J. Edwards, C. F. Mackenzie, P. R. Spackman, D. Jayatilaka, and M. A. Spackman, "Intermolecular interactions in molecular crystals: what's in a name?," *Faraday Discussions*, vol. 203, pp. 93–112, 2017.

[258] P. Macchi, "Crystallographic Approaches for the Investigation of Molecular Materials: Structure Property Relationships and Reverse Crystal Engineering," *CHIMIA International Journal for Chemistry*, vol. 68, no. 1, pp. 31–37, 2014.

[259] A. Krawczuk and M. Gryl, "Qualitative and quantitative crystal engineering of multifunctional co-crystals," in *Multi-component crystals : synthesis, concepts, function* (E. Tiekink and J. Zukerman, eds.), pp. 59–102, De Gruyter, 2017.

[260] G. Saleh, C. Gatti, L. Lo Presti, and J. Contreras-García, "Revealing non-covalent interactions in molecular crystals through their experimental electron densities," *Chemistry - A European Journal*, vol. 18, no. 48, pp. 15523–15536, 2012.

[261] J. Contreras-García, E. R. Johnson, S. Keinan, R. Chaudret, J.-P. Piquemal, D. N. Beratan, and W. Yang, "NCIPLOT: A Program for Plotting Noncovalent Interaction Regions," *Journal of Chemical Theory and Computation*, vol. 7, no. 3, pp. 625–632, 2011.

[262] A. D. Becke and K. E. Edgecombe, "A simple measure of electron localization in atomic and molecular systems," *The Journal of Chemical Physics*, vol. 92, no. 9, pp. 5397–5403, 1990.

[263] D. Jayatilaka and D. J. Grimwood, "Wavefunctions derived from experiment. I. Motivation and theory," *Acta Crystallographica Section A: Foundations of Crystallography*, vol. 57, no. 1, pp. 76–86, 2001.

[264] I. Bytheway, D. J. Grimwood, and D. Jayatilaka, "Wavefunctions derived from experiment. III. Topological analysis of crystal fragments," *Acta Crystallographica Section A: Foundations of Crystallography*, vol. 58, no. 3, pp. 232–243, 2002.

[265] 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 Crystallographica Section A: Foundations of Crystallography*, vol. 58, no. 3, pp. 244–251, 2002.

[266] D. J. Grimwood, I. Bytheway, and D. Jayatilaka, "Wave functions derived from experiment. V. Investigation of electron densities, electrostatic potentials, and electron localization functions for noncentrosymmetric crystals," *Journal of Computational Chemistry*, vol. 24, no. 4, pp. 470–483, 2003.

[267] D. Jayatilaka, P. Munshi, M. J. Turner, J. A. K. Howard, and M. A. Spackman, "Refractive indices for molecular crystals from the response of X-ray constrained Hartree–Fock wavefunctions," *Physical Chemistry Chemical Physics*, vol. 11, no. 33, p. 7209, 2009.

[268] L. ęcińska, S. Mebs, B. Ośmiałowski, A. Zakrzewska, K. Ejsmont, and M. Kohout, "Tuning the Electronic Properties of the Dative N-B with Associated O-B Interaction: Electron Localizability Indicator from X-Ray Wavefunction Refinement," *ChemPhysChem*, vol. 17, no. 15, pp. 2395–2406, 2016.

[269] D. Jayatilaka and D. Grimwood, "Electron localization functions obtained from X-ray constrained Hartree-Fock wavefunctions for molecular crystals of ammonia, urea and alloxan," *Acta Crystallographica Section A: Foundations of Crystallography*, vol. 60, no. 2, pp. 111–119, 2004.

[270] S. Grabowsky, D. Jayatilaka, S. Mebs, and P. Luger, "The Electron Localizability Indicator from X-Ray Diffraction Data-A First Application to a Series of Epoxide Derivatives," *Chemistry - A European Journal*, vol. 16, no. 43, pp. 12818–12821, 2010.

[271] S. Grabowsky, M. Weber, D. Jayatilaka, Y. S. Chen, M. T. Grabowski, R. Brehme, M. Hesse, T. Schirmeister, and P. Luger, "Reactivity differences between α,β-unsaturated carbonyls and hydrazones investigated by experimental and theoretical electron density and electron localizability analyses," *Journal of Physical Chemistry A*, vol. 115, no. 45, pp. 12715–12732, 2011.

[272] A. Genoni, "Molecular Orbitals Strictly Localized on Small Molecular Fragments from X-ray Diffraction Data," *Journal of Physical Chemistry Letters*, vol. 4, pp. 1093–1099, 2013.

[273] A. Genoni, "X-ray Constrained Extremely Localized Molecular Orbitals: Theory and Critical Assessment of the New Technique," *Journal of Chemical Theory and Computation*, vol. 9, no. 7, pp. 3004–3019, 2013.

[274] A. Genoni and B. Meyer, "Chapter seventeen - x-ray constrained wave functions: Fundamentals and effects of the molecular orbitals localization," in *Electron Correlation in Molecules – ab initio Beyond Gaussian Quantum Chemistry* (P. E. Hoggan and T. Ozdogan, eds.), vol. 73 of *Advances in Quantum Chemistry*, pp. 333 – 362, Academic Press, 2016.

[275] A. Genoni, "A first-prototype multi-determinant X-ray constrained wavefunction approach: the X-ray constrained extremely localized molecular orbital–valence bond method," *Acta Crystallographica Section A Foundations and Advances*, vol. 73, no. 4, pp. 312–316, 2017.

[276] N. Casati, A. Genoni, B. Meyer, A. Krawczuk, and P. Macchi, "Exploring charge density analysis in crystals at high pressure: data collection, data analysis and advanced modelling," *Acta Crystallographica Section B Structural Science, Crystal Engineering and Materials*, vol. 73, no. 4, pp. 584–597, 2017.

[277] A. Genoni, D. Franchini, S. Pieraccini, and M. Sironi, "X-ray Constrained Spin-Coupled Wavefunction: a New Tool to Extract Chemical Information from X-ray Diffraction Data," *Chemistry - A European Journal*, vol. 24, no. 58, pp. 15507–15511, 2018.

[278] A. Genoni, G. Macetti, D. Franchini, S. Pieraccini, and M. Sironi, "X-ray constrained spin-coupled technique: theoretical details and further assessment of the method," *Acta Crystallographica Section A Foundations and Advances*, vol. 75, no. 6, pp. 778–797, 2019.

[279] S. F. Boys, "Construction of some molecular orbitals to be approximately invariant for changes from one molecule to another," *Reviews of Modern Physics*, vol. 32, no. 2, pp. 296–299, 1960.

[280] J. M. Foster and S. F. Boys, "Canonical Configurational Interaction Procedure," *Reviews of Modern Physics*, vol. 32, no. 2, pp. 300–302, 1960.

[281] C. Edmiston and K. Ruedenberg, "Localized Atomic and Molecular Orbitals," *Reviews of Modern Physics*, vol. 35, no. 3, pp. 457–464, 1963.

[282] C. Edmiston and K. Ruedenberg, "Localized Atomic and Molecular Orbitals. II," *The Journal of Chemical Physics*, vol. 43, no. 10, pp. S97–S116, 1965.

[283] W. Niessen, "Density localization of atomic and molecular orbitals," *Theoretica Chimica Acta*, vol. 27, no. 1, pp. 9–23, 1972.

[284] J. Pipek and P. G. Mezey, "Dependence ofMO shapes on a continuous measure of delocalization," *International Journal of Quantum Chemistry*, vol. 34, no. S22, pp. 1–13, 1988.

[285] J. Pipek and P. G. Mezey, "A fast intrinsic localization procedure applicable for ab-initio and semiempirical linear combination of atomic orbital wave functions," *The Journal of Chemical Physics*, vol. 90, no. 9, pp. 4916–4926, 1989.

[286] H. Stoll, G. Wagenblast, and H. Preuss, "On the Use of Local Basis Sets for Localized Molecular Orbitals," *Theoretica Chimica Acta*, vol. 57, pp. 169–178, 1980.

[287] A. Genoni, "X-ray Constrained Extremely Localized Molecular Orbitals: Theory and Critical Assessment of the New Technique," *Journal of Chemical Theory and Computation*, vol. 9, no. 7, pp. 3004–3019, 2013.

[288] B. Meyer, B. Guillot, M. F. Ruiz-Lopez, and A. Genoni, "Libraries of Extremely Localized Molecular Orbitals. 1. Model Molecules Approximation and Molecular Orbitals Transferability," *Journal of Chemical Theory and Computation*, vol. 12, no. 3, pp. 1052–1067, 2016.

[289] B. Meyer, B. Guillot, M. F. Ruiz-Lopez, C. Jelsch, and A. Genoni, "Libraries of Extremely Localized Molecular Orbitals. 2. Comparison with the Pseudoatoms Transferability," *Journal of Chemical Theory and Computation*, vol. 12, no. 3, pp. 1068–1081, 2016.

[290] B. Meyer and A. Genoni, "Libraries of Extremely Localized Molecular Orbitals. 3. Construction and Preliminary Assessment of the New Databanks," *The Journal of Physical Chemistry A*, vol. 122, no. 45, pp. 8965–8981, 2018.

[291] D. Arias-Olivares, E. K. Wieduwilt, J. Contreras-García, and A. Genoni, "NCI-ELMO: A New Method To Quickly and Accurately Detect Noncovalent Interactions in Biosystems," *Journal of Chemical Theory and Computation*, vol. 15, no. 11, pp. 6456–6470, 2019.

[292] G. Macetti and A. Genoni, "Quantum Mechanics/Extremely Localized Molecular Orbital Method: A Fully Quantum Mechanical Embedding Approach for Macromolecules," *The Journal of Physical Chemistry A*, vol. 123, no. 43, pp. 9420–9428, 2019.

[293] L. A. Malaspina, E. K. Wieduwilt, J. Bergmann, F. Kleemiss, B. Meyer, M. F. Ruiz-López, R. Pal, E. Hupf, J. Beckmann, R. O. Piltz, A. J. Edwards, S. Grabowsky, and A. Genoni, "Fast and Accurate Quantum Crystallography: From Small to Large, from Light to Heavy," *The Journal of Physical Chemistry Letters*, vol. 8571, pp. 6973–6982, 2019.

[294] N. Casati, P. Macchi, and A. Sironi, "Hydrogen migration in oxalic acid di-hydrate at high pressure?," *Chemical Communications*, no. 19, p. 2679, 2009.

[295] M. F. Guest, I. J. Bush, H. J. J. Van Dam, P. Sherwood, J. M. H. Thomas, J. H. Van Lenthe, R. W. A. Havenith, and J. Kendrick, "The GAMESS-UK electronic structure package: algorithms, developments and applications," *Molecular Physics*, vol. 103, no. 6-8, pp. 719–747, 2005.

[296] Y. Wang, C. J. Tsai, W. L. Liu, and L. D. Calvert, "Temperature-dependence studies of α-oxalic acid dihydrate," *Acta Crystallographica Section B Structural Science*, vol. 41, no. 2, pp. 131–135, 1985.

[297] F. Montisci, "High Pressure as a Tool for Crystal Engineering," *Dissertation at the University of Bern*, pp. 113–115, 2019.

[298] J. Dunitz, J. Lehn, G. Wipff, *et al.*, "Stereochemistry of reaction paths at carbonyl centres," *Tetrahedron*, vol. 30, no. 12, pp. 1563–1572, 1974.

[299] M. G. Voronkov and V. P. Baryshok, "Atranes as a new generation of biologically active substances," *Herald of the Russian Academy of Sciences*, vol. 80, no. 6, pp. 514–521, 2010.

[300] E. Müller and H.-B. Bürgi, "58. Complexes of 2,2',2"-Nitrilotriphenol," *Helvetica Chimica Acta*, vol. 70, pp. 499–510, 1987.

[301] P. Moeckli, D. Schwarzenbach, H.-B. Bürgi, J. Hauser, and B. Delley, "Charge-density study of boron nitrilotriacetate, $C_6H_6BNO_6$, at 100 K: a comparison of polar bonds," *Acta Crystallographica Section B Structural Science*, vol. 44, no. 6, pp. 636–645, 1988.

[302] F. L. Hirshfeld, "Bonded-atom fragments for describing molecular charge densities," *Theoretica Chimica Acta*, vol. 44, no. 2, pp. 129–138, 1977.

[303] R. F. W. Bader, "Bond Paths Are Not Chemical Bonds," *The Journal of Physical Chemistry A*, vol. 113, no. 38, pp. 10391–10396, 2009.

[304] P. Politzer and J. S. Murray, "A look at bonds and bonding," *Structural Chemistry*, vol. 30, no. 4, pp. 1153–1157, 2019.

[305] H. Jacobsen, "Chemical bonding in view of electron charge density and kinetic energy density descriptors," *Journal of Computational Chemistry*, vol. 30, no. 7, pp. 1093–1102, 2009.

[306] M. Rahm and R. Hoffmann, "Distinguishing Bonds," *Journal of the American Chemical Society*, vol. 138, no. 11, pp. 3731–3744, 2016.

[307] D. L. Heinz and R. Jeanloz, "The equation of state of the gold calibration standard," *Journal of applied physics*, vol. 55, no. 4, pp. 885–893, 1984.

[308] C. Chantler, C. Tran, and D. Cookson, "Precise measurement of the lattice spacing of LaB 6 standard powder by the x-ray extended range technique using synchrotron radiation," *Physical Review A*, vol. 69, no. 4, p. 042101, 2004.

[309] A. Le Bail, "Whole powder pattern decomposition methods and applications: A retrospection," *Powder Diffraction*, vol. 20, no. 4, pp. 316–326, 2005.

[310] A. A. Coelho, "TOPAS and TOPAS-Academic: an optimization program integrating computer algebra and crystallographic objects written in C++," *Journal of Applied Crystallography*, vol. 51, no. 1, pp. 210–218, 2018.

[311] R. W. Hooft, L. H. Straver, and A. L. Spek, "Determination of absolute structure using Bayesian statistics on Bijvoet differences," *Journal of Applied Crystallography*, vol. 41, no. 1, pp. 96–103, 2008.

[312] A. L. Spek, "Structure validation in chemical crystallography," *Acta Crystallographica Section D Biological Crystallography*, vol. 65, no. 2, pp. 148–155, 2009.

Declaration of consent

on the basis of Article 28 of the RSL Phil.-nat. 05

Name/First Name: Michelle Ernst

Registration Number: 10-209-153

Study program: Chemistry

Bachelor ☐ Master ☐ Dissertation ☑

Title of the thesis: A Quantum Crystallographic Approach to Study Properties of Molecules in Crystals

Supervisor: Piero Macchi

I declare herewith that this thesis is my own work and that I have not used any sources other than those stated. I have indicated the adoption of quotations as well as thoughts taken from other authors as such in the thesis. I am aware that the Senate pursuant to Article 36 paragraph 1 litera r of the University Act of 5 September, 1996 is authorized to revoke the title awarded on the basis of this thesis.

For the purposes of evaluation and verification of compliance with the declaration of originality and the regulations governing plagiarism, I hereby grant the University of Bern the right to process my personal data and to perform the acts of use this requires, in particular, to reproduce the written thesis and to store it permanently in a database, and to use said database, or to make said database available, to enable comparison with future theses submitted by others.

Place/Date

Signature

Michelle Ernst

Curriculum Vitae

Im Biel 15
5014 Gretzenbach
✉ *michelle.ernst@dcb.unibe.ch*

Personal Information

Citizenship Swiss

Date of birth 9th April 1992

Current Position

Aug.2016-Feb.2020 **Ph.D. Student in Chemistry**, *Department of Chemistry and Biochemistry*, University of Bern, Freiestrasse 3, Bern 3012, Switzerland.

Supervisor Prof. Piero Macchi

Description Investigation of properties and interactions of molecules in crystals by means of quantum crystallographic methodologies.

Education

2014–2016 **M.Sc. in Chemistry**, *University of Bern*, Switzerland.
Specification in Material Science and Physical Chemistry

Thesis Title Predicting the Refractive Index of Metal-Organic Networks

2013–2014 Additional studies in sport sciences to qualify for the teaching diploma, *University of Fribourg*, Switzerland.

2010–2013 **B.Sc. in Chemistry and Sport Sciences**, *University of Fribourg*, Switzerland.

2006–2010 **Matura**, *Kantonsschule Olten*, Switzerland.
Specification in mathematics and physics

Language and computer skills

German Mother tongue

English Fluent

French Fluent

Italian Intermediate

Fortran Intermediate skills

Experiences related to chemistry

2019 Assistent for the excersises of the lecture "Symmetrielehre", *University of Bern*

2017 Tutor for the general chemistry lab, *University of Bern*

2016 Tutor for the general chemistry lab, *University of Bern*

2014 Teaching quantum chemistry exercises to Bachelor students, *University of Fribourg*

2014 Teaching chemistry to high school students, *Kollegium St. Michel, Fribourg*

Schools & Conferences

2019 **Tools for Chemical Bonding**, *Bremen*, Germany.
Poster presentation

2018 **Annual Meeting and General Assembly of the Swiss Society for Crystallography (SGK)**, *PSI, Villigen*, Switzerland.
Poster presentation, awarded with poster price

2018 **Erice International School of Crystallography, Quantum Crystallography**, *Erice, Sicily*, Italy.
Oral and poster presentation

2017 **24th Congress and General Assembly of the International Union of Crystallography**, *Hyderabad*, India.
Oral presentation

2017 **Annual Meeting and General Assembly of the Swiss Society for Crystallography (SGK)**, *Geneva*, Switzerland.
Poster presentation, awarded with poster price

2016 **30th Meeting of the European Crystallographic Association**, *Basel*, Switzerland.

2016 **First Robert F. Stewart school on electron density and related properties**, *Nancy*, France.

2015 **Annual Meeting and General Assembly of the Swiss Society for Crystallography (SGK)**, *Neuchâtel*, Switzerland.

Awards

- "Travel Award for Young Academics" by the Faculty of Science of the University of Bern which allowed for a stay of one month (October 2018) at the University of Lorraine in the research group of Dr. Alessandro Genoni.

Publications

- Michelle Ernst, Alessandro Genoni, Piero Macchi; **Analysis of Crystal Field Effects and Interactions with X-ray Restrained ELMOs** submitted to the *Journal of Molecular Structure*; *2019*.

- Tomasz Poręba, Michelle Ernst, Dominik Zimmer, Piero Macchi, Nicola Casati; **Pressure-induced Polymerization and Electrical Conductivity of a Polyiodide** *Angewandte Chemie International Edition,* 58, 6625-6629; *2019*.

- Michelle Ernst, Leonardo H. R. Dos Santos, A. Krawczuk, Piero Macchi; **Towards a generalized database of atomic polarizabilities** *Chapter 7 in 'Understanding Intermolecular Interactions in Crystals'; Royal Society of Chemistry, 2018*.

- Michelle Ernst, Leonardo H. R. Dos Santos, Piero Macchi; **Optical properties of metal-organic networks from distributed atomic polarizabilities** *CrystEngComm,* 18, 7339–7346; *2016*.