Variational Information-Theoretic Atoms-in-Molecules

VARIATIONAL INFORMATION-THEORETIC ATOMS-IN-MOLECULES

BY

FARNAZ HEIDAR-ZADEH, B.Sc., M.Sc.

A Thesis Submitted to the School of Graduate Studies of McMaster University and Faculty of Sciences of Ghent University In Fulfillment of the Requirements for the Joint Degree of Doctorate in Sciences.

McMaster University DOCTOR OF PHILOSOPHY (2017) Hamilton, Ontario (Sciences)

TITILE	Variational Information-Theoretic Atoms-In-Molecules
AUTHOR	Farnaz Heidar-Zadeh, M.Sc., B.Sc. (Shahid Beheshti University)
SUPERVISORS	Prof. Paul W. Ayers (McMaster University)
	Prof. Patrick Bultinck & Prof. Toon Verstraelen (Ghent University)
PAGES	xiv, 206

Dedicated to my parents, Manouchehr and Elaheh,

for their endless love, support and sacrifice.

Abstract

It is common to use the electron density to partition a molecular system into atomic regions. The necessity for such a partitioning scheme is rooted in the unquestionable role of atoms in chemistry. Nevertheless, atomic properties are not well-defined concepts within the domain of quantum mechanics, as they are not observable. This has resulted in a proliferation of different approaches to retrieve the concept of atoms in molecules (AIM) within the domain of quantum mechanics and *in silico* experiments based on various flavors of model theories.

One of the most popular families of models is the Hirshfeld, or stockholder, partitioning methods. Hirshfeld methods do not produce sharp atomic boundaries, but instead distribute the molecular electron density at each point between all the nuclear centers constituting the molecule. The various flavors of the Hirshfeld scheme differ mainly in how the atomic shares are computed from a reference promolecular density and how the reference promolecular density is defined.

We first establish the pervasiveness of the Hirshfeld portioning by extending its information-theoretic framework. This characterizes the family of *f*-divergence measures as necessary and sufficient for deriving Hirshfeld scheme. Then, we developed a variational version of Hirshfeld partitioning method, called Additive Variational Hirshfeld (AVH). The key idea is finding the promolecular density, expanded as a linear combination of charged and neutral spherically-averaged isolated atomic densities in their ground and/or excited states, that resembles the molecular density as much as possible. Using Kullback-Liebler divergence measure, this automatically guarantees that each atom and proatom have the same number of electrons, and that the partitioning is size consistent. The robustness of this method is confirmed by testing it on various datasets. Considering the mathematical properties and our numerical results, we believe that AVH has the potential to supplant other Hirshfeld partitioning schemes in future.

Acknowledgements

I would like to take this opportunity to express my heartfelt gratefulness to my supervisor Prof. Paul W. Ayers for patiently guiding me through all the ups and downs of graduate school. I would not have been here today without his constant support, scientific advice, and selfless commitment to help his students thrive. I was indeed fortunate to work with him, and his lifelong impact has, without doubt, turned me into a better person and a better scientist and has inspired me to work harder everyday.

In addition, I would like to appreciate my mentors in Ghent University who have always made me feel at home in their groups. I am thankful to Prof. Patrick Bultinck and Prof. Toon Verstraelen for their continuous support and guidance. Specifically, I would like to appreciate Prof. Toon Verstraelen for teaching me how to code.

I would like to acknowledge my international hosts and collaborators for welcoming me in their laboratories: Prof. Esteban Vöhringer-Martinez, Prof. Carlos Cárdenas, Prof. Patricio Fuentealba, Prof. Alberto Vela and Prof. José L. Gázquez.

I owe my deepest gratitude to my beloved parents, Manouchehr and Elaheh, for always believing in me and supporting my dreams, for their continuous encouragement and understanding at every stage in my life. I am lucky to have an amazing sister and brother, Farnoush and Assad, whose love, presence and support sustained me through this long journey, as well as my wonderful brother-in-law, Behrang, whose kind support has always helped me rise.

Last but not least, I would like to thanks Vanier-CGS, NSERC, OGS, McMaster University and Ghent University for their generous financial support.

Table of Contents

Table of Tables

Table of Figures

1 Introduction

1.1 Background

The periodic table of elements is the touchstone of chemistry. It encapsulates the idea that atoms are the building blocks of molecules, and that the properties of molecules are determined by the identity of their constituent atoms. Unfortunately, there is no universally accepted definition for an atom within a molecule.[1-5] This has induced a proliferation of methods for decomposing molecules into atomic subsystems. These methods can be classified based on whether they partition the molecule by dividing the wavefunction in Hilbert space (*e.g.*, the orbital-based approaches of Mulliken[6-9], Löwdin[10-12], Moffitt[13], Weinhold[14-15], Ruedenberg[16-18], and Knizia[19]) or by dividing a molecular descriptor in real space (*e.g.*, the electron-density-based approaches of Politzer[20], Hirshfeld[21], and Bader[22-23]). These methods can also be classified based on whether they are binary (i.e., points in real space, or basis functions in Hilbert space, are fully assigned to a single atom) or fuzzy (i.e., points/basis functions can be shared by several atoms).

Given this imbroglio, it becomes desirable to establish guidelines for developing and assessing atoms-in-molecules (AIM) methods. It is our view that, given the preeminence of the periodic table in chemistry, AIM should be chosen to resemble the isolated atoms enshrined in the periodic table to the greatest possible extent, subject to the

defining constraint that the AIM provide an exhaustive partitioning of the molecule. If, as is conventional, we choose to use the electron density as the fundamental descriptor, then we wish to minimize the dissimilarity D between the electron density of the AIM, $\left\{\rho_A(\mathbf{r})\right\}_{A=1}^{N_{\text{atoms}}}$, and the corresponding electron density of the reference pro-atoms, $\left\{\rho_A^0(\mathbf{r})\right\}_{A=1}^{N_{\text{atoms}}}$,

$$\sum_{A=1}^{N_{\text{atoms}}} D\left[\rho_A, \rho_A^0\right] \tag{1.1}$$

subject to the constraint that the sum of the electron densities of the AIM is equal to the molecular density, $\rho_{\text{mol}}(\mathbf{r})$,

$$\rho_{\text{mol}}(\mathbf{r}) = \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r}) \tag{1.2}$$

and possibly other constraints. In this framework, different partitioning approaches are distinguished by their choice of reference pro-atoms, dissimilarity measure, and imposed constraints on the minimization in Eq. (1.1). Among these measures of dissimilarity, those based on information theory are privileged because they regard the electron density as a probability distribution function, rather than merely a function in Hilbert (or, preferably, Banach) space.[24-27] As we shall see, this imparts desirable features upon information-theoretic partitionings. Many of these desirable features are inherited by the somewhat more general class of dissimilarity measures known as the *f*-divergences. [28-29]

In this introductory chapter, we will first establish sets of criteria that are believed to make an AIM method preferable. In section 1.3, we will discuss various information-theoretic partitioning methods, emphasizing the Hirshfeld family of methods, where the

pro-atom densities are commonly built from the electron densities of the isolated atoms and their ions. In section 1.4, we briefly overview the other popular ways of determining the densities and/or populations of AIM. These are used in section 1.5 to compare atomic populations from different approaches for a few molecules, which we chose to demonstrate key strengths and weaknesses of different methods. Section 1.6 outlines our various efforts to remedy the shortcomings of current Hirshfeld partitioning schemes, which constitutes the content of this thesis.

1.2 Desirable Traits of Atoms In Molecules

Because the atom-in-a-molecule is not a physically observable object, but merely a human-defined object of conceptual utility, it is impossible to say that any specific definition of an AIM is "better" than any other. One can only indicate that a specific definition is more useful in a certain context. (Even so, an atomic partitioning's utility is often strongly dependent on the priorities and biases of the assessor.) The perceived utility of a partitioning method for a given purpose will depend on a balance between its mathematical features (its desirable formal properties) and its chemical utility (its ability to reify chemical intuition and/or elucidate new chemical phenomena). Some partitioning methods are mathematically beautiful but challenge many chemists' intuition. (Bader's Quantum Theory of Atoms in Molecules (QTAIM) is one example.[22, 30]) Other partitioning methods seem somewhat contrived mathematically but apparently give results in excellent agreement with what chemists expect. (The most recent versions of the Density Derived Electrostatic and Chemical (DDEC) net atomic charges are

examples.[31]) To set the stage for developing, assessing and comparing partitioning methods, a list of mathematical, chemical and computational desiderata is introduced below. This, admittedly biased, set of characteristics are what we believe makes the AIM properties reliable. Some of these features resemble the performance goals set by Manz *et. al.* in the recent development of the DDEC6 charge partitioning algorithm.[32-33] The fundamental difference in our methodologies is the strategy one employs to comply with these features, whereas Manz *et. al.* develop methods using a scientific engineering/design approach, we aspire to mathematical elegance and sound theoretical reasoning.

1.2.1 Desirable Mathematical Features

Universality: The AIM partitioning should be definable for any system (including molecules, infinite periodic solids, and infinite disordered systems; including neutral closed-shell molecules, charged systems, and systems with unpaired electrons; including ground and excited electronic states; including equilibrium geometries and strained structures). The partitioning should be computable from any reasonable quantum-mechanical method (Slater-determinant-based methods, correlated wavefunction methods, quantum Monte Carlo, etc.) and any reasonable representation of the molecular wavefunction (basis-set expansion, values on a numerical grid, etc.). Using pseudopotentials and/or taking relativistic effects into account (using the Dirac equation or one of its simplifications) should not be problematic. Going beyond the Born-Oppenheimer approximation or considering

exotic systems should not cause any problem either. (Consider, *e.g.*, recent work extending quantum theory of atoms in molecules (QTAIM) to systems composed of various types of quantum particles.[34-35]) After the atomic partitioning has been performed, *every* atomic property—not just atomic populations and higher electrostatic moments—should be defined in a way that is consistent with the precepts of quantum mechanics.

Foundation in Quantum Mechanics: The AIM partitioning method should have a firm grounding in quantum mechanics. Ideally one should be able to construct a full quantum mechanical framework for the atomic subsystems, as attempted in the quantum theory of atoms in molecules (QTAIM).[22, 30, 36] Failing this, the AIM partitionings should at least be rigorously defined in terms of quantum mechanical observables. (Methods based on quantum-mechanical observables generally meet the aforementioned stipulation of "universality.") Among these, the electron density is special; it quantifies the probability of observing an electron at a point in space, so it is conceptually appealing to define the probability of observing an electron on an atom (i.e., the atomic population) using only the electron density. It also ensures that one's partitioning has a quantum-mechanical basis and, in particular, that all quantum-mechanical observables of an AIM can be computed (using the framework of density-functional theory[37-39]). Extensions of the electron density (e.g., to other types of particles) allow exotic molecules and non-Born-Oppenheimer effects to be included.[34-35, 40-43]

Variational Optimality: The AIM should be the "best possible atoms" in some specified

way. This guides potential users: if an atomic partitioning method is optimal in the way one finds beneficial, it is a good choice for one's problems. If not, then one can seek a partitioning method whose philosophy is more aligned with one's needs. At a practical level, when AIMs are defined by a variational principle, it is straightforward to add constraints. (For example, it is sometimes useful to force the charges on amino acid residues in a polypeptide to be integers, or to force the charges on some atoms to equal their values from a molecular mechanics force field). It is preferable for the optimization problem to be convex, so the variational principle does not have multiple local minima.

Uniqueness: The partitioning should fully and uniquely specify the AIM. One should not have multiple solutions to the equations, or multiple minima in the objective function(s) defining the AIM. This is not only mathematically desirable, but avoids the numerical difficulties and computational expense associated with global optimization. In addition, it eliminates the biased disposal of undesirable solutions, like discarding well-defined Quantum Divided Basins (QDB) in QTAIM in favour of topological atoms.[44-45] A unique solution leaves no room for imposing (possibly controversial) chemical intuition in selecting the relevant answer.

Bias-free: The partitioning method should not require any input beyond the identity of the system being partitioned and its wavefunction. This ensures that the method is immune from any bias that a user may have, and makes it impossible to fudge results. If reference data (e.g., reference densities or reference wavefunctions for

atoms) are necessary, these reference functions should be directly determined by the identity of the atoms that compose the molecule. That is, the reference functions should be prescribed by a simple and physically-motivated procedure that is amenable to automation, and which is not subject to human intervention or bias.

Elegance: The principle of Occam's razor indicates that among all methods with similar performance, the simplest and most elegant method is to be preferred. While mathematical elegance is impossible to quantify, conceptually simple methods that have a compact mathematical description (even if the actual implementation is quite complicated) tend to be elegant. Methods based on variational principles are elegant. Elegant methods work "out of the box," requiring no special knowledge or experience. When intrinsically elegant methods are combined, elegance is compromised. Elegance is also compromised when one must engineer corrections or modifications to an underlying algorithm to explicitly account for "boundary cases" or "exceptions." (Not only is this inelegant, it is also dangerous: the unfathomable diversity of chemistry means that no one can possibly anticipate all the potential problems. If one must explicitly correct for one sort of problem, there are probably other, unanticipated, problems still lurking.) In general, an elegant method can be explained completely in a sentence or two, so much so that one can fully implement the method from its verbal description.

The next two criteria are most relevant to information-theoretic partitioning methods, though similar considerations are sometimes pertinent for other partitioning strategies.

Non-interacting Limit of AIM: If one builds the molecular density by superposing the densities of the isolated atoms (or molecular fragments), then the AIMs' populations should revert to the populations of the pro-atoms (or fragments). That is, if the molecular density is equal to the promolecular density, then the AIMs' densities should be equal to the pro-atoms' densities. This requirement imposes size-consistency (partitioning a molecule composed of noninteracting fragments gives the same results as partitioning the fragments separately), ensuring that separated-atom/fragment limits are sensible and that weak chemical interactions do not induce large shifts in atomic charges.

Distributive Property of Dissimilarity Measure: Since the sum of the AIM densities is the molecular density and the sum of the pro-atomic densities is the promolecular density, the sum of the dissimilarities between the AIMs and pro-atoms should equal the dissimilarity between the molecule and the promolecule. I.e.,

$$D\left[\rho_{\text{mol}};\rho^0_{\text{mol}}\right] = D\left[\sum_{A=1}^{N_{\text{atoms}}}\rho_A;\sum_{A=1}^{N_{\text{atoms}}}\rho^0_A\right] = \sum_{A=1}^{N_{\text{atoms}}} D\left[\rho_A,\rho^0_A\right] \tag{1.3}$$

Since the *raison d'être* of atomic partitioning is to quantify and guide chemists' intuition about atomic properties, the ultimate test of an atomic partitioning is its utility and consistency with respect to chemical observations and the intuitive framework employed by chemists. This leads to the following desiderata, reflecting the chemical application and computational practice of atomic partitioning methods:

1.2.2 Desirable Chemical Features

Chemical Robustness: Atomic charges should be in broad agreement with chemists' expectations based on empirically established atomic electronegativities and oxidation states. That is, the atomic partitioning method should give sensible results for neutral and charged molecules (even highly-charged molecules) as well as molecular excited states. Exact integer charges should never occur, except where required by symmetry or in the infinite separation limit. Atomic partitioning is commonly combined with reactivity indicators to determine the regioselectivity of molecular sites. It is expected for the AIM charges to comply with experimental data on functional group reactivity. Without these features, an atomic partitioning method is unlikely to be useful for elucidating chemical trends.[5, 46]

Transferability: The most important chemical trend is the transferability of atoms and functional groups between similar molecular environments. (This is, in fact, the original motivation for the concept of molecules as being composed of atoms, and it is the primary reason for the utility of the periodic table.) Simply stated: atoms and functional groups in similar environments should have similar properties, and these properties should vary in a systematic way in response to changes in the molecular environment.

Conformational Stability: In accord with chemical intuition, changes in molecular conformation, especially torsional motions and relatively unhindered rotations

around bonds, should not cause large fluctuations in the charges or other properties of the atoms. This property is essential when a partitioning method is used to parameterize a molecular mechanics force field.

Locality and Sensible AIM Densities: The AIMs should be localized around the atomic nucleus, and should not have intricate structures far from their defining nuclear center. This requirement is usually necessary, albeit insufficient, for chemical transferability and conformational stability. It is also expected that each AIM's density should have one and only one cusp, located at the position of the atomic nucleus.[47-48] As one moves away from the atomic nucleus, the atomic density is expected to decrease monotonically.[49] Far from the atomic nuclei, all the atomic densities should share the same asymptotic decay rate, in accord with the electronegativity equalization principle.[50-52]

Accurate Electrostatic Potential: Since the AIMs' "partial charges" are most commonly used to identify the positive and negative regions of a molecule, the electrostatic potential approximated by the AIM charges should accurately approximate the true molecular electrostatic potential on and outside molecular van der Waals surface,

$$\sum_{A=1}^{N_{\text{atoms}}} \frac{Z_A}{\left|\mathbf{r}-\mathbf{R}_A\right|} - \int \frac{\rho_{\text{mol}}\left(\mathbf{r}'\right)}{\left|\mathbf{r}-\mathbf{r}'\right|} d\mathbf{r}' \sim \sum_{A=1}^{N_{\text{atoms}}} \frac{Z_A - N_A}{\left|\mathbf{r}-\mathbf{R}_A\right|} \tag{1.4}$$

where the AIM's population corresponding to atomic number Z_A at position $\mathbf{R}_A$ is defined as

$$N_A = \int \rho_A(\mathbf{r}) d\mathbf{r} \tag{1.5}$$

and the AIM's partial charge is therefore

$$q_A = Z_A - N_A. \tag{1.6}$$

Moreover, if Eq. (1.4) is refined by including contributions from the atomic multipole moments, the electrostatic potential determined by the atomic multipole expansion should rapidly converge to the true molecular electrostatic potential. This, in practice, means that the AIM densities must be nearly spherical. Only atomic partitioning methods that accurately reproduce the long-range portion of the molecular electrostatic potential are convenient for parameterizing molecular mechanics force fields. Partitioning methods that satisfy the distributive property of the dissimilarity measure, Eq. (1.3), tend to give accurate electrostatic potentials.

This feature is specifically important as the primary *quantitative* application of AIM methods is the parameterization of molecular mechanics (MM) force fields to model electrostatic interactions.[53] Much chemical intuition is based on the picture of a molecule as composed of atomic sites with partial charges, linked by spring-like bonds and bond angles, along with rocking motions and hindered rotations around bonds. It is desirable that an AIM partitioning be consistent with this description. The preceding desirable chemical features are also strongly linked to the applicability of a partitioning method for MM parameterization. Note, however, that for degenerate ground states, one needs not just one molecular density (and its underlying atomic densities and charges[54]), but *all* the possible

degenerate molecular densities in order to successfully model the electrostatic potential.[54-57]

1.2.3 Desirable Computational Features

Computational Robustness: The partitioning method should be insensitive to changes in computational parameters. For example, the method should be robust to changes in the electronic structure method (Hartree-Fock, Kohn-Sham DFT, post Hartree-Fock, etc.) and changes in the molecular basis set (even pernicious choices like single-center expansion). (One exception: if improving the electronic structure method or the basis set causes the molecular electronic density to *qualitatively* change, the atomic properties may also change qualitatively.) The method should not be overly sensitive to the choice of initial guess, optimization strategy, and numerical integration grid. Indeed, ideally all integrals could be performed analytically.

Computational Efficiency: The equations that define the atomic partitioning can be solved efficiently and rapidly. Gigantic integration grids should not be required; systems of (non)linear equations that arise should be well-conditioned and have unique solutions; iterative procedures should converge quickly and inexorably to the solution. In addition, partitioning methods should be applicable to large systems like bulk solids, biological networks, and molecular dynamic simulations. In this regard, the possibility for localizing the algorithm or parallelizing the method boosts the performance and applicability.

We know of no single method that possesses all of these desirable properties and, indeed, we suspect that no such method exists. One is forced to compromise. This is why an axiomatic approach to atomic partitioning exists—first specify the features one finds most desirable; then find the partitioning method(s) that possess those features.[27]

However, as will be established in the following sections, the information-theoretic partitioning methods fulfill many of these desiderata: they are universally defined; they are founded in quantum mechanics (though not as fundamentally as claimed by Bader's QTAIM); they are electron-density based; the AIM are uniquely defined by a variational principle. Information-theoretic approaches tend to satisfy the distributive property of dissimilarity measure and the non-interacting limit of AIM constraint whenever these constraints are sensible. While no single information-theoretic method possesses all of the remaining virtues, most information-theoretic methods possess the majority of these features. One exception is the requirement of sensible atomic densities: information-theoretic methods typically have atomic densities with very small, but nonetheless spurious, cusps on other atoms.

1.3 Information-Theoretic Partitionings

1.3.1 The Hirshfeld Partitioning

The genesis of information-theoretic partitioning methods can be traced back to the 1970 paper of Politzer and Harris, who defined a binary real-space partitioning of linear molecules based on the promolecular density.[20] They designed their method so that if the molecular density were equal to the promolecular density, the AIM charges would

match the pro-atom charges. Inspired by this work, in 1977 Hirshfeld proposed a fuzzy real-space partitioning, with the AIM densities,

$$\rho_A(\mathbf{r}) = \rho_A^0(\mathbf{r}) + \left(\frac{\rho_A^0(\mathbf{r})}{\rho_{\text{mol}}^0(\mathbf{r})} \right) \left(\rho_{\text{mol}}(\mathbf{r}) - \rho_{\text{mol}}^0(\mathbf{r}) \right). \tag{1.7}$$

He proposed this definition by analogy to the way that the profit (loss) is shared between the stockholders of a corporation. I.e., if the molecule gains (loses) electron density at point **r** relative to the promolecule, then the AIM gain (lose) molecular electron density in proportion to their contribution to the promolecular density at point **r**.[21]

In 2000, Nalewajski and Parr derived the information-theoretic AIM by minimizing the Kullback-Leibler divergence between the AIM and isolated neutral pro-atoms,

$$\underbrace{\min}_{\left\{ \{\rho_A(\mathbf{r})\} \middle| \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r}) = \rho_{\text{mol}}(\mathbf{r}) \right\}} \sum_{A=1}^{N_{\text{atoms}}} \int \rho_A(\mathbf{r}) \ln \left(\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} \right) d\mathbf{r}\,, \tag{1.8}$$

leading to the AIM density

$$\rho_A(\mathbf{r}) = \left(\frac{\rho_A^0(\mathbf{r})}{\rho_{\text{mol}}^0(\mathbf{r})} \right) \rho_{\text{mol}}(\mathbf{r}). \tag{1.9}$$

This is equivalent to the Hirshfeld definition, Eq. (1.7).[25-26] A detailed derivation of this foundational result is provided in Appendix 7.1. Nalewajski and Parr's information-theoretic approach provides a general theoretical framework for developing new partitioning methods. These methods differ in their choice of (1) dissimilarity measure, (2) reference pro-atoms, and (3) imposed constraints.

Hirshfeld AIM density and charges are well-defined and unique, but the choice of

neutral reference pro-atoms is arbitrary and results in very small charges for Hirshfeld atoms. In addition, as elaborated in Appendix 7.2, Nalewajski and Parr's formulation does not fully comply with the spirit of information theory. To fix these shortcomings, various Hirshfeld-inspired partitioning methods have been developed. In the following sections, we will briefly discuss some of the corresponding algorithms, and the relationships between them. Several related, but more-or-less *ad hoc*, partitioning approaches do not perfectly fit into this information-theoretic framework, so they are not elaborated in detail. Examples of this sort of method include the ever-expanding family of Density Derived Electrostatic and Chemical (DDEC) methods[31-32, 58] and Charge Model 5 (CM5) (developed by mapping Hirshfeld charges onto a new set of charges providing a more accurate monopole approximation of electrostatic potential).[59]

1.3.2 Hirshfeld-I and Hirshfeld-Iλ

The first, and most prevalent, variant of Hirshfeld methods is the iterative Hirshfeld (Hirshfeld-I or HI) partitioning method of Bultinck *et al*..[60] This method lifts the arbitrary selection of neutral pro-atoms criterion, and refines the pro-atoms self-consistently so that, at convergence, the pro-atoms and the AIM have the same population. To perform Hirshfeld-I partitioning,

1. Initialization (n = 0). Obtain initial AIMs using the Hirshfeld partitioning, (1.9), with isolated neutral atom densities as pro-atoms. In practice, the Hirshfeld-I method is insensitive to any reasonable initial guess for the pro-atom populations.
2. Iteration. Until the atomic populations converge (e.g., until $\max_{A\in\text{atoms}} \left(\left| N_A^n - N_A^{n-1} \right| \right) < \varepsilon$),
    a. Update each pro-atom so that it has the same population as the AIM in the previous step,

$n-1$:

$$\rho_A^0(\mathbf{r}) = \left(N_A^{n-1} - \lfloor N_A^{n-1} \rfloor\right)\rho^0_{A;\lceil N_A^{n-1}\rceil}(\mathbf{r}) + \left(\lceil N_A^{n-1}\rceil - N_A^{n-1}\right)\rho^0_{A;\lfloor N_A^{n-1}\rfloor}(\mathbf{r}) \quad (1.10)$$

b. Repeat the Hirshfeld partitioning, (1.9), with the updated pro-atom densities.

Step 2a uses the fact that the only size-consistent way, and therefore the only chemically-sensible way, to define a density with a noninteger number of electrons, N, is to define it as a weighted average of the same system's densities with the next-lower-integer number, $\lfloor N \rfloor$, and the next-higher-integer number, $\lceil N \rceil$, of electrons.[61-64] E.g., the density of a system with 9.4 electrons is a linear combination of the 9-electron and 10-electron systems with the same external potential,

$$\rho_{N=9.4}(\mathbf{r}) = 0.6\rho_{N=9}(\mathbf{r}) + 0.4\rho_{N=10}(\mathbf{r}) \quad (1.11)$$

This elegant refinement results in a set of chemically intuitive reference pro-atoms and produces quality AIM charges. In addition, it provides a simpler information-theoretic interpretation because, at convergence, the AIM and the corresponding pro-atom densities have the same normalization. Numerous studies testify that Hirshfeld-I fulfills many of the desired features, including insensitivity to level of theory,[46] accurate electrostatic potential approximation,[65] and applicability to solids.[66-68] Hirshfeld-I has been extended to Fractional Occupation Hirshfeld-I (FOHI)[69] to calculate atomic spin population and FOHI-D[70] to calculate the atomic charge and atomic dipole self-consistently.

However, the Hirshfeld-I method is not perfect—it is not variational and it gives erratic results for negatively-charged nitrogen atoms and atoms in high oxidation states.

The main issue is that Hirshfeld-I very often requires unbound pro-atom densities; i.e. (di)anions in which the extra electron(s) are not bound to the isolated atom, resulting in AIM densities that extend too far from the nucleus.[71] The first attempt to partially solve these issues was the Hirshfeld-Iλ method, where the Hirshfeld-I form of the pro-atom density of N_A^0-electron system is stipulated,

$$\rho_A^0(\mathbf{r}) = \left(N_A^0 - \lfloor N_A^0 \rfloor\right)\rho^0_{A;\lceil N_A^0 \rceil}(\mathbf{r}) + \left(\lceil N_A^0 \rceil - N_A^0\right)\rho^0_{A;\lfloor N_A^0 \rfloor}(\mathbf{r}), \qquad (1.12)$$

but the atomic density and the pro-atom populations are both selected by minimizing the information loss under the constraint that AIM and pro-atoms have the same population,

$$\underbrace{\min}_{\left\{\begin{array}{c|c} \{\rho_A(\mathbf{r})\} & \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r}) = \rho_{\text{mol}}(\mathbf{r}) \\ \{N_A^0\} & \{N_A^0 = \int \rho_A(\mathbf{r}) d\mathbf{r} = N_A\} \end{array}\right\}} \sum_{A=1}^{N_{\text{atoms}}} \int \rho_A(\mathbf{r}) \ln\left(\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r}, N_A^0)}\right) d\mathbf{r} \qquad (1.13)$$

Theoretically, Hirshfeld-Iλ selects its pro-atoms variationally and adheres to the information-theoretic spirit. (Note, however, that Eq. (1.9) does not hold in Hirshfeld-Iλ.) Unfortunately, Hirshfeld-Iλ is numerically challenging to converge because the objective function is discontinuous at integer pro-atomic populations, it is sensitive to the choice of basis set, and the charges from Hirshfeld-Iλ fail to accurately reproduce the molecular electrostatic potential.[72]

1.3.3 Variational Hirshfeld-I

Lifting the constraint that the pro-atomic and atomic populations be equal, but guaranteeing a correctly normalized promolecule, gives a method we refer to as variational Hirshfeld-I,

$$\underbrace{\min}_{\left\{\{\rho_A(\mathbf{r})\},\{N_A^0\}\;\middle|\;\sum_{A=1}^{N_{\text{atoms}}}\rho_A(\mathbf{r})=\rho_{\text{mol}}(\mathbf{r}),\;\sum_{A=1}^{N_{\text{atoms}}}N_A^0=\sum_{A=1}^{N_{\text{atoms}}}N_A=N_{\text{mol}}\right\}} \sum_{A=1}^{N_{\text{atoms}}}\int\rho_A(\mathbf{r})\ln\left(\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r},N_A^0)}\right)d\mathbf{r} \qquad (1.14)$$

Variational Hirshfeld-I seems to give results that are closer to Hirshfeld-I than Hirshfeld-Iλ, but the optimization is challenging because of the derivative discontinuity in the objective function. In addition, the final pro-atoms do not have the same charges as the AIM, making the information-theoretic argument less elegant.

1.3.4 Extended Hirshfeld (Hirshfeld-E)

The poor performance of the Hirshfeld and Hirshfeld-I partitionings for molecules containing atoms in high oxidation states motivated the Hirshfeld-E method.[73] Hirshfeld-E is based on the decomposition of the *N*-electron density as a sum of the Fukui functions[64, 74-76]

$$f_{A;N}(\mathbf{r}) = \rho_{A;N}^0(\mathbf{r}) - \rho_{A;N-1}^0(\mathbf{r}) \qquad (1.15)$$

$$\rho_{A;N}^0(\mathbf{r}) = \sum_{k=1}^{N} f_{A;k}(\mathbf{r}) \qquad (1.16)$$

This motivates the idea of using the spherically-averaged atomic Fukui functions of the isolated atom as a basis set for expanding the pro-atomic density,

$$\rho_A^0(\mathbf{r}) = \sum_{k=1}^{N_{\text{bound}}} c_k f_{A;k}(\mathbf{r}) \qquad (1.17)$$

where the sum runs over the Fukui functions of all *bound* electronic states. In Hirshfeld-E, step 2a in the Hirshfeld-I algorithm is replaced by:

2a'. Update each pro-atom by performing a least-squares fit of Eq. (1.17) to the atomic density, subject to the constraint that the pro-atom and atom have the same population and that the coefficients of expansion are nonnegative,

$$\underbrace{\min}_{\left\{\{c_k\}\middle|\begin{subarray}{l}c_k\geq 0\\ N_A=\int\rho_A(\mathbf{r})d\mathbf{r}=\sum_{k=1}^{N_{\text{bound}}}c_k\end{subarray}\right\}}\int\left(\rho_A(\mathbf{r})-\sum_{k=1}^{N_{\text{bound}}}c_k f_{A;k}(\mathbf{r})\right)^2 d\mathbf{r} \tag{1.18}$$

The Hirshfeld-E method is more robust for highly-charged atomic sites, but like Hirshfeld-Iλ, the charges from Hirshfeld-E vary erratically with basis set.[73]

1.3.5 Iterative Stockholder Analysis (ISA)

Conceptually, the least appealing aspect of Hirshfeld-I-like partitioning methods is the need to specify a functional form for the pro-atom density. Lillestolen and Wheatley proposed an alternative approach, called iterative stockholder analysis (ISA) to remedy this feature.[77-78] The main idea in ISA is that the pro-atom density is the spherical average of the atomic density,

$$\rho_A^0\left(r=|\mathbf{r}-\mathbf{R}_A|\right)=\int_0^{2\pi}\int_0^{\pi}\rho_A\left(\mathbf{r}-\mathbf{R}_A\right)\sin\theta d\theta d\phi \tag{1.19}$$

By doing this, the need for a set of spherically-averaged reference pro-atom densities is avoided; the reference atom is built as a set of density values on a radial grid and updates in every iteration according to Eq. (1.19). The ISA method is calculated by

1. Initialization. n = 0. Obtain initial AIMs using the Hirshfeld partitioning, (1.9), with isolated neutral pro-atom densities, or spherically symmetric pro-atom densities using, for example, Eq. (1.20) or (1.21).

2. Iteration. Until the atomic populations converge, $\underbrace{\max}_{A\in\text{atoms}}\left(\left|N_A^n - N_A^{n-1}\right|\right) < \varepsilon$,

   a. Perform the Hirshfeld partitioning, (1.9), with the specified pro-atoms.

   b. Update each pro-atom by setting it equal to the spherical average of the atomic density, Eq. (1.19).

Since ISA is a variational method with a unique minimum for pro-atom density; the final partitioning does not depend on how one initializes the pro-atoms.[79] Recognizing this, and desiring to avoid the need for reference atomic densities in the initialization step, Lillestolen and Wheatley suggested using the very simple choice,

$$\rho_A^0(r) = \frac{1}{N_{\text{atoms}}}\int_0^{2\pi}\int_0^{\pi}\rho_{\text{mol}}\left(\mathbf{r}-\mathbf{R}_A\right)\sin\theta\, d\theta\, d\phi\,. \tag{1.20}$$

With this initialization, ISA converges quite slowly, typically requiring several times more iterations than Hirshfeld-I. To try to improve matters, we replaced the average density on each spherical shell with the shell's minimum density,

$$\rho_A^0(r) = \underbrace{\min}_{\theta,\phi}\rho_{\text{mol}}\left(\left|\mathbf{r}-\mathbf{R}_A\right|,\theta,\phi\right) \tag{1.21}$$

While our alternative initialization converges slightly faster, the improvement is not dramatic. In cases where a database of atomic densities is available, it is best to initialize the ISA algorithm with spherically-averaged pro-atom densities or the AIM density of Hirshfeld or Hirshfeld-I methods.

At a chemical level, ISA is not robust when a central atom is surrounded by a spherical shell of other atoms. In this case, the AIM density of the central atom tends to have a blip at the location of the next shell(s) of atoms. This causes the central atom to have too large (often wildly too large) population and leads to an atomic density that is

nonmonotonic, violating the "sensibility" requirement.[79] ISA is also not conformationally stable: a small breaking of the molecular symmetry (so atoms surrounding the central atom lie on the surface of an ellipse, rather than the surface of a sphere) causes the density and population of the central atom to become sensible again.[80]

To overcome these results, one can "hammer down" the spurious blips in the ISA atomic densities, forcing the atomic densities to be monotonic by requiring that the $(n+1)^{st}$ point on a radial grid emanating from the atomic nucleus be no larger than n^{th} point. Mathematically, this replaces Eq. (1.19) with

$$\rho_A^0(r) = \underbrace{\min}_{|\mathbf{r}-\mathbf{R}_A|\le r} \left(\int_0^{2\pi} \int_0^{\pi} \rho_A(\mathbf{r}-\mathbf{R}_A) \sin\theta \, d\theta \, d\phi \right) \tag{1.22}$$

A more numerically efficient approach is to approximate the pro-atomic density from (1.19) as a linear combination of monotonically decreasing basis functions like *s*-type Gaussians,[80]

$$\rho_A^0(r) \approx \sum_k c_k \exp\left(-\alpha_k r^2\right) \tag{1.23}$$

The coefficients can be determined by least-squares fitting subject to the constraint $c_k \ge 0$. This Gaussian-ISA (GISA) method has sensibly monotonic atomic densities, but it still tends to exaggerate the population of atoms at the center of a spherical shell of other atoms. The pro-atom density, instead of descending to a small value and then rising again on the surface of a spherical shell, descends slowly until the spherical shell of atoms is encountered, and descends rapidly thereafter. Because of these slowly-descending atomic densities, the atomic charges are not very transferable and conformational stability is low. The results are also quite sensitive to the choice of the Gaussian expansion functions: for

short expansions, results often seem satisfactory, but in the basis-set limit one recovers the problematic "hammered down" version of ISA in Eq. (1.22).[80]

1.3.6 Density Derived Electrostatic and Chemical (DDEC) Partitioning

ISA variants define AIM that are "as spherical as possible" by some criterion, and therefore the AIM of ISA-related methods have small dipole and higher-order-multipole moments. This ensures that the charges from ISA-type methods provide an excellent description of the molecule's electrostatic potential. One would like to somehow combine this feature of ISA with the favorable transferability and conformational stability features of Hirshfeld-based methods that employ explicit atomic reference densities. This has been attempted in DDEC methods by defining atomic weight function as a weighted-geometric average of the two methods' pro-atomic densities, specifically,[31-32, 58, 81]

$$w_A(\mathbf{r}) = \left(\rho_A^{0;\mathrm{ISA}}(r)\right)^{1-\chi} \left(\rho_A^{0;\mathrm{Hirshfeld\text{-}I}}(\mathbf{r})\right)^{\chi} \tag{1.24}$$

Unfortunately, this method does not fully remedy the problems of conformational stability and spherical-shell bias associated with ISA. Subsequent refinements of the DDEC family of methods have removed these problems, but these refinements involve a complicated hand-tuning of the method and intuitive but nonphysical revisions to the pro-atom densities (especially for anions) to meet certain performance goals. The resulting DDEC methods are inelegant by the criteria of "capable of being compactly described with words alone" and do not satisfy some of the mathematical requirements we believe it is desirable for information-based partitioning methods to possess (e.g., the recovery of the non-interacting limit of AIM and the satisfaction of the distributive property of

dissimilarity measure). This does not diminish the practical utility of the latest DDEC methods for applications, especially in the solid state, where they have been more thoroughly tested than any of the other approaches we discuss.

1.3.7 Minimal Basis Iterative Stockholder (MBIS) Partitioning

The most recent variant proposed by Verstraelen, *et al*. called Minimal Basis Iterative Stockholder (MBIS) method takes a new strategy for modeling pro-atoms.[82] It expands each pro-atom as a weighted sum of normalized *s*-type Slater functions, $\left\{ s_{A,i}\left(\mathbf{r},\sigma_{A,i}\right) \middle| \int s_{A,i}\left(\mathbf{r},\sigma_{A,i}\right) d\mathbf{r} = 1 \right\}_{i=1}^{m_A}$, centered on the atoms,

$$\rho_A^0\left(r=\left|\mathbf{r}-\mathbf{R}_A\right|\right) = \sum_{i=1}^{m_A} N_{A,i} s_{A,i}\left(r,\sigma_{A,i}\right) = \sum_{i=1}^{m_A} \left(\frac{N_{A,i}}{8\pi\sigma_{A,i}^3} \right) e^{\left(-\frac{\left|\mathbf{r}-\mathbf{R}_A\right|}{\sigma_{A,i}}\right)} \qquad (1.25)$$

where $\sigma_{A,i}$ and $N_{A,i}$ denote the width and weight of the i^{th} *s*-type normalized Slater function on atom *A*. The number of Slater functions m_A is dictated by the number of electron shells in an atom with atomic number Z_A, in this regard, the weights $\left\{N_{A,i}\right\}_{i=1}^{m_A}$ can be perceived as shell populations. (One must limit the number of Slater functions because as the number of Slater functions increases, this method approaches "hammered-down" ISA.) The shell widths and populations in MBIS are optimized alongside the AIM densities by constrained minimization of the Kullback-Leibler information loss,

$$\underbrace{\min}_{\left\{ \begin{array}{c} \rho_A(\mathbf{r}) \\ \left\{N_{A,i}\right\}_{i=1}^{m_A},\left\{\sigma_{A,i}\right\}_{i=1}^{m_A} \end{array} \middle| \begin{array}{c} \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r}) = \rho_{\text{mol}}(\mathbf{r}) \\ \left\{ N_A^0 = \sum_{i=1}^{m_A} N_{A,i} = \int \rho_A(\mathbf{r}) d\mathbf{r} = N_A \right\} \end{array} \right\}} \sum_{A=1}^{N_{\text{atoms}}} \int \rho_A\left(\mathbf{r}\right) \ln\left(\frac{\rho_A\left(\mathbf{r}\right)}{\rho_A^0\left(\mathbf{r},\left\{N_{A,i}\right\},\left\{\sigma_{A,i}\right\}\right)} \right) d\mathbf{r} \qquad (1.26)$$

The constraint requiring the pro-atom and AIM to have the same population makes MBIS appealing from the information-theoretic point of view, but couples the minimization of AIM and pro-atoms. However, unlike Hirshfeld-Iλ, the minimization simplifies to the stockholder formula because of the special form of the pro-atoms, and the shell population and shell width of each atom is given by

$$\begin{aligned} N_{A,i} &= \int \frac{N_{A,i} s_{A,i}\left(r,\sigma_{A,i}\right)}{\rho^0_{mol}\left(\mathbf{r},\left\{N_{A,i}\right\},\left\{\sigma_{A,i}\right\}\right)} \rho_{mol}(\mathbf{r})\, d\mathbf{r} \\ \sigma_{A,i} &= \frac{1}{3N_{A,i}} \int \frac{N_{A,i} s_{A,i}\left(r,\sigma_{A,i}\right)}{\rho^0_{mol}\left(\mathbf{r},\left\{N_{A,i}\right\},\left\{\sigma_{A,i}\right\}\right)} \left|\mathbf{r}-\mathbf{R}_A\right| \rho_{mol}(\mathbf{r})\, d\mathbf{r} \end{aligned} \tag{1.27}$$

The identities obtained for pro-atom parameters are specific to the Kullback-Leibler divergence measure, and are used for optimizing the pro-atoms self-consistently.

1. Initialization. n = 0. Obtain initial AIMs using the Hirshfeld partitioning, (1.9), with pro-atom densities modeled in Eq. (1.25). The initial value of the $N_{A,i}$ parameter is set to the number of electrons in the i^{th} shell of atom A. The initial value of $\sigma_{A,i}$ is set to $\frac{a_0}{2Z_A}$ and $\frac{a_0}{2}$ for the innermost and outermost shell of atom A, respectively, where for the intermediate shells, it is assigned by geometric interpolation $\sigma_{A,i} = \frac{a_0}{2Z_A^{\left(1-\frac{i-1}{m_A-1}\right)}}$.
2. Iteration. Until shell parameters converge: $\underbrace{\max}_{\substack{A\in\text{atoms} \\ i\in[1,\cdots,m_A]}} \left(\left|N^n_{A,i} - N^{n-1}_{A,i}\right|\right) < \varepsilon$ and $\underbrace{\max}_{\substack{A\in\text{atoms} \\ i\in[1,\cdots,m_A]}} \left(\left|\sigma^n_{A,i} - \sigma^{n-1}_{A,i}\right|\right) < \varepsilon$
    a. Perform the Hirshfeld partitioning, Eq. (1.9), with the specified pro-atoms, Eq. (1.25)
    b. Update each pro-atom by computing its shell widths $\left\{\sigma_{A,i}\right\}_{i=1}^{m_A}$ and shell populations $\left\{N_{A,i}\right\}_{i=1}^{m_A}$ parameters, Eq. (1.27).

The objective function in Eq. (1.26) is not convex, so the choice of initial guess in step 1

can affect the resulting MBIS charges. The iterative procedure to refine the pro-atoms can easily be implemented with a linear-scaling computational cost for applications to supramolecular systems. In addition, the Slater functions describing the valence electron density of AIM allow better approximations of the electrostatic interaction in force fields.[82-83]

1.4 Other Population Analysis Methods

Although the focus of this work is information-theoretic population analysis methods, we will compare the results to other approaches. For each general family of methods—orbital-based population analysis, topological partitioning, and electrostatic fitting—we have chosen one to compare to what we feel is the best widely available method of that family.

1.4.1 Orbital-Based Population Analysis

Orbital-based partitioning was pioneered by Mulliken.[6-9] Each natural molecular orbital can be expressed as a sum of atomic basis functions,

$$\psi_p(\mathbf{r}) = \sum_{A=1}^{N_{\text{atoms}}} \sum_{i \in A} c_{p;Ai} \chi_{Ai}(\mathbf{r}) \tag{1.28}$$

where $\chi_{Ai}(\mathbf{r})$ denotes the i^{th} basis function on atom A. To divide this molecular orbital into contributions from the atomic basis functions, note that

$$1 = \int \left|\psi_p(\mathbf{r})\right|^2 d\mathbf{r} = \sum_{A=1}^{N_{\text{atoms}}} \sum_{B=1}^{N_{\text{atoms}}} \sum_{i \in A} \sum_{j \in B} c^*_{p;Ai} c_{p;Bj} S_{Ai,Bj} \tag{1.29}$$

where the overlap matrix between the atomic basis functions has been defined as

$$S_{Ai,Bj} = \int \chi^*_{Ai}(\mathbf{r})\chi_{Bj}(\mathbf{r})d\mathbf{r} \tag{1.30}$$

Mulliken proposed decomposing Eq. (1.29) into the (net) populations of the atomic basis functions,

$$\nu_{p;Ai,Ai} = n_p \left|c_{p;Ai}\right|^2 S_{Ai,Ai} \tag{1.31}$$

and "bonding" populations associated with the overlaps between different atomic basis functions,

$$\nu_{p;Ai,Bj} = n_p \left(c^*_{p;Ai}c_{p;Bj}S_{Ai,Bj} + c_{p;Ai}c^*_{p;Bj}S_{Bj,Ai}\right) = 2n_p \,\mathrm{Re}\left(c^*_{p;Ai}c_{p;Bj}S_{Ai,Bj}\right) \tag{1.32}$$

Here we have denoted the occupation number of the p^{th} natural molecular orbital as n_p. To assign atomic populations, one needs to divide the bonding population between the atomic basis functions between the contributing atoms. In the absence of any other information, all one can do is divide the bonding population half-and-half between the contributors, giving the gross populations of the atomic basis functions,

$$\begin{aligned}\pi_{p;Ai} &= \nu_{p;Ai,Ai} + \frac{1}{2}\sum_{Ai\neq Bj}\left(\nu_{p;Ai,Bj} + \nu_{p;Bj,Ai}\right) \\ &= \sum_{B=1}^{N_{\text{atoms}}}\sum_{j\in B}\nu_{p;Ai,Bj}\end{aligned} \tag{1.33}$$

Summing this contribution over all the molecular orbitals and all the basis functions assigned to a given atom gives that atom's population,

$$n_A = \sum_{p=1}^{N_{\text{orbitals}}}\sum_{i\in A}\pi_{p;Ai} \tag{1.34}$$

This can be recast in matrix language by defining the molecular density matrix as

$$P_{Ai,Bj} = \sum_{p=1}^{N_{\text{orbitals}}} n_p C^*_{p;Ai} C_{p;Bj} \tag{1.35}$$

and then the charge-bond matrix as

$$M_{Ai,Bj} = P_{Ai,Bj} S_{Ai,Bj} \tag{1.36}$$

Then the atomic charges are simply

$$n_A = \text{Re}\left(\sum_{i \in A} \sum_{B=1}^{N_{\text{atoms}}} \sum_{j \in B} M_{Ai,Bj} \right) \tag{1.37}$$

Similarly, the electron density of the atom in the molecule can be defined as

$$\rho_A(\mathbf{r}) = \text{Re}\left(\sum_{i \in A} \sum_{B=1}^{N_{\text{atoms}}} \sum_{j \in B} P_{Ai;Bj} \chi^*_{Ai}(\mathbf{r}) \chi_{Bj}(\mathbf{r}) \right) \tag{1.38}$$

Löwdin noted that the preceding analysis could be simplified if the atomic basis functions were orthonormal, $S_{Ai,Bj} = \delta_{AB}\delta_{ij}$, and he proposed choosing orthogonalized atomic basis functions, $\{\chi^{\text{d}}_{Ai}\}$, that were as close as possible to the original (nonorthogonal) basis.[10-12] It is not difficult to deduce that the new basis functions are simply

$$\chi^{\text{d}}_{Ai}(\mathbf{r}) = \sum_{B=1}^{N_{\text{atoms}}} \sum_{j \in B} S^{-1/2}_{Ai,Bj} \chi_{Bj}(\mathbf{r}) \tag{1.39}$$

In the orthogonalized basis, the atomic populations have the simple expression,

$$n^{\text{d}}_A = \sum_{i \in A} P^{\text{d}}_{Ai,Ai} \tag{1.40}$$

These methods for assigning atomic populations are extremely erratic for large and/or unbalanced basis sets. More generally, Mulliken/Löwdin partitioning fails in any

circumstance where the atomic basis functions are not (well-)localized on the atomic centers. Consider, for example, that in the extreme case where one expands the wavefunction with a complete set of basis functions centered on a single atom, all the electrons will be assigned to that atom. To avoid this, it is necessary to ensure that the molecular orbitals and the one-electron reduced density matrix,

$$\gamma(\mathbf{r},\mathbf{r}') = \sum_{p=1}^{N_{\text{orbitals}}} n_p \psi_p(\mathbf{r}) \psi_p^*(\mathbf{r}'), \tag{1.41}$$

are expressed in terms of atomic basis functions that have chemical meaning. The approaches proposed by Weinhold[14-15], Ruedenberg[16-18], and Knizia[19] achieve this by using atomic basis functions that resemble atomic orbitals. We will only consider Weinhold's approach, called natural population analysis (NPA), because it is by far the most widely used of these approaches.

In NPA, one uses the atomic orbitals of the isolated neutral atoms, $\{\phi_{Ai}\}$ as basis functions. The molecular orbitals or, alternatively, the one-electron reduced density matrix can be expanded in this new basis,

$$\omega_{Ai,Bj} = \iint \phi_{Ai}^*(\mathbf{r}) \gamma(\mathbf{r},\mathbf{r}') \phi_{Bj}(\mathbf{r}') d\mathbf{r}'. \tag{1.42}$$

Because the atomic orbitals are not orthogonal, the trace of this matrix is greater than the total number of electrons. The basic idea in natural population analysis is to use an occupation-weighted version of Löwdin's symmetric orthogonalization method. That is, one chooses the orthonormal basis that resembles the existing atomic orbital basis functions to the maximum possible extent, weighted by the occupation of the atomic orbital basis functions,

$$\min \sum_{A=1}^{N_{\text{atoms}}} \sum_{i \in A} \omega_{Ai,Ai} \int \left| \phi_{Ai}(\mathbf{r}) - \phi_{Ai}^{\text{d}}(\mathbf{r}) \right|^2 d\mathbf{r} \tag{1.43}$$

with solution,

$$\phi_{Ai}^{\text{d}}(\mathbf{r}) = \sum_{B=1}^{N_{\text{atoms}}} \sum_{j \in B} \left[\mathbf{D}(\mathbf{DSD})^{-1/2} \right]_{Ai,Bj} \phi_{Bj}(\mathbf{r}) \tag{1.44}$$

where **D** is a diagonal matrix with the entries $d_{Ai,Bj} = \omega_{Ai,Ai}\delta_{Ai,Bj}$ and **S** is the overlap matrix between the atomic orbitals. The natural populations can then be determined, using Eq. (1.40), or, specifically,

$$n_A = \sum_{i \in A} \iint \left(\phi_{Ai}^{\text{d}}(\mathbf{r}) \right)^* \gamma(\mathbf{r}, \mathbf{r}') \phi_{Ai}^{\text{d}}(\mathbf{r}') d\mathbf{r} d\mathbf{r}' \tag{1.45}$$

While this captures the *essence* of the natural population analysis method, the actual approach is significantly more complicated. For example the valence atomic orbitals and the Rydberg atomic orbitals are orthogonalized separately, and special care must be taken when treating atomic orbitals with very small occupations.[14]

In our experience, NPA and other atomic-orbital-based population analysis methods remove most, but not all, of the basis-set sensitivity of the unrefined Mulliken/Löwdin partitioning strategies. However these methods are still sensitive to the quality of the molecular basis set (e.g., they typically work poorly in the extreme case where all the basis functions are located on a single atomic center) and also on the design decisions that were taken in constructing the atomic basis set.

1.4.2 Topological Partitioning

Topological partitioning methods divide space into regions, each of which is then

associated to an atom. For example, in Voronoi partitioning, one assigns each point in space to the closest atomic nucleus, giving atomic regions,

$$\Omega_A^V \equiv \left\{ \mathbf{r} \middle| \forall B \neq A,\ \left|\mathbf{r}-\mathbf{R}_A\right| < \left|\mathbf{r}-\mathbf{R}_B\right| \right\} \tag{1.46}$$

The atomic weight functions are then taken to be the characteristic function of the region,

$$w_A(\mathbf{r}) = \begin{cases} 1 & \mathbf{r} \in \Omega_A \\ 0 & \mathbf{r} \notin \Omega_A \end{cases} \tag{1.47}$$

and so the atomic densities and atomic populations are defined as

$$\rho_A(\mathbf{r}) = w_A(\mathbf{r})\rho_{\text{mol}}(\mathbf{r}) \tag{1.48}$$

and

$$N_A = \int \rho_A(\mathbf{r})\, d\mathbf{r} = \int_{\Omega_A} \rho_{\text{mol}}(\mathbf{r})\, d\mathbf{r} \tag{1.49}$$

The problem with Voronoi-based partitioning is that it depends only on the location of the nuclei, and not on the molecule's electronic structure. Richard Bader realized that there was a natural way to separate a molecule into atoms using the topography of the electron density.[22-23, 84] If one visualizes the molecular density, it looks like a mountain range, with peaks coinciding with the locations of the atomic nuclei and valleys between them. Bader's partititioning, called the quantum theory of atoms in molecules (QTAIM), corresponds to a watershed analysis of the atoms. Imagine that a tiny rain-cloud hovered over the location of an atomic nucleus, and the water from the cloud flowed down the sides of the mountain of electron density associated with the atom. All the points in space that were wet by the rain would be assigned to the atom.

Mathematically, this means that if one takes a point, **r**, and starts to make a steepest-ascent path from that point,

$$\begin{aligned} \mathbf{r}_0 &= \mathbf{r} \\ \mathbf{r}_1 &= \mathbf{r}_0 + \varepsilon \nabla \rho\left(\mathbf{r}_0\right) \\ &\vdots \\ \mathbf{r}_{n+1} &= \mathbf{r}_n + \varepsilon \nabla \rho\left(\mathbf{r}_n\right) \\ &\vdots \end{aligned} \tag{1.50}$$

The point is then assigned to the nucleus at the end of the steepest-ascent path. The boundaries between the atoms correspond to zero-flux surfaces, where the normal to the surface of the atomic volumes is orthogonal to the gradient,

$$\mathbf{r} \in \partial\Omega_A \leftrightarrow \nabla\rho\left(\mathbf{r}\right) \cdot \mathbf{n}_{\Omega_A} = 0 \tag{1.51}$$

Equation (1.51), called local zero-flux condition, establishes that the integral of the Laplacian of the electron density over an atomic region is zero,

$$\int_{\Omega_A} \nabla^2 \rho\left(\mathbf{r}\right) d\mathbf{r} = 0\,. \tag{1.52}$$

Equation (1.52), called net zero-flux condition, in turn, makes it possible to define atomic kinetic energies (and then, by the virial theorem, atomic energies) for this atomic partitioning without excessive sensitivity to the way one chooses to represent the quantum mechanical operators for the atoms.[23, 85-88] This ability to define quantum mechanical operators for atoms is why this approach is usually called the quantum theory of atoms in molecules (QTAIM).[22, 30, 89-90] One should mention that subsequent work makes it clear that regions called quantum divided basins, which satisfy the global zero-flux condition (Eq. (1.52)), but not the local zero-flux condition (Eq. (1.51)), can also be used to define

open quantum subsystems.[44-45] Also, subsequent work has shown that there are mathematically allowable (but arguably chemical unreasonable) definitions for the local kinetic energy for which Eq. (1.52) is insufficient to define unique atomic kinetic energies.[88, 91-92] For these reasons, we believe the strongest justification for QTAIM is topological, based on the intuitive partitioning of space into atomic-density regions.

The mathematical underpinnings of QTAIM are elegant, but it has a few undesirable properties. For example, sometimes there are "extra" atoms associated with maxima in the atomic density that do not coincide with the location of an atomic nucleus.[93-95] Additional non-atomic regions also appear if one partitions the electron density obtained for a pseudopotential calculation, because the absence of the electron density from the atomic cores ruins the mountain-peak structure that the electron density has in all-electron calculations. In pseudopotential calculations, it is advisable to correct the computed electron density by adding back the (approximate) electron density from the atomic core electrons before performing the QTAIM partitioning. If one does not make this correction, then spurious atoms and significant topological complexity are induced by the "volcanic craters" associated with the missing core electrons. Conversely, sometimes there are missing atoms, because a light electron-poor atom is embedded in the electron cloud of a heavier electron-rich atom, so that there is no maximum in the molecular electron density at the location of the light atom's nucleus.

All topological partitioning methods have the problem that the atomic regions have "pointy boundaries" where three or more atomic regions meet. Because the atomic regions and their associated densities are far from spherical, it is hard to represent the

electrostatic potential of the atom with a point charge: one typically needs not only atomic charges, but very high-order atomic multipoles, to describe the molecular electrostatic potential using topological partitioning methods.[96-99]

1.4.3 Electrostatic Potential Fitting

Electrostatic potential fitting is an approach that leads to atomic populations/charges, but not to atomic density distributions.[100-105] Specifically, one tries to find atomic populations that fit the electrostatic potential of the electron density,

$$\int \frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}' \approx \sum_{A=1}^{N_{\text{atoms}}} \frac{n_A}{|\mathbf{r}-\mathbf{R}_A|} \tag{1.53}$$

subject to the constraint that the atomic populations sum up to the total number of electrons,

$$\sum_{A=1}^{N_{\text{atoms}}} n_A = N_{\text{mol}}. \tag{1.54}$$

It is clearly impossible satisfy Eq. (1.53) at every point, but for molecular mechanics force fields it is primarily important that the electrostatic potential be accurately captured at locations in the vicinity of the van der Waals surface and up to three or four van der Waals radii away from the molecule. This suggests that one choose atomic populations by minimizing[102-103, 106]

$$\underbrace{\min}_{\left\{ n_A \middle| \sum_{A=1}^{N_{\text{atoms}}} n_A = N_{\text{mol}} \right\}} \int \omega(\mathbf{r}) \left(\int \frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}' - \sum_{A=1}^{N_{\text{atoms}}} \frac{n_A}{|\mathbf{r}-\mathbf{R}_A|} \right)^2 d\mathbf{r} \tag{1.55}$$

where $\omega(\mathbf{r}) \geq 0$ is a nonnegative weight function that focuses the optimization on the region of chemical interest and which decays rapidly enough to ensure the existence of the integral that defines the objective function.

Different methods for electrostatic potential fitting mainly differ based on the weighting function one uses in Eq. (1.55) and the possible addition of constraints based on intuition about the likely size of atomic charges, equivalence of the populations of chemically similar atoms, etc.. Recognizing that the atomic populations/charges from potential-fitting would behave erratically in response to conformation changes unless the weighting function was perfectly smooth, Hu, Lu, and Yang proposed the objective function[106]

$$w(\mathbf{r}) = \exp\left(-\sigma\left(\ln \rho_{\text{mol}}^{0}(\mathbf{r}) - \ln \rho_{\text{ref}}\right)^{2}\right) \tag{1.56}$$

where $\rho_{\text{mol}}^{0}(\mathbf{r})$ is the promolecular density and the recommended values for the parameters that control the width and location of the weighting function are

$$\sigma = 0.8 \tag{1.57}$$

and

$$\ln \rho_{\text{ref}} = -9, \tag{1.58}$$

respectively. The objective function in Eq. (1.55) with the weighting function (1.56) determines the Hu-Lu-Yang (HLY) populations.[106]

Obviously this method can be extended to dipole and higher-order multipole moments, simply by inserting the appropriate multipole expansion,

$$\int \frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}' \approx \sum_{A=1}^{N_{\text{atoms}}} \frac{n_A}{|\mathbf{r}-\mathbf{R}_A|} + \frac{\mathbf{d}_A \cdot (\mathbf{r}-\mathbf{R}_A)}{|\mathbf{r}-\mathbf{R}_A|^3} + \frac{1}{2} \frac{(\mathbf{r}-\mathbf{R}_A)^T \mathbf{Q}_A (\mathbf{r}-\mathbf{R}_A)}{|\mathbf{r}-\mathbf{R}_A|^5} + \cdots \tag{1.59}$$

in Eq. (1.55). However, electrostatic fitting does not define an atomic density, and is therefore not a true partitioning method. It is possible, however, to reverse-engineer atomic densities that have the correct multipoles and which maximally resemble some atomic reference densities. E.g., one can find the electron densities that satisfy:

$$\underbrace{\min}_{\left\{\rho_A(\mathbf{r}) \middle| \begin{array}{l} \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r}) = \rho_{\text{mol}}(\mathbf{r}) \\ \int \rho_A(\mathbf{r}) d\mathbf{r} = n_A \\ \int \rho_A(\mathbf{r})(\mathbf{r}-\mathbf{R}_A) d\mathbf{r} = \mathbf{d}_A \\ \vdots \end{array}\right\}} \sum_{A=1}^{N_{\text{atoms}}} \int \rho_A(\mathbf{r}) \ln\left(\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})}\right) d\mathbf{r} \tag{1.60}$$

where the pro-atom density should be chosen to have the same atomic population as was assigned by electrostatic fitting, $N_A^0 = n_A$, using Eq. (1.12). If one has only monopoles, then this approach is equivalent to Hirshfeld-Iλ, but without the variational optimization over pro-atom populations.

Unfortunately, the optimization in Eq. (1.55) is numerically ill-conditioned. For example, the objective function in Eq. (1.55) is extremely insensitive to any atomic population that is very far from the molecular van der Waals region that is sampled by the weight function $\omega(\mathbf{r})$; the optimization landscape is therefore extremely flat, and the atomic populations can change significantly due to small changes in electron density, whether due to computational parameters (e.g., electronic structure method or basis set) or geometric changes.

1.5 Shortcomings of Current Schemes: Numerical Assessment

1.5.1 Computational Setup

We compare the performance of various information-theoretic partitioning methods (section 1.3) to more traditional approaches (section 1.4). To do so, we selected examples from our own work and from the literature that reveal specific, usually unfavorable, features of the different partitioning methods. We also selected a set of small molecules (CH_3^+, CH_4, CH_3^-, NH_4^+, NH_3, NH_2^-, H_3O^+, H_2O, OH^-) for investigating the sensitivity of these methods to the one-electron basis set and the type of electronic structure theory method used, and also for assessing how well different partitioning methods recapture chemical trends.

All quantum chemistry calculations were performed by *Gaussian09* (version C.01) software[107] employing ultrafine integration grids and stable=opt keyword to ensure that a (local) minimum of the energy with respect to variations of the orbitals was found. For the small set of molecules, the geometries were optimized at UωB97XD/cc-pVTZ level of theory, followed by single point calculations at UHF, UB3LYP,[108-110] and UωB97XD[111] levels of theory with Dunning's (d-aug-)cc-pVXZ (X=D, T, Q, 5) correlation consistent basis set series.[112-114] The NPA and HLY charges were generated by *Gaussian09*. The QTAIM charges were generated using AIMALL (version 16.01.09 standard) software. The charges from information-theoretic partitioning methods were generated using HORTON 2.0.0.[115]

1.5.2 Hirshfeld partitioning is sensitive to the choice of pro-atom

As noted before, the choice of neutral pro-atoms in the original Hirshfeld method is arbitrary, and the Hirshfeld populations from Eq. (1.9) change significantly when different reference pro-atom charges are used. This is shown in Table 1.1 for lithium chloride, where neutral and charged pro-atoms are used as references. (We have chosen the pro-atoms so that the promolecule has the same number of electrons as the molecule. Note that this is not implicit in the Hirshfeld method: the traditional Hirshfeld method uses neutral promolecules for the population analysis of molecular ions.) The middle column of Table 1.1 contains the conventional Hirshfeld charges, but considering the ionic character of LiCl, the lithium cation and chlorine anion are the chemically intuitive reference pro-atoms. This pro-atom choice produces the higher, and more chemically appealing, Hirshfeld charges in the last column. Notice, however, that no matter what choice one makes for the charges of the pro-atoms, the Hirshfeld charges are semi-insensitive to the choice of quantum chemistry method and basis set. In the remainder of this manuscript, however, whenever we refer to the Hirshfeld partitioning we will be considering only neutral pro-atoms.

Table 1.1 The Hirshfeld charge of the lithium atom in the LiCl molecule for an interatomic separation of 2.045 angstrom. A neutral promolecule has been used, so the chlorine pro-atom has a charge of $q^0_{\mathrm{Cl}} = -q^0_{\mathrm{Li}}$. The same level of theory was used to compute the molecular and pro-atomic electron densities. To show how insensitive Hirshfeld charges are to the choice of basis set, we also tabulate the absolute value of spread in Hirshfeld charges for the basis sets considered.

	Charge of the Li pro-atom, q^0_{Li}				
Level of Theory	**-1.0**	**-0.5**	**0.0**	**0.5**	**1.0**
UHF/cc-pVDZ	0.2345	0.4028	0.5511	0.7331	0.9828
UHF/cc-pVTZ	0.2196	0.3973	0.5525	0.7446	0.9831
UHF/cc-pVQZ	0.2398	0.4063	0.5519	0.7493	0.9848
UHF/cc-pV5Z	0.2428	0.4073	0.5515	0.7507	0.9853
UHF/aug-cc-pVDZ	0.2692	0.4229	0.5549	0.7551	0.9925
UHF/aug-cc-pVTZ	0.2577	0.4162	0.5523	0.7526	0.9859
UHF/aug-cc-pVQZ	0.2579	0.4152	0.5513	0.7518	0.9851
UHF/ aug-cc-pV5Z	0.2587	0.4154	0.5512	0.7517	0.9850
|max(q) - min(q)|	**0.0496**	**0.0256**	**0.0037**	**0.0221**	**0.0097**
UB3LYP/cc-pVDZ	0.1753	0.3507	0.5073	0.6937	0.9683
UB3LYP/cc-pVTZ	0.1612	0.3502	0.5169	0.7143	0.9726
UB3LYP/cc-pVQZ	0.1838	0.3617	0.5184	0.7216	0.9762
UB3LYP/cc-pV5Z	0.1853	0.3617	0.5177	0.7232	0.9763
UB3LYP/aug-cc-pVDZ	0.2100	0.3751	0.5192	0.7295	0.9884
UB3LYP/aug-cc-pVTZ	0.2048	0.3734	0.5193	0.7274	0.9789
UB3LYP/aug-cc-pVQZ	0.2052	0.3727	0.5184	0.7263	0.9775
UB3LYP/aug-cc-pV5Z	0.2053	0.3722	0.5175	0.7251	0.9759
|max(q) - min(q)|	**0.0488**	**0.0249**	**0.0120**	**0.0358**	**0.0202**
UωB97XD/cc-pVDZ	0.2005	0.3760	0.5311	0.7153	0.9769
UωB97XD/cc-pVTZ	0.1846	0.3726	0.5367	0.7326	0.9801
UωB97XD/cc-pVQZ	0.2063	0.3825	0.5364	0.7389	0.9830
UωB97XD/cc-pV5Z	0.2108	0.3841	0.5356	0.7405	0.9837
UωB97XD/aug-cc-pVDZ	0.2364	0.3987	0.5382	0.7447	0.9914
UωB97XD/aug-cc-pVTZ	0.2255	0.3936	0.5374	0.7435	0.9846
UωB97XD/aug-cc-pVQZ	0.2253	0.3919	0.5361	0.7425	0.9837
UωB97XD/aug-cc-pV5Z	0.2259	0.3917	0.5353	0.7418	0.9834
|max(q) - min(q)|	**0.0518**	**0.0261**	**0.0072**	**0.0294**	**0.0145**

1.5.3 Hirshfeld-I is Not Variational

The Hirshfeld-I method fixes the sensitivity of (ordinary) Hirshfeld charges to the choice of pro-atom and ensures that the atom and pro-atom always have the same number of electrons. Although the Hirshfeld-I procedure is not written as a variational method (cf. section 1.3.2), this does not mean that the solution of the Hirshfeld-I procedure cannot be equivalent to the minimization of the Kullback-Leibler divergence or, more generally, some other *f*-divergence measure.

Based on a recent mathematical analysis of the Hirshfeld-I equations, this does not seem to be the case.[116] There are many variational principles that are equivalent to the Hirshfeld-I equations, but they are not written as minimizations of the divergence between the densities of an AIM and a pro-atom. There it is also shown that the Hirshfeld-I solution always exists, and is never unique in a mathematical sense. However, in most (but not all)[33] cases it seems that the solution is unique in a chemical sense, as the spurious mathematical solutions correspond to cases where one or more AIM have zero electrons.

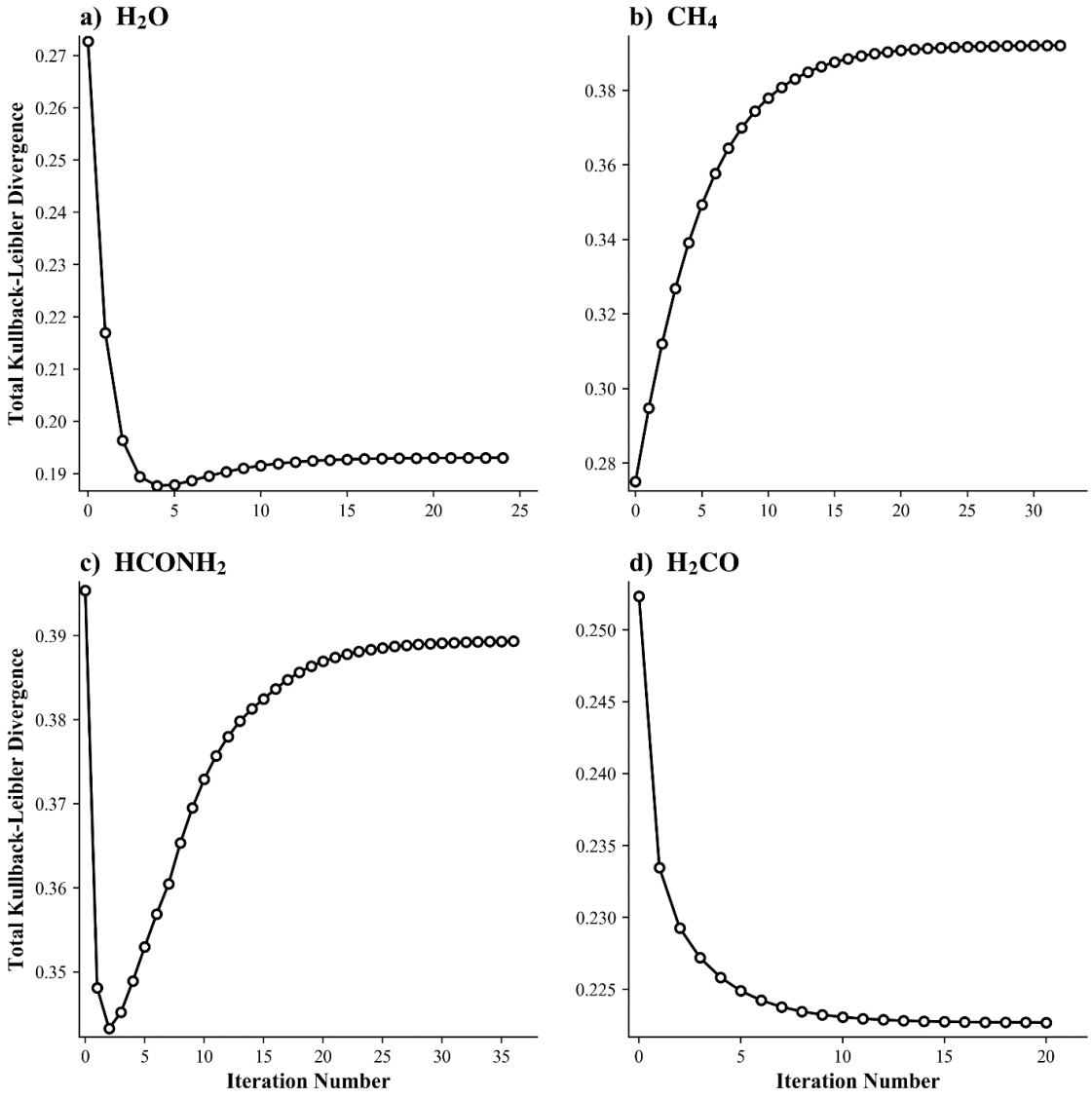
Figure 1.1 shows how the sum of the atomic Kullback-Leibler divergences,

$$\sum_{A=1}^{N_{\text{atoms}}} \int \rho_A(\mathbf{r}) \ln\left(\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})}\right) d\mathbf{r}, \tag{1.61}$$

changes when one performs the iterative refinement of Hirshfeld-I charges. This is in contrast to the iterative stockholder analysis (ISA) method, which is variational, and therefore is associated with steadily decreasing values of Eq. (1.61). We also explored

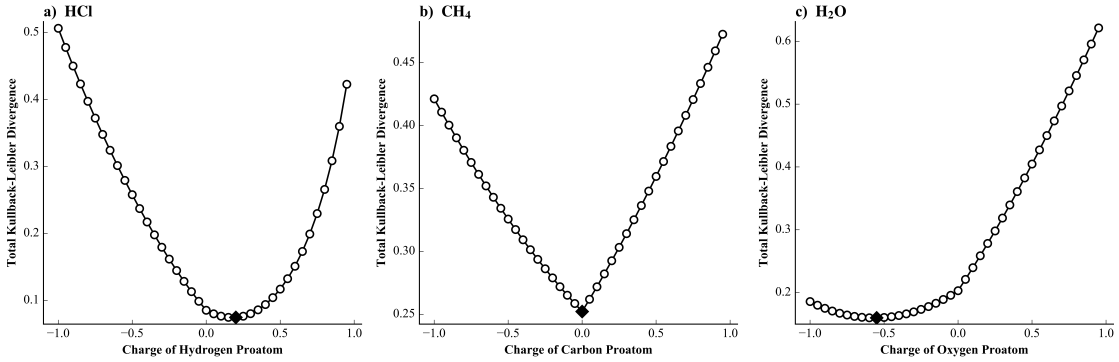
(not shown) the analogue of Eq. (1.61) for other divergences. In those cases, neither Hirshfeld-I nor ISA was variational.

Figure 1.1 The value of the total Kullback-Leibler divergence between the densities of the Hirshfeld I atoms in a molecule and the reference pro-atomic densities, Eq. (1.61), for each iteration of the self-consistent Hirshfeld-I procedure. These are computed for the optimized geometry of (a) water, (b) methane, (c) formamide, and (d) formaldehyde, at the UωB97XD/cc-pVTZ level.



There are ways to refine Hirshfeld-I to be variational. The Hirshfeld-Iλ method is one way to do this, but as mentioned in section 1.3.2, it results in inferior atomic charges[72] and also does not satisfy the distributive property in Eq. (1.3). If one relaxes the requirement that the AIM and pro-atoms have the same charge, variational minimization of Eq. (1.61) using the Hirshfeld-I definition for the pro-atom densities gives the variational Hirshfeld-I method from section 1.3.3 (cf. Eq. (1.14)). The sum of the atomic Kullback-Leibler divergences are computed, as a function of pro-atom charge, for HCl, CH_4, and H_2O in Figure 1.2. Hydrogen Chloride is the favorable case, where the minimum divergence and minimizing pro-atom charge are somewhat reasonable. In methane, the method often gets "locked" at an integer pro-atom charge (which is conceptually unappealing). The nondifferentiability of the objective function also complicates the numerical optimization. While the objective function is only convex in between two consecutive integer pro-atom charges, and so multiple solutions are possible. (For example, we observed that at certain levels of theory, LiCl can have two solutions, with Li pro-atom charges slightly more/less than +1.) In water, the objective function is insensitive to the O pro-atom charge. This makes the results from variational Hirshfeld-I sensitive to the level of theory. E.g., qualitatively insignificant changes in quantum chemistry method (e.g., changing the exchange-correlation functional and/or the basis set) can change the variational Hirshfeld-I charges significantly.

Figure 1.2 The objective function in variational Hirshfeld-I, cf. Eq. (1.13), as a function of pro-atom charges for (a) HCl with a minimum at $q_H^0 = 0.176$, (b) CH_4 with a minimum at $q_C^0 = 0.0$, and (c) H_2O with a minimum at $q_O^0 = -0.568$, at the UωB97XD/cc-pVTZ level of theory. In all cases, the minimum is marked with a black diamond.



1.5.4 Iterative Stockholder Analysis (ISA) is Sensitive to Molecular Conformation

Iterative Stockholder Analysis (ISA) is variational, but it sometimes gives chemically absurd results. Especially when there are several atoms arranged on a spherical shell around a central atom, the central atom tends to become overpopulated.[79] An extreme example of this is the endohedral fullerene where a lithium cation is placed inside a buckyball, $Li^+@C_{60}$. The lithium cation is given an enormous number of electrons, and its population is very sensitive to the size of the numerical integration grid, partly because the population of the lithium "cation" will increase dramatically when one of the radial shells of grid points (nearly) coincides with the position of the C_{60} cage. This is clearly seen in Table 1.2. It also shows that Hirshfeld-I, and the more recent versions of DDEC charges, do not suffer from the same problem. The DDEC methods, however

require very large and costly integration grids.

Table 1.2 Hirshfeld-I, ISA, DDEC3 and DDEC4 charges of lithium atom in $Li^{+}@C_{60}$ for different numerical integration grids. All calculations are performed at B3LYP/6-31G.

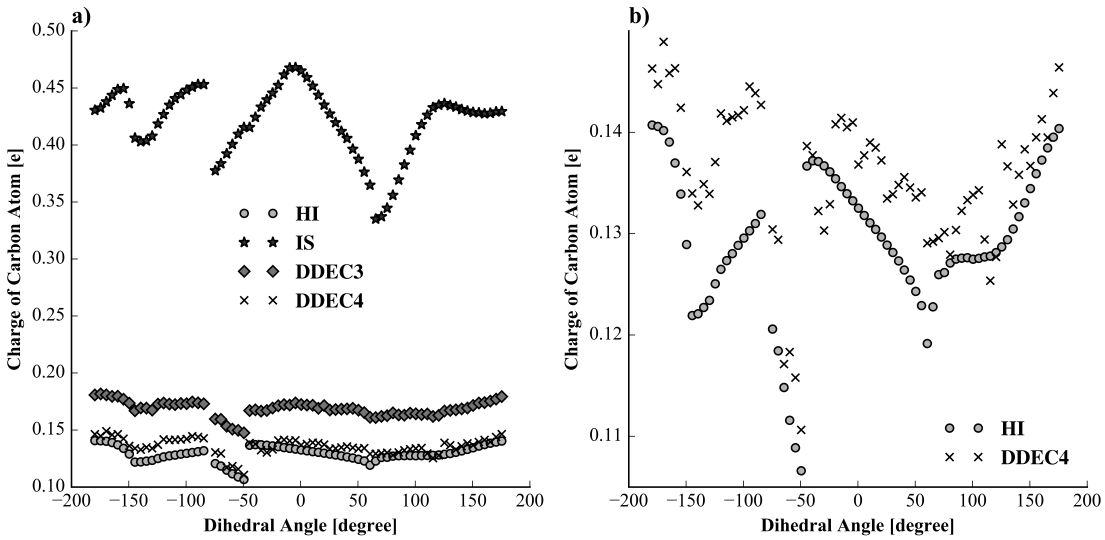
Grid	# Points	Hirshfeld-I	ISA	DDEC3	DDEC4
Coarse	93,576	0.9901	-5.0582		
Medium	152,958	0.9901	-9.0165		
Fine	320,906	0.9899	-7.8583		
Veryfine	518,558	0.9899	-7.5848		
Ultrafine	526,986	0.9899	-7.6278		
Insane	1,958,068	0.9899	-7.2861		
Custom	73,200,000			1.0025	0.9935

ISA charges are also problematic for more typical chemical problems. For example, the tendency for atoms to show a spurious decrease in charge whenever they appear at the center of a (nearly) spherical shell of neighboring atoms causes the ISA charges to be very sensitive to molecular conformational changes. This is chemically unreasonable and it is undesirable for parameterizing molecular mechanics force fields, in which the atomic charges are assumed to be insensitive with respect to molecular torsions.

To show this, Figure 1.3 shows the charge on the central carbon atom in alanine dipeptide versus rotations about the ψ dihedral angle. The ISA charges show a large and unphysical dependence on molecular conformation. Hirshfeld-I, DDEC3, and DDEC4 charges are more reasonable, with little conformation dependence. However, upon closer inspection, it is observed that the DDEC4 charges vary noisily, rather than smoothly, with

respect to the torsion. As the amount of noisiness is small, however, this is more an aesthetic and philosophical issue than a practical one.

Figure 1.3 The charge on the central carbon atom in alanine dipeptide versus the ψ dihedral angle for Hirshfeld-I, ISA, DDEC3, and DDEC4. (b) A more detailed comparison between Hirshfeld-I and DDEC4 charges obtained by zooming in.



1.5.5 Sensitivity to Basis Set and Electronic Structure Method

As mentioned in section 1.2.3, useful population methods are insensitive to changes in the basis set and the electronic structure method. To assess this, we examined the sensitivity of the most promising and popular information-theoretic methods for a set of small molecules (CH_3^+, CH_4, CH_3^-, NH_4^+, NH_3, NH_2^-, H_3O^+, H_2O, OH^-), three different electronic structure methods (UHF, UB3LYP, and UωB97XD) and twelve different basis sets (d-aug-)cc-pVXZ (X=D, T, Q, 5). The molecular anions were not bound at the Hartree-Fock level, so only the density functional theory methods were used to compute charges for CH_3^-, NH_2^-, and OH^-. Figure 1.4 compares the charges from the

most popular and promising information-theoretic methods to the charges from traditional population analysis methods based on orbital-based partitioning (represented by natural population analysis, NPA), topological partitioning (represented by the quantum theory of atoms in molecules, QTAIM), and electrostatic fitting (using the Hu-Lu-Yang procedure, HLY).

None of the methods we consider is very sensitive to the choice of electronic structure method. The traditional methods are relatively insensitive to basis set, with HLY charges being almost invariant to basis set and electronic structure method, and NPA charges being slightly more sensitive, and QTAIM charges showing the greatest dependence on basis set, especially for the molecular cations. In terms of method/basis-set stability, however, the best method by far is the conventional Hirshfeld method. The ISA charges and Hirshfeld-E charges perform well also, though we note that the Hirshfeld-E has shown problematic basis-set sensitivity for inorganic oxides.[73]

The other information-theoretic methods we considered have significantly greater basis-set dependence. The MBIS method performs well for neutral and positively charged molecular ions, but is overly sensitive to basis set for molecular anions. We speculate that this is because the limited number of Slater functions available means that the promolecule density in MBIS is a poor approximation to the slowly decaying molecular density for anions.

As discussed in section 1.5.3, the Achilles heel of variational Hirshfeld-I is its extreme sensitivity to method/basis set, and Figure 1.4 confirms this: variational Hirshfeld-I is the worst population analysis methods we considered by this measure. We wondered

whether using a different divergence measure might alleviate this sensitivity, so we also generated results using the Hellinger-Bhattacharya distance (with $\nu = 2$) as a divergence, i.e., replacing Eq. (1.14) with

$$\underbrace{\min}_{\left\{\{\rho_A(\mathbf{r})\},\{N_A^0\}\middle|\begin{array}{l}\sum_{A=1}^{N_{\text{atoms}}}\rho_A(\mathbf{r})=\rho_{\text{mol}}(\mathbf{r}) \\ \sum_{A=1}^{N_{\text{atoms}}}N_A^0=\sum_{A=1}^{N_{\text{atoms}}}N_A=N_{\text{mol}}\end{array}\right\}} \sum_{A=1}^{N_{\text{atoms}}}\int\left(\left(\rho_A(\mathbf{r})\right)^{1/\nu}-\left(\rho_A^0(\mathbf{r})\right)^{1/\nu}\right)^{\nu} d\mathbf{r}\,. \qquad (1.62)$$

As seen in Figure 1.4, however, this revision did not improve the performance of the variational Hirshfeld-I method.

While the original self-consistent Hirshfeld-I method is not as exquisitely sensitive to changes in basis set as variational-Hirshfeld-I, it still shows excessive dependence on basis set, especially for Nitrogen-containing molecules. (However, Hirshfeld-I is also more sensitive to basis set than one would like for the hydronium atom, H_3O^+.) This problem arises because the nitrogen anion does not exist in nature, and also does not exist at some of the levels of theory we considered. (However, the nitrogen anion is erroneously predicted to be bound for UB3LYP and UωB97XD for sufficiently large and diffuse basis sets.) However, the electron density of the nitrogen anion is an essential ingredient in the Hirshfeld-I procedure when the charge on the nitrogen AIM is negative (and greater than minus two); when the charge on the nitrogen AIM is less than minus two (and greater than minus three), the electron density of the nitrogen dianion is also required. The Hirshfeld-I method uses the electron densities of these basis-set-bound (or otherwise very weakly bound or unbound) anions, and this imparts undesirable basis-set-sensitivity to Hirshfeld-I. In particular, as seen in the Appendix 7.3, the electron

densities of the reference pro-atoms for the oxygen dianion and the nitrogen (di)anion are very sensitive to the presence of diffuse functions in the basis set.

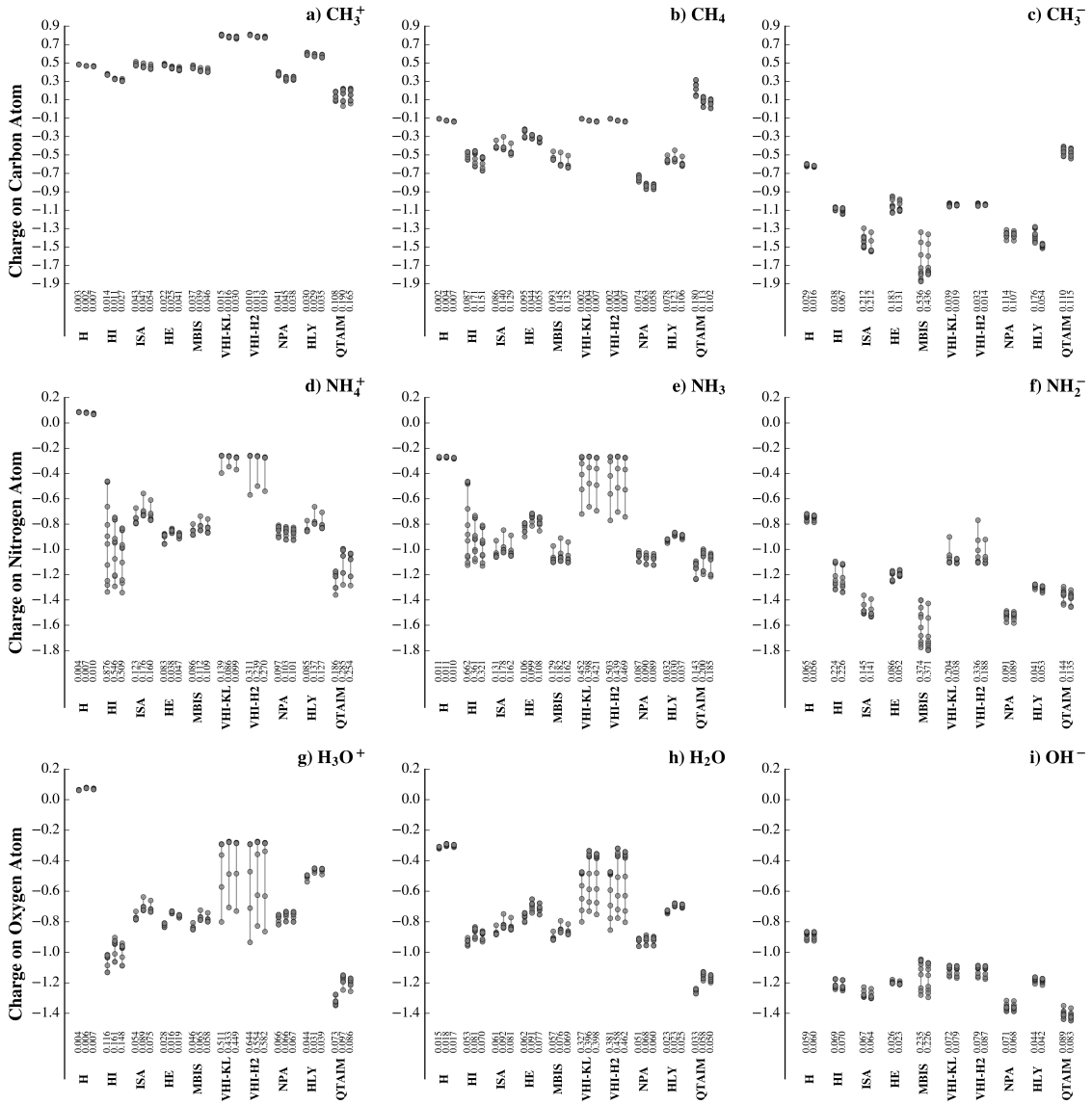
We expect that as one reduces the molecules in our set, the charge on the central atom will decrease. That is, we expect that $q_C(CH_3^+) > q_C(CH_4) > q_C(CH_3^-)$ for the carbon series, $q_N(NH_4^+) > q_N(NH_3) > q_N(NH_2^-)$ for the nitrogen series, and $q_O(H_3O^+) > q_O(H_2O) > q_O(OH^-)$ for the oxygen series). QTAIM and Mulliken charges violate the expected trend in all series, and Hirshfeld-I and Hirshfeld-E violate the trends for the nitrogen and oxygen series. Variational Hirshfeld-I does not entirely violate our intuition, but for many methods/basis sets the charge on the nitrogen atom barely changes when one moves from NH_4^+ to NH_3, which seems questionable, and is inconsistent with the results from most of the other population analysis methods.

The other methods give chemical trends that are largely in agreement with our expectations. This does not mean that those methods are flawless, however. For example, Hirshfeld partitioning is known to give problematic chemical trends for atoms containing large electron-rich atoms.[117]

While it is impossible to say what the "right" value for the charge of an AIM is, there are certain times when the charges fail to conform to our expectations. For example, it seems that Hirshfeld charges are usually "too small" in magnitude and that, conversely, the QTAIM charges are usually "too big". Likewise one can argue that the NPA charges seem to be too negative for the central atom in these molecular anions, and that it is especially counterintuitive to see carbon charges of ~ -0.8 from NPA in methane. The Hirshfeld-I charges in the nitrogen-containing molecules often seem too negative,

probably because the nitrogen pro-atom in these species is too diffuse. As mentioned before, the MBIS charges for CH_3^- and NH_2^- are very sensitive to basis set. In addition, the MBIS charges for these species are anomalously negative.

Figure 1.4 Comparison between partitioning schemes at different levels of theory. For each scheme, three columns plot the charges computed using twelve Dunning basis sets (d-aug-)ccpVXZ with X=D, T, Q, 5 basis functions) at UHF, UB3LYP, and UωB97XD levels of theory, respectively. (Only UB3LYP and UωB97XD give bound molecular anions, so there are only two columns for each partitioning scheme in the last column of figures.) The absolute range of the atomic charges obtained using various basis sets for each level of theory is summarized on the x-axis alongside the partitioning method. The methods used are Hirshfeld (H), Iterative Hirshfeld (HI), Iterative Stockholder Analysis (ISA), Hirshfeld-E (HE), Minimal Basis Stockholder Analysis (MBIS), Variational Hirshfeld-I with Kullback-Leibler divergence (Eq. (1.14); VHI-KL), Variational Hirshfeld-I with generalized Hellinger-Bhattacharya distance with ν = 2 (Eq. (1.62); VHI-H2), Natural Population Analysis (NPA), Hu-Lu-Yang electrostatic fitted charges (HLY), Quantum Theory of Atoms in Molecules (QTAIM).



1.6 Outline

In the first part of this thesis, entitled "Information-Theoretic Atoms in Molecules", we characterize the scope of Hirshfeld-like partitioning methods by precisely delimiting when the Hirshfeld partitioning is obtained, and when it is not.

In chapter 2, we expose conditions on the local divergence measures that are necessary, and sufficient, to recover the popular Hirshfeld partitioning. Specifically, we show that among all local measures of divergence between two probability distribution functions, the Hirshfeld partitioning is obtained only for *f*-divergences.

In chapter 3, this is extended by demonstrating that some nonlocal divergence functionals, namely the statistical divergence measures associated with non-extensive thermodynamic entropy functions like the Tsallis, Réyni, Sharma-Mittal, supraextensive, and *H*-divergences, are associated with the Hirshfeld atoms-in-molecules partitioning as well.

These findings dramatically extend the mathematical framework that one uses for similarity-based atoms-in-molecules partitioning by revealing that many different ways of measuring the divergence between densities lead to the Hirshfeld partitioning. In addition, it subsumes a large body of prior work, where the Hirshfeld partitioning was derived, and re-derived, by using different density divergences[1, 27, 29, 118] and has potential applications in computational algorithms for electronic structure theory, e.g., density-fitting.[82, 119-123] The mathematical tools presented in these chapters are suitable for measuring the divergence between other probability distribution functions that arise in quantum chemistry too. For example, there has been significant recent interest in approaches that

use the shape function,[27, 118, 124-125] instead of the electron density, to describe chemical phenomena.[126-130]

In the second part of this thesis, entitled "Optimal Pro-atom Densities", we address the important issue of selecting the best reference pro-atoms for partitioning the molecule into atoms.

In chapter 4, a general and flexible additive pro-atom density model is introduced. This pro-atom model is variationally optimized so that the promolecular density approximates the molecular density as accurately as possible; these pro-atoms are then used for conventional Hirshfeld partitioning. Inspired by the MBIS approach, we take advantage of the extended Kullback-Leibler divergences to measure the similarity between the molecular and pro-molecular density because: **a)** this results in a size consistent partitioning, **b)** this guarantees that the atom and pro-atom have the same number of electrons (without requiring any constraints), and **c)** this results in a convex optimization problem if the parameters to be optimized are linear. Specifically, we choose to express the pro-atom densities as nonnegative linear combinations of the atomic densities of all bound states of the atoms, which we call the Additive Variational Hirshfeld (AVH) partitioning scheme.

In chapter 5, the multiplicative pro-atom density model is described. This model is based on realization that the additive atomic density model in the chapter 4 can be viewed as a weighted arithmetic average of the spherically-averaged densities of the bound atomic ions. Using instead the weighted geometric average leads to the multiplicative pro-atom model. The advantage of the multiplicative model is that it allows us to easily

control the asymptotic behavior of atomic density. For example, based on electronegativity equalization principle, we can constrain all atoms to have the same ionization potential, and set their common ionization potential to equal the molecule's ionization potential. In addition, through this pro-atom model, one can easily make sure that the pro-atom densities have the correct nuclear cusps. The presentation of the geometric model is followed by a brief numerical assessment of the models laid out in chapters 4 and 5, and a recapitulation of our most important findings.

These flexible pro-atom models provide a new vista on the problem of partitioning the molecular density. Because the pro-atom densities can be variationally optimized concurrently with the density of the AIM, they allow one to add constraints in a straightforward manner. We expect that these elegant mathematical and chemical features improve the quality and transferability of the Hirshfeld charges significantly. Specifically, we believe that the Additive Variational Hirshfeld (AVH) partitioning scheme, with its many superior mathematical and chemical features, may be the best possible Hirshfeld partitioning scheme.

2 Local Divergence Measures

2.1 Background

In 1986, Rychlewski and Parr,[131] suggested that the electron density of an atom in a molecule should be chosen to minimize the deviation of the atom-in-molecule density, $\rho_A(\mathbf{r})$, from a reference pro-atomic density, $\rho_A^0(\mathbf{r})$, subject to the constraint that the sum of the atomic densities is equal to the molecular density,

$$\rho_{\text{molecule}}(\mathbf{r}) = \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r}) \tag{2.1}$$

Specifically, Rychlewski and Parr used the energy-gap between the atom in a molecule and the reference pro-atom to quantify the deviation between them, obtaining the partitioning procedure,

$$\underbrace{\min}_{\left\{\rho_A(\mathbf{r}) \middle| \rho_{\text{molecule}}(\mathbf{r}) = \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r})\right\}} \sum_{A=1}^{N_{\text{atoms}}} \left(\mathcal{E}_{v_A}[\rho_A] - \mathcal{E}_{v_A}[\rho_A^0]\right) \tag{2.2}$$

(Here $\mathcal{E}_v[\rho] = F[\rho] + \langle \rho v \rangle$ is the variational Hohenberg-Kohn energy functional.[37]) While the Rychlewski-Parr procedure, which is often called partition-density functional theory, has reemerged in recent years, it is computationally problematic and can, at least in some cases, give atoms in molecules with delocalized densities.[132-135]

More than a decade later, Nalewajski and Parr revisited this procedure, using the Kullback-Leibler entropy (or information)[136-137] to measure the deviation of the atom-in-molecule density from the reference pro-atomic density,[25-26]

$$\underbrace{\min}_{\left\{\rho_A(\mathbf{r})\middle|\rho_{\text{molecule}}(\mathbf{r})=\sum_{A=1}^{N_{\text{atoms}}}\rho_A(\mathbf{r})\right\}}\sum_{A=1}^{N_{\text{atoms}}}\int\rho_A(\mathbf{r})\ln\left(\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})}\right)d\mathbf{r}\,. \tag{2.3}$$

In stark contrast to the computational difficulties attendant to the Rychlewski-Parr partitioning, the Nalewajski-Parr procedure can be performed *analytically*, giving

$$\rho_A(\mathbf{r})=\left(\frac{\rho_A^0(\mathbf{r})}{\sum_{B=1}^{N_{\text{atoms}}}\rho_B^0(\mathbf{r})}\right)\rho_{\text{molecule}}(\mathbf{r})\,. \tag{2.4}$$

Remarkably, as introduced in previous chapter, Eq. (2.4) is the same partitioning that Hirshfeld had proposed more than two decades earlier, on purely heuristic grounds.[21]

Since then, many researchers have elaborated upon this basic approach, either by generalizing (and even optimizing) the choice of pro-atomic densities[31, 58, 60, 72-73, 78, 80] or by using alternative measures of the deviation between densities.[1, 27, 29, 118] Remarkably, the Hirshfeld partitioning formula, Eq. (2.4) is frequently recovered, even when the deviation between the atomic and pro-atomic densities is measured using functionals that are very dissimilar to the Kullback-Leibler divergence in Eq. (2.3).

So, how pervasive *is* the Hirshfeld partitioning? Does the Hirshfeld partitioning inevitably arise, no matter how the deviation between densities is measured? Clearly not, because the Rychlewski-Parr partitioning does not give Eq. (2.4). Does the Hirshfeld

partitioning arise whenever a local functional is used to measure the deviation between densities? Again, no, because using the squared $\mathbb{L}^2$-distance,

$$\sum_{A=1}^{N_{\text{atoms}}} \int \left(\rho_A(\mathbf{r}) - \rho_A^0(\mathbf{r}) \right)^2 d\mathbf{r}, \tag{2.5}$$

in place of the Kullback-Leibler divergence in Eq. (2.3) gives absurdly delocalized atomic densities, specifically,

$$\rho_A(\mathbf{r}) = \rho_A^0(\mathbf{r}) + \frac{1}{N_{\text{atoms}}} \left(\rho_{\text{molecule}}(\mathbf{r}) - \sum_{B=1}^{N_{\text{atoms}}} \rho_B^0(\mathbf{r}) \right). \tag{2.6}$$

However, the Hirshfeld partitioning is remarkably pervasive. In this chapter, we consider divergence measures that are local functionals of the electron density. These can be written as

$$H_{\text{local}}[\rho] = \int h(\rho(\mathbf{r})) d\mathbf{r} \tag{2.7}$$

where $h(x) : \mathbb{R}^+ \to \mathbb{R}$ is an ordinary function. Equivalently, to evaluate the functional derivative of a local functional at a point, one needs to only know the electron density at that point,

$$\frac{\delta H_{\text{local}}[\rho]}{\delta \rho(\mathbf{r})} = \left. \frac{dh(x)}{dx} \right|_{x=\rho(\mathbf{r})} \tag{2.8}$$

Using a similar approach to Nalewajski and Parr, these measures are studied thoroughly, and it is shown that any *f*-divergence[138-140] between the densities of the atom and pro-atom,

$$\sum_{A=1}^{N_{\text{atoms}}} \int \rho_A(\mathbf{r}) f\left(\frac{\rho_A^0(\mathbf{r})}{\rho_A(\mathbf{r})}\right) d\mathbf{r} \tag{2.9}$$

suffices to recover the Hirshfeld partitioning, Eq. (2.4). Here, f is *any* convex function with $f(1)=0$.[138-140] The Kullback-Leibler divergence used by Nalewajski-Parr in Eq. (2.3), is obviously a special case of an *f*-divergence, corresponding to the choice of $f(x)=-\ln(x)$. We also show that having an *f*-divergence is also necessary for the Hirshfeld partitioning. That is, no other local measure of the deviation between densities recovers the Hirshfeld partitioning. A few especially interesting families of *f*-divergence are also characterized at the end of this chapter.

2.2 *f*-divergence is Sufficient for the Hirshfeld Partitioning

Suppose that one chooses the densities of the atoms in a molecule by minimizing their *f*-divergence from the densities of their corresponding reference pro-atoms, subject to the constraint that the sum of the atomic densities recovers the total molecular density,

$$\underbrace{\min}_{\left\{\rho_A(\mathbf{r}) \middle| \rho_{\text{molecule}}(\mathbf{r}) = \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r})\right\}} \sum_{A=1}^{N_{\text{atoms}}} \int \rho_A(\mathbf{r}) f\left(\frac{\rho_A^0(\mathbf{r})}{\rho_A(\mathbf{r})}\right) d\mathbf{r} \tag{2.10}$$

An explicit equation for the atomic densities is obtained by solving the system of nonlinear equations,

$$\left\{0 = \frac{\delta}{\delta \rho_A(\mathbf{r})}\left[\sum_{A=1}^{N_{\text{atoms}}} \int \rho_A(\mathbf{r}) f\left(\frac{\rho_A^0(\mathbf{r})}{\rho_A(\mathbf{r})}\right) d\mathbf{r} - \int \lambda(\mathbf{r})\left(\rho_{\text{molecule}}(\mathbf{r}) - \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r})\right) d\mathbf{r}\right]\right\}_{A=1}^{N_{\text{atoms}}} \tag{2.11}$$

The term in brackets is the Lagrangian corresponding to the optimization formulated in (2.10), and $\lambda(\mathbf{r})$ is the Lagrange multiplier for the constraint.

Evaluating the functional derivatives in Eq. (2.11), we obtain the result

$$\left\{ g\left(\frac{\rho_A^0(\mathbf{r})}{\rho_A(\mathbf{r})} \right) = \lambda(\mathbf{r}) \right\}_{A=1}^{N_{\text{atoms}}}, \tag{2.12}$$

where

$$g(x) = x\left(\frac{df(x)}{dx} \right) - f(x) \tag{2.13}$$

is the Legendre transform of $f(x)$. Equation (2.12) implies that, for any two atoms,

$$g\left(\frac{\rho_A^0(\mathbf{r})}{\rho_A(\mathbf{r})} \right) = g\left(\frac{\rho_B^0(\mathbf{r})}{\rho_B(\mathbf{r})} \right). \tag{2.14}$$

But, because *f* is convex, *g* is monotonically increasing, and therefore invertible. Since *g* is invertible, Eq. (2.14) is equivalent to the simpler statement

$$\frac{\rho_A^0(\mathbf{r})}{\rho_A(\mathbf{r})} = \frac{\rho_B^0(\mathbf{r})}{\rho_B(\mathbf{r})} \tag{2.15}$$

Equation (2.15) is the key relation from which the Hirshfeld partitioning follows.[26-27] This is most easily seen by rewriting Eq. (2.15) as

$$\rho_B(\mathbf{r})\rho_A^0(\mathbf{r}) = \rho_B^0(\mathbf{r})\rho_A(\mathbf{r}). \tag{2.16}$$

Summing both sides over all atoms *A* and using the constraint that the sum of atomic densities is the molecular density recovers Eq. (2.4). A detailed derivation can be found in Appendix 7.1.

2.3 *f*-divergence is Necessary for the Hirshfeld Partitioning

2.3.1 Local Measures of Deviation Between Densities

In section 2.2, we showed that whenever the deviation between the electron densities of atoms from their corresponding reference pro-atoms is quantified by an *f*-divergence, the Hirshfeld partitioning inexorably arises. We now show that the converse is also true.

To prove the converse, we must characterize the set of all "reasonable" ways of quantifying the deviation between two densities, and show that the Hirshfeld partitioning only arises when the formula for the deviation is an *f*-divergence.

To motivate our specific approach, note that the Hirshfeld partitioning is local: the atomic density of atom A at the point $\mathbf{r}$ depends only on the molecular density at $\mathbf{r}$ and the density of the pro-atoms at $\mathbf{r}$. (Not every partitioning satisfies this requirement. For example, the Rychlewski-Parr partitioning (a.k.a. partition DFT) does not.) We therefore choose to impose an *axiom of locality*: the density of an atom at $\mathbf{r}$ is a local property, and does not depend on the density of the molecule or pro-atoms at points $\mathbf{r}' \neq \mathbf{r}$. The axiom of locality is equivalent to assuming that the deviation between the densities of the atom and pro-atom is quantified by a local functional, with the general form

$$H\left[\rho,\rho^0\right] \equiv \int h\left(\rho(\mathbf{r}),\rho^0(\mathbf{r})\right) d\mathbf{r} \,. \tag{2.17}$$

This is only a sensible measure of the deviation between two densities if H satisfies

$$H\left[\rho,\rho^0\right] > H\left[\rho^0,\rho^0\right] = 0 \,, \tag{2.18}$$

whenever the density and the pro-density have the same number of electrons,

$$\int \rho(\mathbf{r})\, d\mathbf{r} = \int \rho^0(\mathbf{r})\, d\mathbf{r}\,. \tag{2.19}$$

2.3.2 Local Measure of Deviation Giving Hirshfeld Partitioning

To characterize which functions, $\{h(x,y)\}$, give the Hirshfeld partitioning, we notice that if

$$\frac{\delta H[\rho,\rho^0]}{\delta\rho(\mathbf{r})} = \left(\frac{\partial h(x,y)}{\partial x}\right)_{\substack{x=\rho(\mathbf{r}) \\ y=\rho^0(\mathbf{r})}} = \phi\left(\frac{\rho(\mathbf{r})}{\rho^0(\mathbf{r})}\right), \tag{2.20}$$

where $\phi(x)$ is an invertible function for $x \geq 0$, then Hirshfeld partitioning is recovered because

$$\begin{gathered} \underbrace{\min}_{\left\{\rho_A(\mathbf{r}) \middle| \rho_{\text{molecule}}(\mathbf{r}) = \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r})\right\}} \sum_{A=1}^{N_{\text{atoms}}} \int h\left(\rho_A(\mathbf{r}), \rho_A^0(\mathbf{r})\right) d\mathbf{r} \\ \phi\left(\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})}\right) = \phi\left(\frac{\rho_B(\mathbf{r})}{\rho_B^0(\mathbf{r})}\right) = \lambda(\mathbf{r}) \\ \frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} = \frac{\rho_B(\mathbf{r})}{\rho_B^0(\mathbf{r})} \end{gathered} \tag{2.21}$$

This last equation implies the Hirshfeld partitioning, Eq. (2.4).

The converse is also true: the Hirshfeld partitioning is not obtained unless the last line of Eq. (2.21) is true. The last line of Eq. (2.21) is, furthermore, equivalent to the second line. The second line, however, leads to Eq. (2.20). To show this, consider what would happen if the Eq. (2.20) were replaced by the more general result,

$$\frac{\delta H\left[\rho,\rho^0\right]}{\delta\rho(\mathbf{r})}=\left(\frac{\partial h(x,y)}{\partial x}\right)_{\substack{x=\rho(\mathbf{r})\\ y=\rho^0(\mathbf{r})}}=\gamma\left(\rho(\mathbf{r}),\rho^0(\mathbf{r})\right) \tag{2.22}$$

Then the second line of Eq. (2.21) would be replaced by the more general equation,

$$\gamma\left(\rho_A(\mathbf{r}),\rho_A^0(\mathbf{r})\right)=\gamma\left(\rho_B(\mathbf{r}),\rho_B^0(\mathbf{r})\right) \tag{2.23}$$

The solution to this equation is the last line of Eq. (2.21) (i.e., the Hirshfeld partitioning) only if $\gamma\left(\rho(\mathbf{r}),\rho^0(\mathbf{r})\right)$ is, in fact, merely a function of $\rho(\mathbf{r})/\rho^0(\mathbf{r})$. Therefore we may rewrite the function as $\gamma\left(\rho(\mathbf{r})/\rho^0(\mathbf{r})\right)$. In addition, this function must be invertible, for otherwise there would—at least in some cases—be alternative, non-Hirshfeld partitionings consistent with Eq. (2.23).

2.3.3 Functional Form of a Local Measure of Deviation Giving Hirshfeld Partitioning

We now need to show that all local measures of the deviation between two densities that satisfy Eqs. (2.18) and (2.20) are *f*-divergences. To do this, we must determine when the functional derivative relation in Eq. (2.20) is satisfied. For a point **r** and pro-density ρ^0, Eq. (2.20) is an ordinary differential equation in x with the form:

$$\frac{dh(x,y)}{dx}=\phi\left(\frac{x}{y}\right), \tag{2.24}$$

where y is a constant and $h(y,y) = 0$ (from Eq. (2.18)). This equation may be formally solved by separation of variables, giving,

$$h(x,y) = y\left(\Phi\left(\frac{x}{y}\right) - \Phi(1)\right) \tag{2.25}$$

where (cf. Eq. (2.20))

$$\phi(u) = \frac{d\Phi(u)}{du} \tag{2.26}$$

is required to be invertible. That is, the Hirshfeld partitioning is obtained whenever the deviation between densities is measured using the local functional,

$$H\left[\rho,\rho^0\right] = \int \rho^0(\mathbf{r})\left[\Phi\left(\frac{\rho(\mathbf{r})}{\rho^0(\mathbf{r})}\right) - \Phi(1)\right] d\mathbf{r} \geq 0 \tag{2.27}$$

The function ϕ is invertible only if it is monotonic: i.e., it is either a strictly increasing or strictly decreasing function of u. This means, in turn, that Φ must be either strictly convex or strictly concave.

However, Eq. (2.18) implies that Φ is convex. To see this, consider that if Φ were concave, then Φ would be less than or equal to the value of its tangent line at $u = 1$,

$$\Phi(u) \leq \phi(1)u + \left(\Phi(1) - \phi(1)\right). \tag{2.28}$$

Therefore,

$$\begin{aligned}\int \rho^0(\mathbf{r})\left[\Phi\left(\frac{\rho(\mathbf{r})}{\rho^0(\mathbf{r})}\right) - \Phi(1)\right] d\mathbf{r} &\leq \int \phi(1)\rho^0(\mathbf{r})\left[\frac{\rho(\mathbf{r})}{\rho^0(\mathbf{r})} - 1\right] d\mathbf{r} \\ &\leq \phi(1)\int\left(\rho(\mathbf{r}) - \rho^0(\mathbf{r})\right) d\mathbf{r}\end{aligned} \tag{2.29}$$

Consider the special case where the density and the pro-density have the same number of electrons, Eq. (2.19). In this case, Eq. (2.29) implies that $H[\rho,\rho^0] \leq 0$, contradicting our initial assumptions (cf. Eq. (2.18)).

2.3.4 *f*-Divergences are Local Measures of Deviation Giving the Hirshfeld Partitioning

In the previous section, we showed that if a local measure of the deviation between two densities recovers the Hirshfeld partitioning, it has the form (2.27), where Φ is convex. We now show that this is equivalent to assuming that $H[\rho,\rho^0]$ is an *f*-divergence.

Equation (2.27) measures the deviation between densities by a local functional with the form

$$\int \rho^0(\mathbf{r}) F\left(\frac{\rho(\mathbf{r})}{\rho^0(\mathbf{r})}\right) d\mathbf{r} \geq 0, \tag{2.30}$$

where F is a convex functional with $F(1) = 0$.

Equation (2.30) is the form of an *f*-divergence (cf. Eq. (2.9)) except that the roles of the density and the pro-density are interchanged. However, choose F to have the special form

$$F\left(\frac{\rho(\mathbf{r})}{\rho^0(\mathbf{r})}\right) = \frac{\rho(\mathbf{r})}{\rho^0(\mathbf{r})} f\left(\frac{\rho^0(\mathbf{r})}{\rho(\mathbf{r})}\right), \tag{2.31}$$

where $f(x)$ is a convex function with $f(1) = 0$. Then Eq. (2.30) is equivalent to the usual form of *f*-divergence,

$$\int \rho(\mathbf{r}) f\left(\frac{\rho^0(\mathbf{r})}{\rho(\mathbf{r})}\right) d\mathbf{r} \geq 0. \tag{2.32}$$

To show that the only local measures of the deviation between densities that recover the Hirshfeld partitioning are *f*-divergences, we need to show that Eq. (2.31) is

always an allowable form of the function F. Since $f(1) = 0$, $F(1) = 0$. So we need only show that F is convex. If we assume that F is twice-differentiable, then

$$\frac{d^2F}{du^2} = \frac{d^2}{du^2}\left(uf\left(u^{-1}\right)\right) = u^{-3}\frac{d^2 f\left(u^{-1}\right)}{d\left(u^{-1}\right)^2} > 0\,, \tag{2.33}$$

which establishes the convexity of the form of $F(u)$. A slightly more general derivation starts from the derivative,

$$\frac{dF}{du} = -\frac{d}{du}\left(uf\left(u^{-1}\right)\right) = -\left(u^{-1}f'\left(u^{-1}\right) - f\left(u^{-1}\right)\right) \tag{2.34}$$

The right-hand side of this equation is minus one times the Legendre transform of $f(u^{-1})$, so it is an increasing function of u^{-1}. Since $\mathrm{d}F/\mathrm{d}u$ is an increasing function of u, $F(u)$ is convex.

Summarizing, among all the possible local functionals that might be used to measure the deviation between two densities, only the *f*-divergences give the Hirshfeld partitioning.

2.4 Characterizing *f*-divergences

Because the family of *f*-divergences is so diverse, and because the preceding treatment is abstract, it seems beneficial to present some specific members of the family that have nice properties. Subsections 2.4.1 and 2.4.2 present two families of *f*-divergence that can be given explicit parameterizations. Specifically, in subsection 2.4.1, we assume that the *f*-divergence has a Taylor series expansion (entire *f*-divergences). In subsection 2.4.2, we assume that the *f*-divergence is also a Bregman divergence (α-divergences).

Like most *f*-divergences, these families treat the densities of atoms and pro-atoms inequivalently, which means that one cannot interpret the *f*-divergence as a measure of "distance" between the atom and the pro-atom. In subsection 2.4.3, we discuss ways to symmetrize the *f*-divergence. While most symmetrized *f*-divergences are not distance metrics (they do not satisfy the triangle inequality), appropriately symmetrized α-divergences are squared metrics.[141-142]

2.4.1 Entire *f*-divergences

In this subsection we assume that $f(x)$ is entire, which means that it can be expressed as a Taylor series. To construct the Taylor series for $f(x)$, we recall that every positive polynomial can be written as the sum of the squares of two polynomials,

$$\begin{aligned} P(x) &= \big(p(x)\big)^2 + \big(q(x)\big)^2 \geq 0 \\ p(x) &= a_0 + a_1 x + a_2 x^2 + \cdots \\ q(x) &= b_0 + b_1 x + b_2 x^2 + \cdots \end{aligned} \tag{2.35}$$

Therefore,

$$\begin{aligned} P(x) &= \left(a_0^2 + b_0^2\right) + \left(2a_1 a_0 + 2b_1 b_0\right)x + \cdots + \left(\sum_{i=0}^{n}\left(a_i a_{n-i} + b_i b_{n-i}\right)\right)x^n + \cdots \\ &\geq 0 \end{aligned} \tag{2.36}$$

By choosing $P(x)$ to be the second derivative of $f(x)$, we ensure that $f(x)$ is convex. Integrating $P(x)$ twice, we obtain

$$\begin{aligned} f(x) = A + Bx + \tfrac{1}{2}\left(a_0^2 + b_0^2\right)x^2 + \tfrac{1}{6}\left(2a_1 a_0 + 2b_1 b_0\right)x^3 + \cdots \\ \cdots + \frac{1}{(n+2)(n+1)}\left(\sum_{i=0}^{n}\left(a_i a_{n-i} + b_i b_{n-i}\right)\right)x^{n+2} + \cdots \end{aligned} \tag{2.37}$$

The value of A is determined by the requirement that $f(1) = 0$,

$$A = -\left(\begin{array}{l} B + \frac{1}{2}\left(a_0^2 + b_0^2\right) + \frac{1}{6}\left(2a_1a_0 + 2b_1b_0\right) + \cdots \\ \qquad \cdots + \dfrac{1}{(n+2)(n+1)}\left(\sum_{i=0}^{n} \left(a_i a_{n-i} + b_i b_{n-i}\right) \right) + \cdots \end{array} \right), \tag{2.38}$$

but all the other parameters in Eq. (2.37) are arbitrary real numbers; different choices for these parameters lead to different members of the family of entire *f*-divergences. This seems to be the largest family of *f*-divergences that can be explicitly parameterized.

2.4.2 α-divergences

In the literature, the most popular choices for the *f*-divergence are members of the family of α-divergences, with

$$f(x) = \frac{x^\alpha - 1}{\alpha(\alpha - 1)} \tag{2.39}$$

where α is a real number. (Note that the α-divergences are not entire except when α = 2,3,…) The α-divergences are equivalent to the Tsallis divergences considered in ref. [118]. I.e.,

$$\begin{aligned} H_\alpha\left[\rho, \rho^0\right] &= \frac{1}{\alpha(\alpha-1)} \int \rho(\mathbf{r}) \left(\left(\frac{\rho^0(\mathbf{r})}{\rho(\mathbf{r})} \right)^\alpha - 1 \right) d\mathbf{r} \\ &= \frac{1}{\alpha(\alpha-1)} \int \left(\rho(\mathbf{r})\right)^{1-\alpha} \left(\rho^0(\mathbf{r})\right)^\alpha - \rho(\mathbf{r})\, d\mathbf{r} \end{aligned} \tag{2.40}$$

In the α→0 limit, this is an indeterminate form, and one has the Kullback-Leibler directed divergence,

$$H_{\alpha=0}\left[\rho,\rho^0\right]=\int\rho(\mathbf{r})\ln\left(\frac{\rho(\mathbf{r})}{\rho^0(\mathbf{r})}\right)d\mathbf{r} \tag{2.41}$$

The form is unbounded (not indeterminant) as $\alpha\to1$ unless the pro-molecular density and molecular density have the same number of electrons. When that is true, however, Eq. (2.40) reduces to

$$H_{\alpha=1}\left[\rho,\rho^0\right]=\int\rho^0(\mathbf{r})\ln\left(\frac{\rho^0(\mathbf{r})}{\rho(\mathbf{r})}\right)d\mathbf{r}\,. \tag{2.42}$$

Among the innumerable possibilities for the *f*-divergence, the family of α-divergences is special, and deserves further scrutiny. If **(1)** the α-divergence is symmetrized as discussed in section 2.4.3, **(2)** the deviation between atomic and pro-atomic densities is measured relative to their average, and **(3)** the molecular density and pro-molecular density contain the same number of electrons, then the resulting *f*-divergence is the square of a distance metric. This enriches our understanding of the Hirshfeld partitioning, by clarifying what we mean when we say that the densities of Hirshfeld atoms are "as close as possible" to the densities of the corresponding pro-atoms. Specifically, we may say that the Hirshfeld atomic densities minimize the α-distance to the pro-atomic densities.

2.4.3 Symmetrized *f*-divergences

The specific *f*-divergences considered in the preceding sections are *directed* divergences: the density and the pro-density enter the formula in different ways, so in general, $H[\rho,\rho^0]\neq H[\rho^0,\rho]$. However, as we showed in section 2.3.4, for any asymmetric *f*-divergence, one can formulate another *f*-divergence in which the roles of the density and

the pro-density are interchanged. Averaging these two *f*-divergences gives a symmetric *f*-divergence.

Therefore, for any (possibly asymmetric) *f*-divergence, there exists a corresponding symmetrized *f*-divergence, with $f_{\text{sym}}\left(x\right)=\frac{1}{2}f\left(x\right)+\frac{1}{2}xf\left(x^{-1}\right)$. This divergence has the form

$$\begin{aligned} H_{\text{sym}}\left[\rho,\rho^{0}\right] &= \int \frac{\rho\left(\mathbf{r}\right)}{2} f\left(\frac{\rho^{0}\left(\mathbf{r}\right)}{\rho\left(\mathbf{r}\right)}\right)+\frac{\rho^{0}\left(\mathbf{r}\right)}{2} f\left(\frac{\rho\left(\mathbf{r}\right)}{\rho^{0}\left(\mathbf{r}\right)}\right) d\mathbf{r} \\ &= \tfrac{1}{2}\left(H\left[\rho,\rho^{0}\right]+H\left[\rho^{0},\rho\right]\right) \end{aligned} \tag{2.43}$$

The symmetrized *f*-divergence in Eq. (2.43) is still not a distance metric because it does not, in general, satisfy the triangle inequality. (The Hellinger distance, $f\left(x\right)=\left(\sqrt{x}-1\right)^{2}$, is an exception.[29, 143]) Consider what happens if we also symmetrize the reference densities, so that both of the divergences in Eq. (2.43) are measured relative to the average of the atomic and pro-atomic densities,

$$\begin{aligned} H_{\text{sym}}\left[\rho,\rho^{0}\right] &= \int \frac{\rho\left(\mathbf{r}\right)}{2} f\left(\frac{\frac{1}{2}\left(\rho\left(\mathbf{r}\right)+\rho^{0}\left(\mathbf{r}\right)\right)}{\rho\left(\mathbf{r}\right)}\right)+\frac{\rho^{0}\left(\mathbf{r}\right)}{2} f\left(\frac{\frac{1}{2}\left(\rho\left(\mathbf{r}\right)+\rho^{0}\left(\mathbf{r}\right)\right)}{\rho^{0}\left(\mathbf{r}\right)}\right) d\mathbf{r} \\ &= \tfrac{1}{2}\left(H\left[\rho,\tfrac{1}{2}\left(\rho+\rho^{0}\right)\right]+H\left[\rho^{0},\tfrac{1}{2}\left(\rho+\rho^{0}\right)\right]\right) \end{aligned} \tag{2.44}$$

This is still an *f*-divergence because if $f(x)$ is convex and has $f(1) = 0$, then $f_{\text{symref}}\left(x\right)=f\left(\frac{1}{2}+\frac{x}{2}\right)$ also has those properties. For the special case where $f(x)$ is an α-divergence and the pro-molecular density and molecular density contain the same number of electrons, Eq. (2.44) is the square of a distance measure,[141-142]

$$D_{\alpha}^{2}\left(\rho,\rho^{0}\right)=\frac{\int\left(\left[\left(\rho\left(\mathbf{r}\right)\right)^{1-\alpha}+\left(\rho^{0}\left(\mathbf{r}\right)\right)^{1-\alpha}\right]\left[\frac{1}{2}\left(\rho\left(\mathbf{r}\right)+\rho^{0}\left(\mathbf{r}\right)\right)\right]^{\alpha}-\left(\rho\left(\mathbf{r}\right)+\rho^{0}\left(\mathbf{r}\right)\right)\right)d\mathbf{r}}{2\alpha\left(\alpha-1\right)} \qquad (2.45)$$

In the cases where this is an indeterminant form, one recovers the Jensen-Shannon divergence,

$$D_{\alpha=\{0,1\}}^{2}\left(\rho,\rho^{0}\right)=\frac{1}{2}\int\rho\left(\mathbf{r}\right)\ln\left(\frac{\rho\left(\mathbf{r}\right)}{\frac{1}{2}\left(\rho\left(\mathbf{r}\right)+\rho^{0}\left(\mathbf{r}\right)\right)}\right)+\rho^{0}\left(\mathbf{r}\right)\ln\left(\frac{\rho^{0}\left(\mathbf{r}\right)}{\frac{1}{2}\left(\rho\left(\mathbf{r}\right)+\rho^{0}\left(\mathbf{r}\right)\right)}\right)d\mathbf{r}\,. \qquad (2.46)$$

As discussed in ref. [29], there are interpretative advantages when one uses a distance metric to measure the deviation between the atomic densities and their corresponding pro-atomic densities.

2.5 Selecting Reference Pro-atoms

A desirable feature of the Rychlewski-Nalewajski-Parr approach to defining atoms in a molecule is that, by minimizing the deviation between the atomic densities and the pro-atomic densities, one ensures that the properties of the pro-atoms are "as transferable as possible" to the atoms in a molecule. (The optimality of this transferability can be precisely specified, but it obviously depends on the way one measures the deviation between the atomic and pro-atomic densities.[27]) Given a choice of several reference pro-atomic densities (e.g., the densities of isolated atoms with various charges), we can maximize the transferability of these reference atoms' properties to the atoms in the molecule by minimizing the *f*-divergence as a function of the reference pro-atoms. This is extensively addressed in chapters 4 and 5.

For concreteness, consider the special case where the pro-atomic densities depend only on the number of electrons in the pro-atoms, denoted by $\rho_A^0\left(\mathbf{r},N_A^0\right)$. The optimal pro-atoms can be obtained by minimizing the expression in Eq. (2.10) with respect to the populations of the reference pro-atoms,

$$\underbrace{\min}_{\left\{N_A^0\right\}}\ \underbrace{\min}_{\left\{\rho_A(\mathbf{r})\middle|\rho_{\text{molecule}}(\mathbf{r})=\sum_{A=1}^{N_{\text{atoms}}}\rho_A(\mathbf{r})\right\}}\ \sum_{A=1}^{N_{\text{atoms}}}\int\rho_A\left(\mathbf{r}\right)f\left(\frac{\rho_A^0\left(\mathbf{r},N_A^0\right)}{\rho_A\left(\mathbf{r}\right)}\right)d\mathbf{r} \tag{2.47}$$

The inner minimization can be performed analytically, with the solution in Eq. (2.4). Substituting this into Eq. (2.47),

$$\begin{aligned}&\underbrace{\min}_{\left\{N_A^0\right\}}\sum_{A=1}^{N_{\text{atoms}}}\int\left(\rho_A^0\left(\mathbf{r},N_A^0\right)\frac{\rho_{\text{molecule}}\left(\mathbf{r}\right)}{\sum_{B=1}^{N_{\text{atoms}}}\rho_B^0\left(\mathbf{r},N_B^0\right)}\right)f\left(\frac{\rho_A^0\left(\mathbf{r},N_A^0\right)\sum_{B=1}^{N_{\text{atoms}}}\rho_B^0\left(\mathbf{r},N_B^0\right)}{\rho_A^0\left(\mathbf{r},N_A^0\right)\rho_{\text{molecule}}\left(\mathbf{r}\right)}\right)d\mathbf{r}\\&=\underbrace{\min}_{\left\{N_A^0\right\}}\sum_{A=1}^{N_{\text{atoms}}}\int\rho_{\text{molecule}}\left(\mathbf{r}\right)f\left(\frac{\rho_{\text{molecule}}^0\left(\mathbf{r},\left\{N_A^0\right\}\right)}{\rho_{\text{molecule}}\left(\mathbf{r}\right)}\right)d\mathbf{r}\end{aligned} \tag{2.48}$$

In the second line, the pro-molecular density is defined as

$$\rho_{\text{molecule}}^0\left(\mathbf{r},\left\{N_A^0\right\}\right)=\sum_{A=1}^{N_{\text{atoms}}}\rho_A^0\left(\mathbf{r},N_A^0\right) \tag{2.49}$$

This indicates that—at least from the standpoint of transferability—the optimal pro-atoms should be chosen so that the deviation of the molecular density from the pro-molecular density is as small as possible. This suggests a two-step approach to atomic partitioning procedure based on the *f*-divergence:

1. Determine the optimal pro-atomic densities by minimizing the *f*-divergence between the pro-molecular and molecular densities, Eq. (2.48). Different *f*-divergences will give different optimal pro-atoms.
2. Determine the densities of the atoms in a molecule that maximally resemble the specified reference pro-atoms using the explicit Eq. (2.4). For a given choice of pro-atoms, different *f*-divergences give the same atomic densities.

In accord with the axiom of locality, the atomic density at a point in space, **r**, depends only on the molecular density and the pro-atomic densities at **r**. However, the value of the pro-atomic density at **r** depends on the molecular density (and the density of the other pro-atoms) at all points in space, through the integral in Eq. (2.48). It seems unclear which *f*-divergences give pro-atoms whose locality is most consistent with chemical intuition. The choice is *f*-divergence is discussed more in chapter 4.

However, it is reasonable to speculate that *f*-divergences that exaggerate the nearly-inevitable large asymptotic deviations between the molecular and pro-molecular densities (e.g., the Kullback-Leibler form, where $f(x\to 0)$ diverges) are less than ideal. It may be preferable to choose an *f*-divergence that prevents divergent values of x from contributing disproportionately to the integral in Eq. (2.48) (e.g., $f\left(x\right)=2^{1-x}-1$ or $f\left(x\right)=\left(1-x\right)/\left(1+x\right)$).

One can also constrain the pro-atoms to have desirable properties by imposing constraints on the optimization in Eq. (2.47). The variational Hirshfeld method in ref. [72], which forces atoms and pro-atoms to have the same charge, does this. (However, because that method constrains the pro-atomic charges directly, it couples the inner and outer

minimization in Eq. (2.47), and so Eq. (2.48) is not valid for that method.) It would be very useful to constrain the pro-molecule and the molecule to have the same number of electrons. (However, for two inequivalent, well-separated, subsystems, imposing any property of the entire system as a constraint leads to a population analysis method that is not size-consistent.) More generally, one may constrain certain atoms or functional groups to have specified charges (perhaps to ensure consistency with a molecular mechanics force field).

2.6 Conclusion

In this chapter, we have shown that among all local measures of divergence, the family of *f*-divergence measures is necessary and sufficient to recover the Hirshfeld partitioning scheme. Special cases of *f*-divergence measures were characterized. And it was demonstrated that for Hirshfeld atoms-in-molecule, the total *f*-divergence of all atomic densities relative to reference pro-atom density is equivalent to the *f*-divergence between molecular and pro-molecule densities.

3 Nonlocal Divergence Measures

3.1 Background

This work was initiated when our numerical investigations revealed that optimizing the pro-atoms (as will be discussed in chapters 4 and 5), gave the same results for the Tsallis and Réyni divergences. We were surprised that the Réyni divergence, even though it is not an *f*-divergence, gave back the Hirshfeld partitioning. This led us to explore what other sorts of nonlocal divergence measures would recover the Hirshfeld partitioning. This chapter reports the results of that exploration.

Here, we consider divergence measures that cannot be written as local functionals of the electron density. Specifically, we explore nonlocal functionals that are functions of local functionals discussed in chapter 2, i.e.,

$$G_{\text{nonlocal}}\left[\rho\right] = g\left(F^{(1)}_{\text{local}}\left[\rho\right], F^{(2)}_{\text{local}}\left[\rho\right], \ldots\right) \tag{3.1}$$

Of particular interest are divergence measures that are based on non-extensive functionals for the thermodynamic entropy. (Entropy functionals which are nonlocal are inherently non-extensive.) Extending the results of chapter 2, we show that a more general family of divergences, which are closely related to the α-divergence, gives rise to Hirshfeld atoms as well. This is the first time the Hirshfeld partitioning has been obtained from nonlocal divergence functionals.

The nonlocal functionals denoted in Eq. (3.1) are not *obviously f*-divergences. To assess whether or not these recover Hirshfeld partitioning, specifically, we consider the directed divergence measures associated with the Tsallis divergence,[118, 144-145]

$$I_T^{\alpha}\left[\{\rho_A\}\middle|\{\rho_A^0\}\right]=\frac{1}{\alpha-1}\left[\int\sum_{A=1}^{N_{\text{atoms}}}\rho_A(\mathbf{r})\left(\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})}\right)^{\alpha-1}d\mathbf{r}-\int\sum_{A=1}^{N_{\text{atoms}}}\rho_A(\mathbf{r})\,d\mathbf{r}\right] \quad (3.2)$$

the Réyni divergence,[145-147]

$$I_R^{\alpha}\left[\{\rho_A\}\middle|\{\rho_A^0\}\right]=\frac{1}{\alpha-1}\ln\left(\frac{\int\sum_{A}^{N_{\text{atoms}}}\left(\rho_A(\mathbf{r})\left(\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})}\right)^{\alpha-1}\right)d\mathbf{r}}{\int\sum_{A}^{N_{\text{atoms}}}\rho_A(\mathbf{r})\,d\mathbf{r}}\right) \quad (3.3)$$

the Sharma-Mittal divergence,[148-151]

$$I_{SM}^{\alpha,\beta}\left[\{\rho_A\}\middle|\{\rho_A^0\}\right]=\frac{1}{\beta-1}\left[\left(\int\sum_{A=1}^{N_{\text{atoms}}}\rho_A(\mathbf{r})\left(\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})}\right)^{\alpha-1}d\mathbf{r}\right)^{\frac{\beta-1}{\alpha-1}}-\left(\int\sum_{A=1}^{N_{\text{atoms}}}\rho_A(\mathbf{r})\,d\mathbf{r}\right)^{\frac{\beta-1}{\alpha-1}}\right] \quad (3.4)$$

a recently proposed supraextensive divergence,[152]

$$I_{SE}^{\alpha,\beta}\left[\{\rho_A\}\middle|\{\rho_A^0\}\right]=\frac{1}{\alpha-1}\left[\left(\int\sum_{A=1}^{N_{\text{atoms}}}\rho_A(\mathbf{r})\,d\mathbf{r}+\frac{\beta-1}{\alpha-1}\ln\left(\frac{\int\sum_{A=1}^{N_{\text{atoms}}}\rho_A(\mathbf{r})\left(\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})}\right)^{\alpha-1}d\mathbf{r}}{\int\sum_{A=1}^{N_{\text{atoms}}}\rho_A(\mathbf{r})\,d\mathbf{r}}\right)\right)^{\frac{\alpha-1}{\beta-1}}-\left(\int\sum_{A=1}^{N_{\text{atoms}}}\rho_A(\mathbf{r})\,d\mathbf{r}\right)^{\frac{\alpha-1}{\beta-1}}\right] \quad (3.5)$$

and the very general family of *H*-divergences,[153]

$$H_{h,\varphi_1,\varphi_2}\left[\{\rho_A\}\middle|\{\rho_A^0\}\right] = h\left(\frac{\int \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r})\varphi_1\left(\frac{\rho_A^0(\mathbf{r})}{\rho_A(\mathbf{r})}\right)d\mathbf{r}}{\int \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r})\varphi_2\left(\frac{\rho_A^0(\mathbf{r})}{\rho_A(\mathbf{r})}\right)d\mathbf{r}}\right) \tag{3.6}$$

These *H*-divergences are not a valid divergence measure for every choice for the $\varphi_1(x)$, $\varphi_2(x)$, and $h(x)$ functions. It suffices, however, for $\varphi_1(x)$ to be convex with $\varphi_1(1)=0$ (as for an *f*-divergence), $\varphi_2(x)>0$, and $h(x)$ to be monotonic, $h'(x)>0$, and $h(0)=0$.

There are further extensions (e.g., corresponding to position-dependent values, $\alpha(\mathbf{r})$, for the parameter in Tsallis divergence)[153-154] but we choose not to explore those generalizations here. We also omit consideration of divergence measures that are invariant to coordinate rotations (e.g., the total Bregman divergence).[155-157] Finally, we note that divergence measures in Eqs. (3.2)-(3.6) are slightly different from the usual form of these divergence measures. This revision is needed because atomic electron densities are normalized to the number of electrons, while the traditional divergence measures only apply to probability distribution functions that are normalized to one.

The Tsallis divergence is known to be an *f*-divergence and, in particular, is closely related to the special type of *f*-divergences called the α-divergences,[28, 138-140, 145, 158]

$$\begin{aligned} I_f^{\alpha}\left[\{\rho_A\}\middle|\{\rho_A^0\}\right] &= \int \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r})\left(\left(\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})}\right)^{\alpha-1} - 1\right)d\mathbf{r} \\ &= \int \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r})\left(\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})}\right)^{\alpha-1} d\mathbf{r} - N_{\text{mol}} \end{aligned} \tag{3.7}$$

where

$$N_{\mathrm{mol}} = \int \rho_{\mathrm{mol}}(\mathbf{r})\, d\mathbf{r} = \int \sum_{A=1}^{N_{\mathrm{atoms}}} \rho_A(\mathbf{r})\, d\mathbf{r} \tag{3.8}$$

is the number of electrons in the molecule. For convenience, we have chosen a different normalization of the α-divergence from the usual form. While we regard Eq. (3.7) as merely a notational convenience, we note that in the absence of prefactors, I_f^α is not a valid divergence measure for $0 \le \alpha \le 1$, because it is not convex.

Specifically, the Tsallis divergence is proportional to the α-divergence[145]

$$I_T^\alpha = \frac{I_f^\alpha}{\alpha - 1}, \tag{3.9}$$

Similarly, the Réyni divergence can be written as

$$I_R^\alpha = \frac{1}{\alpha - 1} \ln\left(1 + \frac{I_f^\alpha}{N_{\mathrm{mol}}}\right). \tag{3.10}$$

The α-divergence is also closely related to the Sharma-Mittal divergence,

$$I_{SM}^{\alpha,\beta} = \frac{N_{\mathrm{mol}}^{\frac{\beta-1}{\alpha-1}}}{\beta - 1}\left[\left(1 + \frac{I_f^\alpha}{N_{\mathrm{mol}}}\right)^{\frac{\beta-1}{\alpha-1}} - 1\right], \tag{3.11}$$

and the supraextensive divergence,

$$I_{SE}^{\alpha,\beta} = \frac{N_{\mathrm{mol}}^{\frac{\alpha-1}{\beta-1}}}{\alpha - 1}\left[\left(1 + \frac{\beta - 1}{N_{\mathrm{mol}}(\alpha - 1)} \ln\left(1 + \frac{I_f^\alpha}{N_{\mathrm{mol}}}\right)\right)^{\frac{\alpha-1}{\beta-1}} - 1\right]. \tag{3.12}$$

Notice that the Réyni, Sharma-Mittal, and supraextensive divergences are functions of local functionals (cf. Eq. (3.1)). They are therefore nonlocal density functionals, not *f*-divergences.

3.2 Non-Extensive Entropy Measures

Suppose one is given an information loss function that has the general form,

$$I_{\text{gen}}^{\alpha}\left[\left\{\rho_A\right\}\middle|\left\{\rho_A^0\right\}\right]=g\left(N_{\text{mol}},I_f^{\alpha}\right) \tag{3.13}$$

This form clearly encompasses and generalizes the Tsallis, Réyni, Sharma-Mittal, and supraextensive divergence measures. We then determine the atoms in molecule by the usual procedure,

$$\underbrace{\min}_{\left\{\rho_A(\mathbf{r})\middle|\sum_{A=1}^{N_{\text{atoms}}}\rho_A(\mathbf{r})=\rho_{\text{mol}}(\mathbf{r})\right\}} I_{\text{gen}}^{\alpha}\left[\left\{\rho_A\right\}\middle|\left\{\rho_A^0\right\}\right] \tag{3.14}$$

Introducing the constraint with a Lagrange multiplier, the Lagrangian is,

$$\Lambda\left[\left\{\rho_A\right\}\right]=g\left(N_{\text{mol}},I_f^{\alpha}\right)-\int\lambda(\mathbf{r})\left(\sum_{A=1}^{N_{\text{atoms}}}\rho_A(\mathbf{r})-\rho(\mathbf{r})\right)d\mathbf{r} \tag{3.15}$$

and the stationary condition for the minimum is,

$$0=\frac{\delta\Lambda}{\delta\rho_B(\mathbf{r})}=\frac{\partial g}{\partial N_{\text{mol}}}\frac{\delta N_{\text{mol}}}{\delta\rho_B(\mathbf{r})}+\frac{\partial g}{\partial I_f^{\alpha}}\frac{\delta I_f^{\alpha}}{\delta\rho_B(\mathbf{r})}-\lambda(\mathbf{r}) \tag{3.16}$$

where

$$\frac{\delta N_{\text{mol}}}{\delta\rho_B(\mathbf{r})}=1 \tag{3.17}$$

$$\frac{\delta I_f^{\alpha}}{\delta\rho_B(\mathbf{r})}=(\alpha-1)\left(\frac{\rho_B(\mathbf{r})}{\rho_B^0(\mathbf{r})}\right)^{\alpha-1} \tag{3.18}$$

The equation can then be written

$$\lambda(\mathbf{r}) - \frac{\partial g}{\partial N_{\text{mol}}} = (\alpha - 1)\frac{\partial g}{\partial I_f^\alpha}\left(\frac{\rho_B(\mathbf{r})}{\rho_B^0(\mathbf{r})}\right)^{\alpha-1} \tag{3.19}$$

As long as $\alpha \neq 1$ and $\partial g / \partial I_f^\alpha \neq 0$, this identity gives the key relation from which the Hirshfeld atom is derived, namely that $\rho_B(\mathbf{r})/\rho_B^0(\mathbf{r})$ is the same for all atoms. (For example, it is sufficient to have a strictly monotonic $g\left(N_{\text{mol}}, I_f^\alpha\right)$ with respect to $I_f^\alpha > 0$.) For the Tsallis, Rényi, and Sharma-Mittal divergences,

$$\frac{\partial g_T}{\partial I_f^\alpha} = \frac{1}{\alpha - 1} \neq 0 \tag{3.20}$$

$$\frac{\partial g_R}{\partial I_f^\alpha} = \frac{1}{(\alpha - 1)\left(N_{\text{mol}} + I_f^\alpha\right)} \neq 0 \tag{3.21}$$

$$\frac{\partial g_{SM}}{\partial I_f^\alpha} = \frac{N_{\text{mol}}^{\frac{\beta-\alpha}{\alpha-1}}}{\alpha - 1}\left[\left(1 + \frac{I_f^\alpha}{N_{\text{mol}}}\right)^{\frac{\beta-\alpha}{\alpha-1}}\right] \neq 0 \tag{3.22}$$

For the supraextensive entropy,

$$\frac{\partial g_{SE}}{\partial I_f^\alpha} = \frac{N_{\text{mol}}^{\frac{\alpha-\beta}{\beta-1}}}{\beta - 1}\left[\left(N_{\text{mol}} + \frac{\beta - 1}{N_{\text{mol}}(\alpha - 1)}\ln\left(1 + \frac{I_f^\alpha}{N_{\text{mol}}}\right)\right)^{\frac{\alpha-\beta}{\beta-1}}\right]\left(\frac{\beta - 1}{(\alpha - 1)\left(N_{\text{mol}} + I_f^\alpha\right)}\right) \tag{3.23}$$

Since this expression cannot be equal to zero, one must have $\beta \neq 1$. In all these expressions, we have used the fact that $I_f^\alpha \geq 0$, which presumes that the sum of the atomic densities and the sum of the reference pro-atomic densities have the same normalization.

For local divergence functionals, one sometimes uses the fact that the densities of the reference pro-atoms, $\left\{\rho_A^0(\mathbf{r})\right\}_{A=1}^{N_{\text{atoms}}}$, can be optimized to make the density of the so-called promolecule,

$$\rho_{\text{mol}}^0(\mathbf{r}) = \sum_{A=1}^{N_{\text{atoms}}} \rho_A^0(\mathbf{r}) \tag{3.24}$$

as close as possible to the density of the molecule, $\rho_{\text{mol}}(\mathbf{r})$.[28, 82] (This can remove the ambiguity associated with picking the reference pro-atoms.) This can also be done for these measures. To see this, notice that the key Hirshfeld criterion,

$$\frac{\rho_B(\mathbf{r})}{\rho_B^0(\mathbf{r})} = h(\mathbf{r}), \tag{3.25}$$

for some function $h(\mathbf{r})$, can be written as

$$\sum_{B=1}^{N_{\text{atoms}}} \rho_B(\mathbf{r}) = h(\mathbf{r}) \sum_{B=1}^{N_{\text{atoms}}} \rho_B^0(\mathbf{r}) \tag{3.26}$$

Therefore

$$h(\mathbf{r}) = \frac{\rho_B(\mathbf{r})}{\rho_B^0(\mathbf{r})} = \frac{\rho_{\text{mol}}(\mathbf{r})}{\rho_{\text{mol}}^0(\mathbf{r})} \tag{3.27}$$

and Eq. (3.7) can be rewritten as

$$I_f^\alpha\left[\rho_{\text{mol}}\middle|\rho_{\text{mol}}^0\right] = \int \rho_{\text{mol}}(\mathbf{r})\left(\frac{\rho_{\text{mol}}(\mathbf{r})}{\rho_{\text{mol}}^0(\mathbf{r})}\right)^{\alpha-1} d\mathbf{r} - N_{\text{mol}} = I_f^\alpha\left[\left\{\rho_A\right\}\middle|\left\{\rho_A^0\right\}\right] \tag{3.28}$$

where in Eqs. (3.26) and (3.28) we have used the constraint that the atom-in-molecule densities add up to the total molecular density. The pro-molecule density can therefore be optimized by the two-step procedure,

$$\underbrace{\min}_{\left\{\rho_A^0(\mathbf{r})\right\}}\ \underbrace{\min}_{\left\{\rho_A(\mathbf{r})\left|\sum_{A=1}^{N_{\text{atoms}}}\rho_A(\mathbf{r})=\rho_{\text{mol}}(\mathbf{r})\right.\right\}} I_f^\alpha\left[\left\{\rho_A\right\}\middle|\left\{\rho_A^0\right\}\right] = \underbrace{\min}_{\left\{\rho_A^0(\mathbf{r})\right\}} I_f^\alpha\left[\rho_{\text{mol}}\middle|\sum_{A=1}^{N_{\text{atoms}}}\rho_A^0\right] \qquad (3.29)$$

The identity (3.28) and the strategy in Eq. (3.29) clearly extend to any of the generalized α-divergences in this paper. While these formulas generalize the *f*-divergences considered in chapter 2 somewhat, they do not contradict the results in that paper because these divergences are not local functionals of the electron density.[28] Their generalizations are also not very consequential, since one still obtains the Hirshfeld partitioning. However, while the Tsallis and Réyni divergences give the same pro-atoms (because both objective functions are minimized when $I_f^\alpha\left[\rho_{\text{mol}}\middle|\rho_{\text{mol}}^0\right]$ is made as small as possible), this is not necessarily true for the Sharma-Mittal and supraextensive divergences.

3.3 *H*-Divergences

The divergence measures we considered in the previous section are all based on non-extensive entropy formulas. The *H*-divergence formula in Eq. (3.6) generalizes these equations as well as the *f*-divergence. For example, the *H*-divergence is an *f*-divergence (up to a choice of normalization) if $\varphi_1(1)=0$, $\varphi_1(x)$ is convex, $\varphi_2(x)=1$, and $h(x)=x$.

As mentioned before, not every choice of functions in Eq. (3.6) is allowed. In this chapter, we consider only *H*-divergences which satisfy the requirements:

- $h(x)$ is monotonically increasing, $h'(x)>0$, and $h(0)=0$.
- $\varphi_1(x)$ is convex, $\varphi_1''(x)>0$, and $\varphi_1(1)=0$.

- $\varphi_2(x) > 0$.

This gives $H_{h,\varphi_1,\varphi_2}\left[\left\{\rho_A\right\}\middle|\left\{\rho_A^0\right\}\right] \geq H_{h,\varphi_1,\varphi_2}\left[\left\{\rho_A^0\right\}\middle|\left\{\rho_A^0\right\}\right] = 0$, for the densities with the same normalization, which is one of the essential properties of a divergence measure. The analogous *H*-divergence derivation of the Hirshfeld atoms-in-molecules partitioning is found by minimizing

$$\underbrace{\min}_{\left\{\rho_A(\mathbf{r}) \middle| \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r}) = \rho_{\text{mol}}(\mathbf{r})\right\}} H_{h,\varphi_1,\varphi_2}\left[\left\{\rho_A\right\}\middle|\left\{\rho_A^0\right\}\right] \tag{3.30}$$

with the solution

$$\lambda(\mathbf{r}) = \frac{1}{G_{\varphi_2}^2} \cdot h'\left(\frac{G_{\varphi_1}}{G_{\varphi_2}}\right) \cdot \left[G_{\varphi_2} \frac{\delta G_{\varphi_1}}{\delta \rho_B(\mathbf{r})} - G_{\varphi_1} \frac{\delta G_{\varphi_2}}{\delta \rho_B(\mathbf{r})}\right] \tag{3.31}$$

where we have defined the convenient notation,

$$\begin{aligned} G_{\varphi_1} &= \int \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r}) \varphi_1\left(\frac{\rho_A^0(\mathbf{r})}{\rho_A(\mathbf{r})}\right) d\mathbf{r} \\ G_{\varphi_2} &= \int \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r}) \varphi_2\left(\frac{\rho_A^0(\mathbf{r})}{\rho_A(\mathbf{r})}\right) d\mathbf{r} \end{aligned}, \tag{3.32}$$

Note that by requiring that $\varphi_1(x)$ is a convex function with $\varphi_1(1) = 0$, we ensure that G_{φ_1} is an *f*-divergence. G_{φ_2} is not an *f*-divergence, but a type of normalization factor. Possible choices include $\varphi_2(x) = x^\alpha$ $(0 \leq \alpha \leq 1)$, $\varphi_2(x) = x/(x+1)$, $\varphi_2(x) = \ln(x+1)$, $\varphi_2(x) = \tanh(x)$. All of these functions are concave for $x \geq 0$, $\varphi_2''(x) < 0$. This is not required for *H*-divergence to be a valid divergence measures, but later it will turn out to be useful.

Inserting the functional derivatives,

$$\begin{aligned}\frac{\delta G_{\varphi_1}}{\delta \rho_B(\mathbf{r})} &= \varphi_1\left(\frac{\rho_B^0(\mathbf{r})}{\rho_B(\mathbf{r})}\right) - \frac{\rho_B^0(\mathbf{r})}{\rho_B(\mathbf{r})}\varphi_1'\left(\frac{\rho_B^0(\mathbf{r})}{\rho_B(\mathbf{r})}\right) \\ \frac{\delta G_{\varphi_2}}{\delta \rho_B(\mathbf{r})} &= \varphi_2\left(\frac{\rho_B^0(\mathbf{r})}{\rho_B(\mathbf{r})}\right) - \frac{\rho_B^0(\mathbf{r})}{\rho_B(\mathbf{r})}\varphi_2'\left(\frac{\rho_B^0(\mathbf{r})}{\rho_B(\mathbf{r})}\right)\end{aligned} \tag{3.33}$$

into Eq. (3.31), we obtain the expression

$$\lambda(\mathbf{r}) = h'\left(\frac{G_{\varphi_1}}{G_{\varphi_2}}\right)\begin{pmatrix} \frac{1}{G_{\varphi_2}}\cdot\left(\varphi_1\left(\frac{\rho_B(\mathbf{r})}{\rho_B^0(\mathbf{r})}\right) - \frac{\rho_B(\mathbf{r})}{\rho_B^0(\mathbf{r})}\varphi_1'\left(\frac{\rho_B(\mathbf{r})}{\rho_B^0(\mathbf{r})}\right)\right) \\ -\frac{G_{\varphi_1}}{G_{\varphi_2}^2}\cdot\left(\varphi_2\left(\frac{\rho_B(\mathbf{r})}{\rho_B^0(\mathbf{r})}\right) - \frac{\rho_B(\mathbf{r})}{\rho_B^0(\mathbf{r})}\varphi_2'\left(\frac{\rho_B(\mathbf{r})}{\rho_B^0(\mathbf{r})}\right)\right) \end{pmatrix} \tag{3.34}$$

Eq. (3.25), which leads to the Hirshfeld partitioning, is *a* solution to this equation. However, it may not be the only solution. In general, Eq. (3.34) gives an equation relating the densities of every atom-pair in the molecule,

$$g\left(\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})}\right) = g\left(\frac{\rho_B(\mathbf{r})}{\rho_B^0(\mathbf{r})}\right) \tag{3.35}$$

where

$$g(x) = \frac{1}{G_{\varphi_2}}\left(\varphi_1(x) - x\varphi_1'(x)\right) - \frac{G_{\varphi_1}}{G_{\varphi_2}^2}\left(\varphi_2(x) - x\varphi_2'(x)\right). \tag{3.36}$$

If $g(x)$ is invertible for $x \geq 0$, then the unique solution to Eq. (3.35) is

$$\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} = \frac{\rho_B(\mathbf{r})}{\rho_B^0(\mathbf{r})}, \tag{3.37}$$

which leads to the Hirshfeld partitioning. If we assume that all the functions are at least twice-differentiable, it is sufficient that $g(x)$ be monotonic. Therefore, for $x > 0$,

$$g'(x) = \frac{x}{G_{\varphi_2}^2}\left(G_{\varphi_2}\varphi_1''(x) - G_{\varphi_1}\varphi_2''(x)\right) > 0 \tag{3.38}$$

The conditions stipulated at the beginning of this section are almost sufficient to satisfy this equation because they ensure that G_{φ_2} is positive, that $\varphi_1''(x)$ are positive, and that G_{φ_1} is nonnegative. If we further require $\varphi_2''(x)$ to be nonpositive, then the Hirshfeld partitioning is the unique solution to the variational procedure (3.30). These conditions also suffice to derive the analogue of the identity in Eq. (3.28), namely that for the atom-in-molecule densities obtained from Eq. (3.30),

$$I_{h,\varphi_1,\varphi_2}\left[\{\rho_A\}\middle|\{\rho_A^0\}\right] = I_{h,\varphi_1,\varphi_2}\left[\rho_{\text{mol}}\middle|\rho_{\text{mol}}^0\right] = h\left(\frac{\int \rho_{\text{mol}}(\mathbf{r})\,\varphi_1\left(\dfrac{\rho_{\text{mol}}^0}{\rho_{\text{mol}}(\mathbf{r})}\right)d\mathbf{r}}{\int \rho_{\text{mol}}(\mathbf{r})\,\varphi_2\left(\dfrac{\rho_{\text{mol}}^0}{\rho_{\text{mol}}(\mathbf{r})}\right)d\mathbf{r}}\right). \tag{3.39}$$

3.4 Conclusions

In this chapter, extending on the results of chapter 2, we have shown that several nonlocal divergence measures, like Réyni, Sharma-Mittal, and supraextensive divergence measures, all lead to the Hirshfeld partitioning. These functionals are very closely linked to the α-divergence. This is desirable in the sense that it ensures that these measures are closely linked to a very popular and useful family of *f*-divergence measures, but it is undesirable insofar as it means that optimizing the pro-atom does not give significantly different results for these approaches.

The *H*-divergence in represents a much more general class of measures. While it is difficult to find necessary conditions for the *H*-divergence that gives the Hirshfeld atom, it is sufficient to require the following properties for $x > 0$:

- $h(x)$ is monotonically increasing, $h'(x) > 0$. Also $h(0) = 0$.
- $\varphi_1(x)$ is convex, $\varphi_1''(x) > 0$, and $\varphi_1(1) = 0$. (same requirements as for an *f*-divergence)
- $\varphi_2(x) > 0$ and is nonconvex, $\varphi_2''(x) \leq 0$.

Note that this family of *H*-divergences is closely related to the *f*-divergences, but extends that set in a nontrivial way.

4 Theory of Variational Hirshfeld Partitioning

4.1 Background

In this chapter, we explore the freedom in selecting the pro-atoms in the Hirshfeld partitioning scheme by introducing the very simple and flexible additive pro-piece model. As mentioned in the introduction, the Hirshfeld definition can be extended beyond atoms and functional groups to include any suitably defined molecular components. For example, in the minimal basis iterative stockholder (MBIS) partitioning, the components are atomic shells.[82]. The additive pro-piece model generalizes the MBIS approach for defining pro-density, and has many similarities with other extensions of the original Hirshfeld approach, e.g. the iterative Hirshfeld method,[46, 60] the extended Hirshfeld method,[73, 159] variational Hirshfeld-I, etc.[31-33, 58, 72, 81], but has better mathematical properties. This model is combined with the theoretical framework laid in previous chapters to introduce the Additive Variational Hirshfeld (AVH) method, which variationally optimizes both the atoms and protoms and has desirable mathematical and chemical properties.

4.2 Mathematical Formulation

Instead of approximating the molecular electron density as a sum of pro-atoms, as is typically done in the Hirshfeld family of methods, let us approximate the electron

density as a sum of nonnegative basis states, $\{b_k(\boldsymbol{\alpha}_k;\mathbf{r})\}_{k=1}^{K}$, which for convenience we choose to be normalized to one,

$$0 \le b_k(\boldsymbol{\alpha}_k;\mathbf{r}). \tag{4.1}$$

$$1 = \int b_k(\boldsymbol{\alpha}_k;\mathbf{r})\,d\mathbf{r} \tag{4.2}$$

Here $\boldsymbol{\alpha}_k$ is an optional vector of parameters that can be optimized to improve the quality of the basis. The molecular electron density is approximately expanded as a linear combination of the basis states,

$$\rho_{\text{mol}}(\mathbf{r}) \approx \sum_{k=1}^{K} c_k b_k(\boldsymbol{\alpha}_k;\mathbf{r}) \equiv \rho_{\text{mol}}^{0}\left(\{c_k,\boldsymbol{\alpha}_k\};\mathbf{r}\right) \tag{4.3}$$

As is typical, this approximation of the molecular density is called the pro-molecular density, which in this case, is the sum of the reference densities of the individual pieces called the pro-pieces,

$$\rho_k^0(c_k,\boldsymbol{\alpha}_k;\mathbf{r}) \equiv c_k b_k(\boldsymbol{\alpha}_k;\mathbf{r}) \tag{4.4}$$

We wish for the pro-molecular density to resemble the molecular density as strongly as possible, subject to the constraint that the pro-molecular density is normalized to the total number of electrons,

$$N_{\text{mol}} = \int \rho_{\text{mol}}(\mathbf{r})\,d\mathbf{r} = \int \rho_{\text{mol}}^{0}\left(\{c_k,\boldsymbol{\alpha}_k\};\mathbf{r}\right)d\mathbf{r} = \sum_{k=1}^{K} c_k\,. \tag{4.5}$$

Based on the theoretical framework established in chapter 2, we can find the best pro-pieces by minimizing the deviation between molecular and pro-molecular density.

That is, using a directed *f*-divergence to measure this similarity, we can use the variational principle to find the pro-pieces,

$$\underbrace{\min}_{\left\{c_k,\boldsymbol{\alpha}_k \middle| N=\sum_{k=1}^{K} c_k\right\}} \int \rho_{\text{mol}}(\mathbf{r}) f\left(\frac{\rho_{\text{mol}}^0\left(\{c_k,\boldsymbol{\alpha}_k\};\mathbf{r}\right)}{\rho_{\text{mol}}(\mathbf{r})}\right) d\mathbf{r} \tag{4.6}$$

where f is any convex function with $f(1)=0$.[28] The constraint that the pro-molecular density and the molecular density have the same number of electrons is required not only for chemical sensibility but also mathematically: otherwise Eq. (4.6) might not be a valid divergence measure.

We would also like for the pieces of the molecule, $\rho_k(\mathbf{r})$, to resemble as closely as possible the pro-pieces. This can also be expressed using variational principle,

$$\underbrace{\min}_{\left\{c_k,\boldsymbol{\alpha}_k \middle| \rho_{\text{mol}}(\mathbf{r})=\sum_{k=1}^{K}\int \rho_k(\mathbf{r})\, d\mathbf{r}\right\}} \underbrace{\min}_{\rho_k(\mathbf{r})} \sum_{k=1}^{K} \int \rho_k(\mathbf{r}) f\left(\frac{\rho_k^0\left(c_k,\boldsymbol{\alpha}_k;\mathbf{r}\right)}{\rho_k(\mathbf{r})}\right) d\mathbf{r} \tag{4.7}$$

where the constraint can be simplified into $N=\sum_{k=1}^{K} c_k$. Because the minimization of molecular piece and pro-piece are not coupled, the variational principle in Eq. (4.7) is equivalent to Eq. (4.6), and the densities of the molecular pieces are defined by the Hirshfeld, or stockholder, partitioning,[21, 28]

$$\rho_k(\mathbf{r}) = \frac{\rho_k^0\left(c_k,\boldsymbol{\alpha}_k;\mathbf{r}\right)}{\rho_{\text{mol}}^0\left(\{c_k,\boldsymbol{\alpha}_k\};\mathbf{r}\right)} \rho_{\text{mol}}(\mathbf{r}) \tag{4.8}$$

Note that while we derived this using a divergence directed from the pro-density to the density, as elaborated in chapter 2, this treatment includes the alternative cases where the

roles of the density and pro-density are interchanged (consider $\tilde{f}(x) = xf(x^{-1})$) and where the density and pro-density are treated symmetrically (consider $\tilde{f}(x) = \frac{1}{2}\left(f(x) + xf(x^{-1})\right)$).[28] There is also no constraint guaranteeing that the molecular piece and pro-piece have the same number of electrons, i.e.

$$\int \rho_k(\mathbf{r})\, d\mathbf{r} = N_k = c_k = \int \rho_k^0(c_k, \alpha_k; \mathbf{r})\, d\mathbf{r} \tag{4.9}$$

This is a valid mathematical requirement, but having such a constraint makes the optimizations in Eq. (4.7) inseparable and does not result in conventional Hirshfeld scheme of Eq. (4.8).[72] As a result, we will especially focus on the family of *extended* divergence measures, $\tilde{f}(x) = f(x) - f'(1)(x-1)$, for which the constraint of Eq. (4.9) is not required.[160-163] Extended *f*-divergences satisfy $f(1) = f'(1) = 0$ and therefore have the desirable property that

$$\int \rho_k(\mathbf{r}) f\left(\frac{\rho_k^0(c_k, \boldsymbol{\alpha}_k; \mathbf{r})}{\rho_k(\mathbf{r})}\right) d\mathbf{r} \tag{4.10}$$

is a valid measure of the divergence of a molecular piece from its corresponding pro-piece even when the piece and the pro-piece are not normalized to the same number of electrons. Notice that $f(1) = f'(1) = 0$ is automatically satisfied for any symmetric *f*-divergence, so they are also plausible divergence measures for densities with different normalizations.

The parameters in the pro-molecule are determined by differentiating the Lagrangian corresponding to Eq. (4.6),

$$\mathrm{L}\left(\{c_k,\boldsymbol{\alpha}_k\},\mu\right)=\int\rho_{\mathrm{mol}}(\mathbf{r})f\left(\frac{\sum_{k=1}^{K}c_k b_k(\boldsymbol{\alpha}_k;\mathbf{r})}{\rho_{\mathrm{mol}}(\mathbf{r})}\right)d\mathbf{r}-\mu\left(\sum_{k=1}^{K}c_k-N_{\mathrm{mol}}\right) \quad (4.11)$$

giving a system of nonlinear equations,

$$0=\frac{\partial \mathrm{L}}{\partial c_k}=\int f'\left(\frac{\sum_{k=1}^{K}c_k b_k(\boldsymbol{\alpha}_k;\mathbf{r})}{\rho_{\mathrm{mol}}(\mathbf{r})}\right)b_k(\boldsymbol{\alpha}_k;\mathbf{r})\,d\mathbf{r}-\mu \quad (4.12)$$

$$0=\frac{\partial \mathrm{L}}{\partial \boldsymbol{\alpha}_k}=\int f'\left(\frac{\sum_{k=1}^{K}c_k b_k(\boldsymbol{\alpha}_k;\mathbf{r})}{\rho_{\mathrm{mol}}(\mathbf{r})}\right)c_k\frac{\partial b_k(\boldsymbol{\alpha}_k;\mathbf{r})}{\partial \boldsymbol{\alpha}_k}d\mathbf{r} \quad (4.13)$$

To solve for the Lagrange multiplier, we multiply Eq. (4.12) by c_k and sum over k to obtain

$$\mu=\frac{1}{N_{\mathrm{mol}}}\int\rho^0_{\mathrm{mol}}\left(\{c_k,\boldsymbol{\alpha}_k\};\mathbf{r}\right)f'\left(\frac{\rho^0_{\mathrm{mol}}\left(\{c_k,\boldsymbol{\alpha}_k\};\mathbf{r}\right)}{\rho_{\mathrm{mol}}(\mathbf{r})}\right)d\mathbf{r} \quad (4.14)$$

Eqs. (4.12) and (4.13) can then be rewritten as

$$\mu=\int b_k(\boldsymbol{\alpha};\mathbf{r})f'\left(\frac{\rho^0_{\mathrm{mol}}\left(\{c_k,\boldsymbol{\alpha}_k\};\mathbf{r}\right)}{\rho_{\mathrm{mol}}(\mathbf{r})}\right)d\mathbf{r} \quad (4.15)$$

$$0=\int\frac{\partial b_k(\boldsymbol{\alpha};\mathbf{r})}{\partial\boldsymbol{\alpha}}f'\left(\frac{\rho^0_{\mathrm{mol}}\left(\{c_k,\boldsymbol{\alpha}_k\};\mathbf{r}\right)}{\rho_{\mathrm{mol}}(\mathbf{r})}\right)d\mathbf{r} \quad (4.16)$$

In deriving the second equation we assumed that $c_k\neq 0$.

There can be many local minima in Eq. (4.6), and so there will usually be multiple solutions to these equations. However, in the absence of nonlinear parameters $\{\boldsymbol{\alpha}_k\}$, the second derivative of the Lagrangian is

$$\begin{aligned}\frac{\partial^2 \mathcal{L}}{\partial c_k \, \partial c_l} &= \int \frac{1}{\rho_{\mathrm{mol}}(\mathbf{r})} f''\left(\frac{\rho^0_{\mathrm{mol}}(\{c_k\};\mathbf{r})}{\rho_{\mathrm{mol}}(\mathbf{r})} \right) b_k(\mathbf{r}) b_l(\mathbf{r}) \, d\mathbf{r} \\ &= \iint b_k(\mathbf{r}) \left[\frac{1}{\rho_{\mathrm{mol}}(\mathbf{r})} f''\left(\frac{\rho^0_{\mathrm{mol}}(\{c_k\};\mathbf{r})}{\rho_{\mathrm{mol}}(\mathbf{r})} \right) \delta(\mathbf{r}-\mathbf{r}') \right] b_l(\mathbf{r}') \, d\mathbf{r} \, d\mathbf{r}'\end{aligned} \tag{4.17}$$

The integral kernel in the second line is positive-definite because $f''(x) > 0$. In such cases, the objective function is convex and the variational principle has a unique solution *for a given value of* μ. One may then solve the equations by optimizing the coefficients (uniquely) and perform a subsequent one-dimensional search for the appropriate value of μ.

It is, unfortunately, difficult to generalize this argument to the case where there are nonlinear parameters. The corresponding blocks of the second derivative matrix are:

$$\begin{aligned}\frac{\partial^2 \mathcal{L}}{\partial c_k \, \partial \boldsymbol{\alpha}_l} &= \int \frac{1}{\rho_{\mathrm{mol}}(\mathbf{r})} f''\left(\frac{\rho^0_{\mathrm{mol}}(\{c_k\};\mathbf{r})}{\rho_{\mathrm{mol}}(\mathbf{r})} \right) b_k(\mathbf{r}) c_l \frac{\partial b_l(\mathbf{r})}{\partial \boldsymbol{\alpha}_l} \, d\mathbf{r} \\ &\quad + \delta_{kl} \int f'\left(\frac{\rho^0_{\mathrm{mol}}(\{c_k\};\mathbf{r})}{\rho_{\mathrm{mol}}(\mathbf{r})} \right) \frac{\partial b_l(\mathbf{r})}{\partial \boldsymbol{\alpha}_l} \, d\mathbf{r}\end{aligned} \tag{4.18}$$

$$\begin{aligned}\frac{\partial^2 \mathcal{L}}{\partial \boldsymbol{\alpha}_k \, \partial \boldsymbol{\alpha}_l} &= \int \frac{1}{\rho_{\mathrm{mol}}(\mathbf{r})} f''\left(\frac{\rho^0_{\mathrm{mol}}(\mathbf{r})}{\rho_{\mathrm{mol}}(\mathbf{r})} \right) c_k \frac{\partial b_k(\mathbf{r})}{\partial \boldsymbol{\alpha}_k} c_l \frac{\partial b_l(\mathbf{r})}{\partial \boldsymbol{\alpha}_l} \, d\mathbf{r} \\ &\quad + \delta_{kl} \int f'\left(\frac{\rho^0_{\mathrm{mol}}(\mathbf{r})}{\rho_{\mathrm{mol}}(\mathbf{r})} \right) \frac{\partial^2 b_l(\mathbf{r})}{\partial \boldsymbol{\alpha}_l^2} \, d\mathbf{r}\end{aligned} \tag{4.19}$$

The same argument for the convexity of the objective function would hold if the second terms in these equations (the $k = l$ terms) were not present. Notice, however, that the second term in Eq. (4.18) automatically vanishes if the gradient is zero. This means that if the second term in Eq. (4.19) is nonnegative, the second derivative is positive semidefinite at every critical point. Assuming sufficient differentiability, then, one may apply the Poincare-Hopf theorem to conclude that there can be only one minimum. (To have more than one minimum requires the presence of another critical point, typically a saddle point.)

However, it is not possible to ensure that the second term in Eq. (4.19) is always positive. One can argue for this mathematically, but it is intuitively obvious: given one local minimum, another local minimum can be found by permuting the basis functions. It seems difficult to determine whether there are additional minima beyond these trivial solutions. Note, however, that it is easy to verify whether one has discovered a local minimum (rather than a saddle point) using inequality (4.23). (This inequality is only sufficient, not necessary. If condition (4.23) is not satisfied, one might still have a local minimum, but verifying this requires evaluating the full second derivative of the Lagrangian, Eqs. (4.17)-(4.19).)

4.3 Extended Kullback-Leibler Divergence

As mentioned in the previous section, we specifically focus our attention on the extended Kullback-Leibler divergence measures, because they do not require the density and pro-density to have the same number of electrons, i.e. not extra constraints are needed

to make the variational principle in Eq. (4.7) mathematically plausible. In addition, the results from the previous section have an especially pleasing form in MBIS-style partitioning, where one chooses the extended Kullback-Leibler divergence, $f(x) = -\ln(x) + x - 1$, obtaining the Lagrangian

$$\mathrm{L}\left(\{c_k, \boldsymbol{\alpha}_k\}, \mu\right) = \int \rho_{\mathrm{mol}}(\mathbf{r}) \ln\left(\frac{\rho_{\mathrm{mol}}(\mathbf{r})}{\sum_{k=1}^{K} c_k b_k(\boldsymbol{\alpha}_k;\mathbf{r})}\right) + \sum_{k=1}^{K} c_k b_k(\boldsymbol{\alpha}_k;\mathbf{r}) - \rho_{\mathrm{mol}}(\mathbf{r})\, d\mathbf{r} - \mu\left(\sum_{k=1}^{K} c_k - N_{\mathrm{mol}}\right) \tag{4.20}$$

In this case $\mu = 0$ (cf. Eq. (4.14)), indicating that constraining the molecule and pro-molecule to have the same number of electrons is unnecessary: this constraint is already satisfied when the extended Kullback-Leibler divergence is used. The equations for deriving the parameters become (cf. Eqs. (4.15) and (4.16))

$$1 = \int b_k(\boldsymbol{\alpha};\mathbf{r}) \frac{\rho_{\mathrm{mol}}(\mathbf{r})}{\rho^0_{\mathrm{mol}}\left(\{c_k, \boldsymbol{\alpha}_k\};\mathbf{r}\right)} d\mathbf{r} \tag{4.21}$$

$$0 = \int \frac{\partial b_k(\boldsymbol{\alpha};\mathbf{r})}{\partial \boldsymbol{\alpha}} \left(1 - \frac{\rho_{\mathrm{mol}}(\mathbf{r})}{\rho^0_{\mathrm{mol}}\left(\{c_k, \boldsymbol{\alpha}_k\};\mathbf{r}\right)}\right) d\mathbf{r} \tag{4.22}$$

In deriving these equations we used the result $f'(x) = -x^{-1} + 1$, the normalization constraint in Eq. (4.2), and the result $\mu = 0$. The Lagrangian is convex if there are no nonlinear parameters in the basis function, or if the nonlinear parameters satisfy the equation

$$\int\left(1-\frac{\rho_{\text{mol}}(\mathbf{r})}{\rho_{\text{mol}}^{0}(\mathbf{r})}\right)\frac{\partial b_k(\boldsymbol{\alpha}_k;\mathbf{r})}{\partial \boldsymbol{\alpha}_k^2}d\mathbf{r}\geq 0 \tag{4.23}$$

While this is probably not true in general, this term will be small when the pro-molecular density is a good approximation to the molecular density. Under such circumstances, the (manifestly positive definite) first term in Eq. (4.19) is expected to be dominant.

A beautiful property that is specific to MBIS-like partitioning is that the pro-pieces and pieces of the molecule have the same normalization. This desirable feature is an outcome of using the (extended) Kullback-Leibler divergence. To see this, multiply both sides of Eq. (4.21) by c_k and use the normalization condition, (4.2), the definition of the pro-pieces, (4.4), and the result for the densities of the pieces of the molecule, (4.8), to conclude that

$$\begin{aligned} c_k &= \int c_k b_k(\boldsymbol{\alpha}_k;\mathbf{r})\frac{\rho_{\text{mol}}(\mathbf{r})}{\rho_{\text{mol}}^{0}(\mathbf{r})}d\mathbf{r} \\ \int \rho_k^0(\mathbf{r})d\mathbf{r} &= \int \rho_k^0(\mathbf{r})\frac{\rho_{\text{mol}}(\mathbf{r})}{\rho_{\text{mol}}^{0}(\mathbf{r})}d\mathbf{r} \\ \int \rho_k^0(\mathbf{r})d\mathbf{r} &= \int \rho_k(\mathbf{r})d\mathbf{r} \end{aligned} \tag{4.24}$$

It is remarkable that this result, which is ordinarily imposed in the Hirshfeld-E,[53] Hirshfeld-I,[60] and Hirshfeld-Iλ methods,[72] arises automatically here.

Notice, however, that this result does not hold for an arbitrary *f*-divergence and is specific to the (extended) Kullback-Leibler family. To show the class of *f*-divergences for which the pieces and pro-pieces have the same population, we reverse the argument in Eq. (4.24). That is, the pieces and pro-pieces have the same normalization only if Eq. (4.21) is true. Comparing Eqs. (4.15) and (4.21), we have

$$\mu - \int b_k(\boldsymbol{\alpha};\mathbf{r}) f'\left(\frac{\rho^0_{\mathrm{mol}}\left(\{c_k,\boldsymbol{\alpha}_k\};\mathbf{r}\right)}{\rho_{\mathrm{mol}}(\mathbf{r})} \right) d\mathbf{r} = 0 = 1 - \int b_k(\boldsymbol{\alpha};\mathbf{r}) \frac{\rho_{\mathrm{mol}}(\mathbf{r})}{\rho^0_{\mathrm{mol}}\left(\{c_k,\boldsymbol{\alpha}_k\};\mathbf{r}\right)} d\mathbf{r} \tag{4.25}$$

or, equivalently

$$\mu - 1 = \int b_k(\boldsymbol{\alpha};\mathbf{r}) \left[f'\left(\frac{\rho^0_{\mathrm{mol}}\left(\{c_k,\boldsymbol{\alpha}_k\};\mathbf{r}\right)}{\rho_{\mathrm{mol}}(\mathbf{r})} \right) - \frac{\rho_{\mathrm{mol}}(\mathbf{r})}{\rho^0_{\mathrm{mol}}\left(\{c_k,\boldsymbol{\alpha}_k\};\mathbf{r}\right)} \right] d\mathbf{r} \tag{4.26}$$

Notice that the left-hand-side of this equation is a constant for *any* choice of basis function. This requires that

$$f'\left(\frac{\rho^0_{\mathrm{mol}}\left(\{c_k,\boldsymbol{\alpha}_k\};\mathbf{r}\right)}{\rho_{\mathrm{mol}}(\mathbf{r})} \right) = a \frac{\rho_{\mathrm{mol}}(\mathbf{r})}{\rho^0_{\mathrm{mol}}\left(\{c_k,\boldsymbol{\alpha}_k\};\mathbf{r}\right)} + b \tag{4.27}$$

where a and b are constants. Therefore $f(x) = a\ln x + b(x-1)$. The requirement $f''(x) > 0$ corresponds to the requirement $a < 0$. For this general form, the Lagrange multiplier is (cf. Eq. (4.14))

$$\mu = \frac{1}{N^0_{\mathrm{mol}}} \int \rho^0_{\mathrm{mol}}\left(\{c_k,\boldsymbol{\alpha}_k\};\mathbf{r}\right) \left(a \frac{\rho_{\mathrm{mol}}(\mathbf{r})}{\rho^0_{\mathrm{mol}}\left(\{c_k,\boldsymbol{\alpha}_k\};\mathbf{r}\right)} + b \right) d\mathbf{r} = a + b \tag{4.28}$$

where in the last line we have employed the constraint $N_{\mathrm{mol}} = N^0_{\mathrm{mol}}$. Inserting the value of the Lagrange multiplier into Eq. (4.15), one has

$$a + b = \int b_k(\boldsymbol{\alpha};\mathbf{r}) \left(a \frac{\rho_{\mathrm{mol}}(\mathbf{r})}{\rho^0_{\mathrm{mol}}\left(\{c_k,\boldsymbol{\alpha}_k\};\mathbf{r}\right)} + b \right) d\mathbf{r} \tag{4.29}$$

which, using the normalization of the basis functions, simplifies to Eq. (4.21). The extended *f*-divergence corresponds to the choice $a = -1$ and $b = 1$.

4.4 Size-Consistency

A partitioning is size consistent if by performing the method for two molecules, A and B, that are infinitely separated, one obtains the same pieces and pro-pieces as when one treats the A and B separately. Presuming that the basis functions for the molecular pieces are local and therefore can be clearly assigned to one of the molecules, optimizing the super-system (i.e., A and B infinitely apart) corresponds to the Lagrangian,

$$\mathcal{L}_{\mathrm{A}\cdots\mathrm{B}}\left(\left\{c_k,\boldsymbol{\alpha}_k\right\},\mu\right)=\int\rho_{\mathrm{A}}\left(\mathbf{r}\right)f\left(\frac{\sum_{k\in\mathrm{A}}c_k b_k\left(\boldsymbol{\alpha}_k;\mathbf{r}\right)}{\rho_{\mathrm{A}}\left(\mathbf{r}\right)}\right)d\mathbf{r}+\int\rho_{\mathrm{B}}\left(\mathbf{r}\right)f\left(\frac{\sum_{k\in\mathrm{B}}c_k b_k\left(\boldsymbol{\alpha}_k;\mathbf{r}\right)}{\rho_{\mathrm{B}}\left(\mathbf{r}\right)}\right)d\mathbf{r} \\ -\mu_{\mathrm{A}\cdots\mathrm{B}}\left(\sum_{k\in\mathrm{A}}c_k+\sum_{k\in\mathrm{B}}c_k-N_{\mathrm{A}}-N_{\mathrm{B}}\right) \tag{4.30}$$

while optimizing the subsystems separately corresponds to the Lagrangian,

$$\mathcal{L}_{\mathrm{A}}\left(\left\{c_k,\boldsymbol{\alpha}_k\right\},\mu\right)+\mathcal{L}_{\mathrm{B}}\left(\left\{c_k,\boldsymbol{\alpha}_k\right\},\mu\right)=\int\rho_{\mathrm{A}}\left(\mathbf{r}\right)f\left(\frac{\sum_{k\in\mathrm{A}}c_k b_k\left(\boldsymbol{\alpha}_k;\mathbf{r}\right)}{\rho_{\mathrm{A}}\left(\mathbf{r}\right)}\right)d\mathbf{r}-\mu_{\mathrm{A}}\left(\sum_{k\in\mathrm{A}}c_k-N_{\mathrm{A}}\right) \\ +\int\rho_{\mathrm{B}}\left(\mathbf{r}\right)f\left(\frac{\sum_{k\in\mathrm{B}}c_k b_k\left(\boldsymbol{\alpha}_k;\mathbf{r}\right)}{\rho_{\mathrm{B}}\left(\mathbf{r}\right)}\right)d\mathbf{r}-\mu_{\mathrm{B}}\left(\sum_{k\in\mathrm{B}}c_k-N_{\mathrm{B}}\right) \tag{4.31}$$

The two optimizations will give different results unless the Lagrange multiplier is a constant that is independent of the system. The only *f*-divergences that satisfy this constraint are the (extended) Kullback-Leibler family considered in previous section, $f\left(x\right)=a\ln x+b\left(x-1\right)$ with $a<0$. For these divergences, $\mu=a+b$.

4.5 Iterative Solution of Variational Principle

In the MBIS procedure, one solves the Eqs. (4.15) and (4.16) using fixed-point iteration. That procedure can be generalized to an arbitrary *f*-divergence by writing the update formula

$$c_k^{\text{new}} = \int \rho_k^{0;\text{old}}(\mathbf{r}) f'\left(\frac{\rho^{0;\text{old}}(\mathbf{r})}{\rho(\mathbf{r})}\right) d\mathbf{r} \tag{4.32}$$

If the basis functions are varied, a sufficiently good guess for the nonlinear parameters is required. If one assumes that the basis functions have the form of a normalization function times a functional form,

$$b_k(\alpha_k;\mathbf{r}) = A_k(\alpha_k) g_k(\alpha_k;\mathbf{r}) \tag{4.33}$$

then one can derive the equation

$$\frac{1}{A(\alpha_k)} \frac{\partial A_k(\alpha_k)}{\partial \alpha_k} = -\frac{\int \frac{\partial g_k(\alpha_k;\mathbf{r})}{\partial \alpha_k} f'\left(\frac{\rho^0(\mathbf{r})}{\rho(\mathbf{r})}\right) d\mathbf{r}}{\int g_k(\alpha_k;\mathbf{r}) f'\left(\frac{\rho^0(\mathbf{r})}{\rho(\mathbf{r})}\right) d\mathbf{r}} \tag{4.34}$$

which can be expressed as an update formula,

$$\frac{1}{A_k(\alpha_k^{\text{new}})} \frac{\partial A_k(\alpha_k^{\text{new}})}{\partial \alpha_k} = -\frac{\int \frac{\partial \rho_k^{0;\text{old}}(\mathbf{r})}{\partial \alpha_k} f'\left(\frac{\rho^{0;\text{old}}(\mathbf{r})}{\rho(\mathbf{r})}\right) d\mathbf{r}}{\int \rho_k^{0;\text{old}}(\mathbf{r}) f'\left(\frac{\rho^{0;\text{old}}(\mathbf{r})}{\rho(\mathbf{r})}\right) d\mathbf{r}} \tag{4.35}$$

These update formulas correspond to the MBIS equations in the special case of the Kullback-Leibler information with the basis functions

$$b_k(\alpha_k;\mathbf{r}) = A(\alpha_k)\exp\left(-\alpha_k|\mathbf{r}-\mathbf{R}_k|\right). \tag{4.36}$$

4.6 Additive Variational Hirshfeld (AVH) Method

In existing Hirshfeld parititioning methods, the promolecule density is written as a linear combination of the spherically-averaged ground state densities of atoms and their ions. This suggests that we use the shape functions of atoms and atomic ions as the pro-pieces. That is,

$$\left\{b^0_{A,n}(\mathbf{r}) = \sigma^0_{A,n}(\mathbf{r}-\mathbf{R}_A)\right\}^{N_{\text{atoms}},N_{\max,A}}_{A=1,n=1} \tag{4.37}$$

where the atomic shape functions are the unit-normalized atomic densities (cf. Eq. (4.2))[1, 27, 118, 124-126]

$$\sigma^0_{A,n}(\mathbf{r}) = \frac{\rho^0_{A,n}(\mathbf{r})}{n} \tag{4.38}$$

where $\rho^0_{A,n}(\mathbf{r})$ is the spherically averaged electron density of the atomic ion with nuclear charge Z_A and charge $Z_A - n$, $\mathbf{R}_A$ is the location of this atomic nucleus in the molecule, and $N_{\max,A}$ is the maximum number of electrons (either $N_{\max,A} = Z_A$ or $N_{\max,A} = Z_A + 1$) that can be bound by this atom. Note that, in contrast to methods like Hirshfeld-I, the pro-pieces all correspond to bound atoms. (This is favorable since, as demonstrated in chapter 1, Hirshfeld-I can behave erratically when unbound pro-atomic reference densities are used.) We choose to allow contributions from *all* the bound ions of each atom when we form the pro-molecular density,

$$\rho^0_{\mathrm{mol}}\left(\left\{c_{A,n}\right\};\mathbf{r}\right)=\sum_{A=1}^{N_{\mathrm{atoms}}}\sum_{n=1}^{N_{\mathrm{max},A}} c_{A,n}\sigma^0_{A,n}\left(\mathbf{r}-\mathbf{R}_A\right) \tag{4.39}$$

In our computational tests, however, we observe that the shape functions of very highly charged atom ions have zero contribution, and can be neglected for computational expediency.

We also observed that the iterative approach based on Eq. (4.32) converges very slowly, requiring thousands of iterations. It was much more efficient to use the fundamental variational procedure,

$$\underbrace{\min}_{\left\{c_{A,n}\geq 0 \middle| N=\sum_{A=1}^{N_{\mathrm{atoms}}}\sum_{n=1}^{N_{\mathrm{max},A}} c_{A,n}\right\}} g\left(\left\{c_{A,n}\right\}\right) \tag{4.40}$$

$$g\left(\left\{c_{A,n}\right\}\right)=\int \rho_{\mathrm{mol}}\left(\mathbf{r}\right) f\left(\frac{\rho^0_{\mathrm{mol}}\left(\left\{c_{A,n}\right\};\mathbf{r}\right)}{\rho_{\mathrm{mol}}\left(\mathbf{r}\right)}\right) d\mathbf{r} \tag{4.41}$$

and the first and second derivatives of the objective function are (compare Eqs. (4.12) and (4.17)):

$$\frac{\partial g\left(\left\{c_{A,n}\right\}\right)}{\partial c_{A,m}}=\int \sigma^0_{A,m}\left(\mathbf{r}-\mathbf{R}_A\right) f'\left(\frac{\rho^0_{\mathrm{mol}}\left(\left\{c_{A,n}\right\};\mathbf{r}\right)}{\rho_{\mathrm{mol}}\left(\mathbf{r}\right)}\right) d\mathbf{r} \tag{4.42}$$

$$\frac{\partial g\left(\left\{c_{A,n}\right\}\right)}{\partial c_{A,m}\partial c_{B,n}}=\int \frac{\sigma^0_{A,m}\left(\mathbf{r}-\mathbf{R}_A\right)\sigma^0_{B,n}\left(\mathbf{r}-\mathbf{R}_B\right)}{\rho_{\mathrm{mol}}\left(\mathbf{r}\right)} f''\left(\frac{\rho^0_{\mathrm{mol}}\left(\left\{c_{A,n}\right\};\mathbf{r}\right)}{\rho_{\mathrm{mol}}\left(\mathbf{r}\right)}\right) d\mathbf{r} \tag{4.43}$$

We call this the additive variational Hirshfeld (AVH) method.

In AVH scheme, we concentrate on two specific families of *f*-divergence, namely the extended α-divergence[28, 138-139, 145, 158, 164]

$$f_{\alpha,\mathrm{ext}}(x)=\frac{x^{\alpha}-\alpha x+\alpha-1}{\alpha(\alpha-1)} \tag{4.44}$$

$$I_{\alpha,\mathrm{ext}}\left[\rho,\rho^{0}\right]=\frac{1}{\alpha(\alpha-1)}\int\rho(\mathbf{r})\left(\frac{\rho^{0}(\mathbf{r})}{\rho(\mathbf{r})}\right)^{\alpha}+(\alpha-1)\rho(\mathbf{r})-\alpha\rho^{0}(\mathbf{r})\,d\mathbf{r} \tag{4.45}$$

and the symmetrized α-divergence

$$f_{\alpha,\mathrm{sym}}(x)=\frac{x^{\alpha}-1+x^{1-\alpha}-x}{2\alpha(\alpha-1)} \tag{4.46}$$

$$I_{\alpha,\mathrm{sym}}\left[\rho,\rho^{0}\right]=\frac{1}{2\alpha(\alpha-1)}\int\rho(\mathbf{r})\left(\frac{\rho^{0}(\mathbf{r})}{\rho(\mathbf{r})}\right)^{\alpha}+\rho^{0}(\mathbf{r})\left(\frac{\rho(\mathbf{r})}{\rho^{0}(\mathbf{r})}\right)^{\alpha}-\left(\rho(\mathbf{r})+\rho^{0}(\mathbf{r})\right)d\mathbf{r} \tag{4.47}$$

The extended Kullback-Leibler and symmetrized Kullback-Leibler divergences correspond to the choice $\alpha=1$. A good discussion of the interpretation and significance of different values of α can be found in the report of Minka.[164]

4.7 Extensions

One advantage of a variational formulation of Hirshfeld partitioning is that it facilitates the addition of constraints. This is especially useful if one wishes to develop and use atomic charges in a molecular mechanics force field. For example, if one wished to adapt atomic charges to the geometry of a protein then,[165] unless one wishes to reparameterize the force field entirely, one should conserve (a) the total charges of the individual amino acid residues and (b) the charges of the backbone atoms. The constraints have the same general form as the constraint that the promolecule has the correct charge,

which is already included in Eq. (4.40). Specifically, these constraints on the atomic charges of an atom or group can be expressed as

$$N_{\text{group}} = \sum_{A \in \text{group}} \sum_{n=1}^{N_{\max,A}} c_{A,n} \tag{4.48}$$

This sort of linear constraint is easily incorporated into the optimization in Eq. (4.40), and the minimum of a convex function with respect to linear constraints still has a unique minimum.

Another advantage of this approach is that it is easily extended to additional states. For example, the neutral carbon atom has a $1s^2 2s^2 2p^2$ electron configuration, but one might speculate that a $1s^2 2s^1 2p^3$ electron configuration is a more appropriate reference for the carbon atom in saturated hydrocarbons. One advantage of this approach is that excited-state pro-atoms can be easily included in the sum. One merely extends Eq. (4.39) to include those states,

$$\rho^0_{\text{mol}}\left(\left\{c_{A,n,k}\right\};\mathbf{r}\right) = \sum_{A=1}^{N_{\text{atoms}}} \sum_{n=1}^{N_{\max,A}} \sum_{k=1}^{N_{\text{excited},A,n}} c_{A,n,k}\sigma^0_{A,n,k}\left(\mathbf{r}-\mathbf{R}_A\right). \tag{4.49}$$

It is not desirable, however, to include *all* the possible excited states. Including all excited states gives the pro-atom density,

$$\rho^0_A\left(\mathbf{r}\right) = \sum_{n=1}^{N_{\max,A}} \sum_{k=1}^{N_{\text{excited},A,n}} c_{A,n,k}\sigma^0_{A,n,k}\left(\mathbf{r}-\mathbf{R}_A\right) \tag{4.50}$$

too much flexibility. Indeed, *any* spherically symmetric function can be described with the expansion in Eq. (4.49), including pro-atom densities that are not monotonically decreasing.[49, 166-168] Therefore, if one includes all possible excited states, the AVH method becomes the *f*-divergence extension of the iterative stockholder analysis (ISA),[77-78] where

one minimizes the Kullback-Leibler divergence with respect to all possible spherically-symmetric reference densities.[79] As illustrated in chapter 1, ISA has well-known shortcomings: for atoms that are surrounded by a spherical shell of atoms (e.g., endohedral fullerenes), it gives atomic populations that are far too large;[79-80] for large floppy molecules like polypeptides, ISA charges show an erratic dependence on molecular conformation.[80] (Constraining the pro-atom densities to be monotonic is an inadequate remedy to these problems.[80])

If excited states are to be included in the model for the promolecular density, it is therefore essential to include only those excited states that correspond to low-energy electron configurations. Identifying which excited states to include can be challenging, especially for multiconfigurational correlated wavefunctions. A useful heuristic is to include the lowest bound excited state of each spin-multiplicity. For example, for the carbon atom one would include the ^3P, ^1D, and ^5S states. (The lowest-energy septuplet, corresponding to the electron configuration $1s^1 2s^1 2p^3 3s^1$, is unbound because it is much higher in energy than the ground state of the carbon cation, C^+.)

4.8 Conclusion

In this chapter, we introduced the additive pro-piece model to represent pro-atom density. The parameters in this model were variationally optimized to provide the most accurate approximation to the molecular density. This led to Additive Variational Hirshfeld (AVH) partitioning scheme, which is convex (has a unique solution), size

consistent and easily extendable for including additional constraints and atomic excited states.

5 Variational Hirshfeld Extensions and Case Study

5.1 Background

One strategy for defining an atom in a molecule (AIM) is to define AIMs so that their properties reproduce the properties of a reference pro-atom, typically selected to be an isolated atom or atomic ion, as strongly as possible.[1, 27] This maximizes the transferability of intuition from isolated atoms to AIM. Since AIM with similar electron densities will have similar properties, this strategy can be implemented by forcing the electron density of the AIMs to maximally resemble the electron densities of the reference pro-atoms, subject to the obvious constraint that the sum of the electron densities of the AIMs is equal to the total molecular density.[1, 25-29, 118] That is, the AIM densities are obtained by partitioning the molecular density.

In this chapter, the mathematical framework associated with minimizing the divergence between the molecular and promolecular density for the additive pro-atom model and multiplicative pro-atom models and any given *f*-divergence are explored. We present a computational strategy appropriate for these models, thereby providing a concrete realization of the additive variational Hirshfeld (AVH) and multiplicative variational Hirshfeld (MVH) partitioning schemes. After discussing some nuances

associated with the choice of divergence measure and constraints on multiplicative pro-atom densities, we present numerical results.

5.2 Mathematical Formulation

To implement this strategy mathematically, one minimizes the total divergence between the AIM densities, $\{\rho_A(\mathbf{r})\}_{A=1}^{N_{\text{atoms}}}$, and the reference pro-atom densities, $\{\rho_A^0(\mathbf{r})\}_{A=1}^{N_{\text{atoms}}}$, subject to the constraint that the AIM densities partition the molecular density, $\rho_{\text{mol}}(\mathbf{r})$.[1, 25-27] I.e.,

$$\underbrace{\min}_{\left\{\rho_A(\mathbf{r}) \middle| \rho_{\text{mol}}(\mathbf{r}) = \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r})\right\}} \sum_{A=1}^{N_{\text{atoms}}} D\left[\rho_A \middle| \rho_A^0\right] \tag{5.1}$$

Here $D\left[\rho_A \middle| \rho_A^0\right]$ is a mathematical divergence measure, which has the property that

$$D\left[\rho_A \middle| \rho_A^0\right] \geq D\left[\rho_A^0 \middle| \rho_A^0\right] = 0 \tag{5.2}$$

whenever the number of electrons in the AIM and the pro-atom are the same,

$$\int \rho_A(\mathbf{r}) = N_A = N_A^0 = \int \rho_A^0(\mathbf{r}) d\mathbf{r} \tag{5.3}$$

Usually it is not required that Eq. (5.3) hold for every AIM, but it is a convenient and desirable feature.[46, 72, 82] In this chapter we will focus on extended divergence measures where Eq. (5.2) is true for all nonnegative integrable functions, $\rho_A(\mathbf{r}) \geq 0$ and $\rho_A^0(\mathbf{r}) \geq 0$, regardless of their normalization.

5.2.1 Divergence Measures

The first,[25-26] and most popular,[31-33, 46, 58, 72-73, 77-78, 80-82, 169] divergence measure to be used in this context is the extended Kullback-Leibler directed divergence,

$$D_{\mathrm{KL,ext}}\left[\rho_A\middle|\rho_A^0\right]=\int\rho_A(\mathbf{r})\ln\left(\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})}\right)-\rho_A(\mathbf{r})+\rho_A^0(\mathbf{r})\,d\mathbf{r} \tag{5.4}$$

and its symmetrized version[27]

$$D_{\mathrm{KL,sym}}\left[\rho_A\middle|\rho_A^0\right]=\int\frac{\rho_A(\mathbf{r})-\rho_A^0(\mathbf{r})}{2}\ln\left(\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})}\right)d\mathbf{r} \tag{5.5}$$

The somewhat unusual form of directed divergence in Eq. (5.4) is the appropriate generalization of Kullback-Leibler directed divergence that ensures that Eq. (5.2) is always true, regardless of normalization.[160-162] Other measures have been also considered in chapter 3, including the (generalized) Hellinger distance,[29] the Tsallis entropy,[118] nonextensive entropies,[170] as well as the Bregman divergence.[171] Many of these results arise, however, as special cases of the *f*-divergence,[28, 138-140]

$$D_f\left[\rho_A\middle|\rho_A^0\right]=\int\rho_A(\mathbf{r})\,f\left(\frac{\rho_A^0(\mathbf{r})}{\rho_A(\mathbf{r})}\right)d\mathbf{r} \tag{5.6}$$

where $f(x)$ is any convex function with $f(1)=f'(1)=0$. The requirement $f'(1)=0$ is needed to ensure Eq. (5.2) holds even when $N_A\neq N_A^0$. Every *f*-divergence can be "extended" so that it can be used for non-normalized densities by defining $f_{\mathrm{extended}}(x)=f(x)-f'(1)(x-1)$. Symmetrized *f*-divergences like Eq. (5.5) are associated

with the additional identity $f(x) = \frac{1}{2} f(x) + \frac{1}{2} x f\left(x^{-1}\right)$. Every symmetrized *f*-divergence is also an extended *f*-divergence, but the converse is not true.

For every *f*-divergence, the AIM densities have the "stockholder" form,

$$\rho_A(\mathbf{r}) = \frac{\rho_A^0(\mathbf{r})}{\sum_{B=1}^{N_{\text{atoms}}} \rho_B^0(\mathbf{r})} \tag{5.7}$$

This form was first proposed by Hirshfeld,[21] building on the work of Politzer,[20] on heuristic grounds. The quantity

$$\rho_{\text{mol}}^0(\mathbf{r}) = \sum_{B=1}^{N_{\text{atoms}}} \rho_B^0(\mathbf{r}) \tag{5.8}$$

is usually called the promolecular density. As discussed in chapter 2, *every* divergence measure that (a) is a local functional of $\rho_A(\mathbf{r})$ and $\rho_A^0(\mathbf{r})$ and (b) gives the stockholder partitioning (Eq. (5.7)) when used in minimization (5.1) is an *f*-divergence.[28]

Moreover, for an *f*-divergence, one has

$$D_f\left[\rho_{\text{mol}} \middle| \rho_{\text{mol}}^0\right] = D_f\left[\rho_{\text{mol}} \middle| \sum_{A=1}^{N_{\text{atoms}}} \rho_A^0\right] = \sum_{A=1}^{N_{\text{atoms}}} D_f\left[\rho_A \middle| \rho_A^0\right] \tag{5.9}$$

This suggests that the optimal pro-atoms should be determined by minimizing the divergence between the molecular and promolecular densities. This allows one to define adaptive, molecule-specific, pro-atoms.[28]

5.2.2 Pro-atom Density Models

The importance of choosing pro-atoms that are adapted to the molecule being partitioned was first recognized in the work of Bultinck *et al.*,[46] who used the ground-

state atoms with fractional charge in an iterative, non-variational method.[172,173] We propose, however, to use a variational procedure,

$$\underbrace{\min}_{\{\mathbf{c}\}} D\left[\rho_{\mathrm{mol}} \middle| \rho^0_{\mathrm{mol}}(\mathbf{c})\right], \tag{5.10}$$

where **c** is a list of parameters upon which the promolecular density depends. We wish to retain the conceptually useful picture of AIM densities that maximally resemble pro-atom densities, so we consider only promolecular densities that can be expressed as Eq. (5.8). (Not every Hirshfeld-like partitioning respects this choice.[31, 58, 77-78, 80, 82]) Similarly, we wish to retain the picture that the pro-atom densities correspond to suitably chosen reference states of the isolated atom, including relevant ions and possibly low-lying excited states. The optimized pro-atom then gives us information about the dominant charge and excited (promoted) reference states of the atoms, facilitating a valence-bond-like interpretation of molecular electronic structure.[174-177] For example, in the previous chapter we chose

$$\begin{aligned} \rho^0_A(\mathbf{c}_A, \mathbf{r}) &= \sum_{n=1}^{N_{\max,A}} \sum_{k=0}^{N_{\mathrm{excite},A}} c_{A,n,k} \rho^0_{A,n,k}(\mathbf{r}) \\ c_{A,n,k} &\geq 0 \end{aligned} \tag{5.11}$$

where $\rho^0_{A,n,k}(\mathbf{r})$ is the spherically-averaged electron density of the k^{th} included excited state of atom A when it has n electrons (and, therefore, charge $q_A = Z_A - n$, where Z_A is the atomic number of atom A). Variational minimization of Eq. (5.10) using the additive pro-atom model in Eq. (5.11) is called the additive variational Hirshfeld (AVH) method.

In AVH, the pro-atom densities are expressed as a (non-normalized) weighted average of the electron densities of the isolated atom states. If one generalizes this formula to the *p*-mean, one has

$$\rho_A^0(\mathbf{c}_A,\mathbf{r}) = \left(\sum_{n=1}^{N_{\max A}} \sum_{k=0}^{N_{\text{excite},A}} c_{A,n,k} \left(\rho_{A,n,k}^0(\mathbf{r}) \right)^p \right)^{1/p} \qquad c_{A,n,k} \geq 0 \tag{5.12}$$

Equation (5.11) corresponds to the ordinary arithmetic mean, $p = 1$. Assuming that none of the coefficients in Eq. (5.12) are exactly zero, then for $p > 0$, the pro-atom density decays very slowly asymptotically, with its decay controlled by the ionization potential of the most-weakly-bound atomic state in the sum, i.e.,[178-181]

$$\rho_A^0(\mathbf{c},\mathbf{r}) :\ \exp\left(-r \sqrt{8\text{IP}_{N_{\max,A} N_{\text{excite},A}, N_{\max}}} \right) \tag{5.13}$$

For $p < 0$, the pro-atom density decays very rapidly asymptotically, with its decay controlled by the ionization potential of the most-strongly-bound atomic state in the sum,

$$\rho_A^0(\mathbf{c},\mathbf{r}) :\ \exp\left(-r \sqrt{8\text{IP}_{1,0}} \right). \tag{5.14}$$

These asymptotic decay rates seem un-chemical. Based on the electronegativity equalization principle, we expect that all the AIM have the same ionization potential as the molecule as a whole (and each other). We also expect that the pro-atoms should have the same, or at least very similar, ionization potentials. This suggests that the ionization potentials of the pro-atoms should lie between the extreme limits in Eqs. (5.13) and (5.14) . We note that this is not merely a formal problem: some of the failures of Hirshfeld methods are often attributed the pro-atom densities having very different asymptotic

decays, because then the slowly-decaying pro-atoms take excessive electron density from neighboring atoms.[117] In extreme cases, this can lead to a "runaway charges" effect, and is associated with electropositive atoms that are far too positively charged, even as electronegative atoms become far too negatively charged.

The asymptotic decay of the pro-atoms changes if one uses the geometric mean, corresponding to $p = 0$,

$$\rho_A^0(\mathbf{c}_A, \mathbf{r}) = c_{A,0,0} \prod_{n=1}^{N_{\max,A}} \prod_{k=0}^{N_{\text{excite},A}} \left(\rho_{A,n,k}^0(\mathbf{r})\right)^{c_{A,n,k}} \qquad c_{A,n,k} \geq 0 \tag{5.15}$$

We have slightly extended the $p = 0$ mean by including the multiplicative scaling factor $c_{A,0,0}$. The asymptotic decay of this multiplicative pro-atom is

$$\rho_A^0(\mathbf{c}_A, \mathbf{r}) \sim \exp\left[-r\left(\sum_{n=1}^{N_{\max,A}} \sum_{k=0}^{N_{\text{excite},A,n}} c_{A,n,k}\sqrt{8\text{IP}_{A,n,k}}\right)\right]. \tag{5.16}$$

The rationale for including a multiplicative scaling factor is that it ensures that the AIM and pro-atoms have the same charge, cf. Eq. (5.3), when the extended Kullback-Leibler divergence is selected. (See section 5.2.6) We call minimizing the divergence between the promolecular and molecular densities (cf. Eq. (5.10)) using the multiplicative pro-atom model in Eq. (5.15) the multiplicative variational Hirshfeld (MVH) method.

5.2.3 Lagrangian and Its Derivatives

We wish to define the pro-atom densities by minimizing the divergence between the molecular and promolecular densities, as in Eq. (5.10), subject to the constraint that the molecular and promolecular densities contain the same number of electrons,

$$\int \rho_{\text{mol}}(\mathbf{r})\, d\mathbf{r} = N_{\text{mol}} = N_{\text{mol}}^{0} = \int \rho_{\text{mol}}^{0}(\mathbf{r})\, d\mathbf{r} \tag{5.17}$$

This constraint is chemically intuitive. If one is not using an extended or symmetrized *f*-divergence, a constraint like this is *essential* because otherwise the objective function is unbound from below. We will consider pro-atoms that are defined by both the additive, Eq. (5.11), and multiplicative, Eq. (5.15), formulas. In the appendix 7.2, we catalogue multiple divergence formulas that we find especially interesting. The Lagrangian is,

$$\Lambda(\mathbf{c},\mu) = \int \rho_{\text{mol}}(\mathbf{r})\, f\left(\frac{\rho_{\text{mol}}^{0}(\mathbf{c},\mathbf{r})}{\rho_{\text{mol}}(\mathbf{r})}\right) d\mathbf{r} - \mu\int\left(\rho_{\text{mol}}(\mathbf{r}) - \rho_{\text{mol}}^{0}(\mathbf{c},\mathbf{r})\right) d\mathbf{r} \tag{5.18}$$

The gradient of this Lagrangian is

$$\begin{aligned} 0 &= \frac{\partial\Lambda(\mathbf{c},\mu)}{\partial c_{A,m,k}} = \int f'\left(\frac{\rho_{\text{mol}}^{0}(\mathbf{c},\mathbf{r})}{\rho_{\text{mol}}(\mathbf{r})}\right)\left(\frac{\partial\rho_{\text{mol}}^{0}(\mathbf{c},\mathbf{r})}{\partial c_{A,m,k}}\right) d\mathbf{r} + \mu\int\left(\frac{\partial\rho_{\text{mol}}^{0}(\mathbf{c},\mathbf{r})}{\partial c_{A,m,k}}\right) d\mathbf{r} \\ 0 &= \frac{\partial\Lambda(\mathbf{c},\mu)}{\partial\mu} = \int\left(\rho_{\text{mol}}^{0}(\mathbf{c},\mathbf{r}) - \rho_{\text{mol}}(\mathbf{r})\right) d\mathbf{r} \end{aligned} \tag{5.19}$$

For a given value of Lagrange multiplier μ, the Hessian is

$$\begin{aligned} \frac{\partial^2\Lambda(\mathbf{c},\mu)}{\partial c_{A,m,k}\partial c_{B,n,l}} &= \int\left(\frac{1}{\rho_{\text{mol}}(\mathbf{r})}\right) f''\left(\frac{\rho_{\text{mol}}^{0}(\mathbf{c},\mathbf{r})}{\rho_{\text{mol}}(\mathbf{r})}\right)\left(\frac{\partial\rho_{\text{mol}}^{0}(\mathbf{c},\mathbf{r})}{\partial c_{A,m,k}}\right)\left(\frac{\partial\rho_{\text{mol}}^{0}(\mathbf{c},\mathbf{r})}{\partial c_{B,n,l}}\right) d\mathbf{r} \\ &+ \int\left[f'\left(\frac{\rho_{\text{mol}}^{0}(\mathbf{c},\mathbf{r})}{\rho_{\text{mol}}(\mathbf{r})}\right) + \mu\right]\left(\frac{\partial^2\rho_{\text{mol}}^{0}(\mathbf{c},\mathbf{r})}{\partial c_{A,m,k}\partial c_{B,n,l}}\right) d\mathbf{r} \end{aligned} \tag{5.20}$$

For the additive pro-atom model, the derivatives of the promolecular density with respect to its parameters have the simple expressions,

$$\frac{\partial\rho_{\text{mol}}^{0}(\mathbf{c}_A,\mathbf{r})}{\partial c_{A,m,k}} = \rho_{A,m,k}^{0}(\mathbf{r}) \tag{5.21}$$

$$\frac{\partial^2 \rho^0_{\mathrm{mol}}(\mathbf{c}_A, \mathbf{r})}{\partial c_{A,m,k}\, \partial c_{B,n,l}} = 0 \tag{5.22}$$

Since

$$\left(\frac{1}{\rho_{\mathrm{mol}}(\mathbf{r})}\right) f''\left(\frac{\rho^0_{\mathrm{mol}}(\mathbf{c}, \mathbf{r})}{\rho_{\mathrm{mol}}(\mathbf{r})}\right) > 0, \tag{5.23}$$

the Hessian is positive definite, the objective function is convex, and the minimum of the additive pro-atom model is unique.

For the multiplicative pro-atom model, the derivatives of the promolecular density with respect to its parameters have the expressions

$$\begin{aligned} \frac{\partial \rho^0_{\mathrm{mol}}(\mathbf{c}_A, \mathbf{r})}{\partial c_{A,0,0}} &= \frac{\rho^0_A(\mathbf{c}_A, \mathbf{r})}{c_{A,0,0}} \\ \frac{\partial \rho^0_{\mathrm{mol}}(\mathbf{c}_A, \mathbf{r})}{\partial c_{A,m,k}} &= \rho^0_A(\mathbf{c}_A, \mathbf{r}) \ln \rho^0_{A,m,k}(\mathbf{r}) \end{aligned} \tag{5.24}$$

$$\begin{aligned} \frac{\partial^2 \rho^0_{\mathrm{mol}}(\mathbf{c}_A, \mathbf{r})}{\partial c^2_{A,0,0}} &= 0 \\ \frac{\partial^2 \rho^0_{\mathrm{mol}}(\mathbf{c}_A, \mathbf{r})}{\partial c_{A,m,k}\, \partial c_{A,0,0}} &= \frac{\rho^0_A(\mathbf{c}_A, \mathbf{r}) \ln \rho^0_{A,m,k}(\mathbf{r})}{c_{A,0,0}} \\ \frac{\partial^2 \rho^0_{\mathrm{mol}}(\mathbf{c}_A, \mathbf{r})}{\partial c_{A,m,k}\, \partial c_{B,n,l}} &= \delta_{AB} \rho^0_A(\mathbf{c}_A, \mathbf{r}) \ln \rho^0_{A,m,k}(\mathbf{r}) \ln \rho^0_{B,n,l}(\mathbf{r}) \end{aligned} \tag{5.25}$$

where $\rho^0_A(\mathbf{c}_A, \mathbf{r})$ is defined in Eq. (5.15). The optimization of multiplicative pro-atoms is not always convex, but it will tend to be convex when the "diagonal" atom blocks of the Hessian are predominately positive

$$\rho^0_A(c_A, \mathbf{r}) \left[\left(\frac{\rho^0_A(c_A, \mathbf{r})}{\rho_{\mathrm{mol}}(\mathbf{r})}\right) f''\left(\frac{\rho^0_{\mathrm{mol}}(\mathbf{c}, \mathbf{r})}{\rho_{\mathrm{mol}}(\mathbf{r})}\right) + f'\left(\frac{\rho^0_{\mathrm{mol}}(\mathbf{c}, \mathbf{r})}{\rho_{\mathrm{mol}}(\mathbf{r})}\right) \right] \geq 0 \tag{5.26}$$

Notice that the first term in square brackets is always positive, and that because we are considering models with $f'(1)=0$, the second term is nearly zero whenever the promolecular density is an accurate approximation to the true density. This suggests that the multiplicative pro-atom model should be unproblematic whenever an adequate initial guess is available.

5.2.4 Lagrange Multiplier: Explicit Formulas

As in the recently proposed minimal basis iterative stockholder (MBIS), the Lagrange multiplier μ can be solved for explicitly.[82] For the additive model, inserting Eq. (5.21) into Eq. (5.19), multiplying by $c_{A,m,k}$, and summing over all the pro-atom pieces gives,

$$\begin{aligned} 0 &= \sum_{A,m,k} c_{A,m,k} \int f'\left(\frac{\rho^0_{\text{mol}}(\mathbf{c},\mathbf{r})}{\rho_{\text{mol}}(\mathbf{r})} \right) \rho^0_{A,m,k}(\mathbf{r})\, d\mathbf{r} + \mu \int \rho^0_{A,m,k}(\mathbf{r})\, d\mathbf{r} \\ \mu &= -\frac{\int f'\left(\frac{\rho^0_{\text{mol}}(\mathbf{c},\mathbf{r})}{\rho_{\text{mol}}(\mathbf{r})} \right) \rho^0_{\text{mol}}(\mathbf{c},\mathbf{r})\, d\mathbf{r}}{\int \rho^0_{\text{mol}}(\mathbf{c},\mathbf{r})\, d\mathbf{r}} \end{aligned} \tag{5.27}$$

Notice that the Lagrange multiplier is almost zero when the promolecular density is very similar to the molecular density,

$$\begin{aligned} \mu &= -\frac{\int \left(f'(1) + \left(\frac{\rho^0_{\text{mol}}(\mathbf{c},\mathbf{r})}{\rho_{\text{mol}}(\mathbf{r})} - 1 \right) f''(1) + \cdots \right) \rho^0_{\text{mol}}(\mathbf{c},\mathbf{r})\, d\mathbf{r}}{\int \rho^0_{\text{mol}}(\mathbf{c},\mathbf{r})\, d\mathbf{r}} \\ &\approx -\frac{\int \frac{\rho^0_{\text{mol}}(\mathbf{c},\mathbf{r})}{\rho_{\text{mol}}(\mathbf{r})} \left(\rho^0_{\text{mol}}(\mathbf{c},\mathbf{r}) - \rho_{\text{mol}}(\mathbf{r}) \right) d\mathbf{r}}{\int \rho^0_{\text{mol}}(\mathbf{c},\mathbf{r})\, d\mathbf{r}} \approx 0 \end{aligned} \tag{5.28}$$

Here again we have assumed that $f'(1)=0$, as it is for extended and symmetric *f*-divergences. For the multiplicative pro-atom model, we likewise insert the expression for the gradient, Eq. (5.24), into Eq. (5.19), multiply by $c_{A,m,k}$, and sum over all the pro-atom pieces. This gives

$$\begin{aligned} 0 &= \int f'\left(\frac{\rho^0_{\text{mol}}(\mathbf{c},\mathbf{r})}{\rho_{\text{mol}}(\mathbf{r})}\right)\left[\sum_{A=1}^{N_{\text{atoms}}}\rho^0_A(\mathbf{c},\mathbf{r})\left(1+\ln\rho^0_A(\mathbf{c},\mathbf{r})-\ln c_{A,0,0}\right)\right]d\mathbf{r} \\ &\quad +\mu\left[\int\sum_{A=1}^{N_{\text{atoms}}}\rho^0_A(\mathbf{c},\mathbf{r})\left(1+\ln\rho^0_A(\mathbf{c},\mathbf{r})-\ln c_{A,0,0}\right)d\mathbf{r}\right] \\ \mu &= -\frac{\int f'\left(\frac{\rho^0_{\text{mol}}(\mathbf{c},\mathbf{r})}{\rho_{\text{mol}}(\mathbf{r})}\right)\left[\sum_{A=1}^{N_{\text{atoms}}}\rho^0_A(\mathbf{c},\mathbf{r})\left(1+\ln\rho^0_A(\mathbf{c},\mathbf{r})-\ln c_{A,0,0}\right)\right]d\mathbf{r}}{\int\sum_{A=1}^{N_{\text{atoms}}}\rho^0_A(\mathbf{c},\mathbf{r})\left(1+\ln\rho^0_A(\mathbf{c},\mathbf{r})-\ln c_{A,0,0}\right)d\mathbf{r}} \end{aligned} \tag{5.29}$$

As for the additive pro-atom model, $\mu\approx 0$ if the promolecular and molecular densities are very similar.

5.2.5 Computational Details

We use gradient-based optimization to optimize the Lagrangian in Eq. (5.18). As an initial guess, we use a method we call the scaled Hirshfeld (SH) method. In the scaled Hirshfeld method, the pro-atom densities are the spherically-averaged neutral-atom densities, scaled by a multiplicative constant. The promolecule density is therefore

$$\rho^0_{\text{mol}}(\mathbf{r}) = \sum_{A=1}^{N_{\text{atoms}}} c^0_A \rho^0_{A,Z_A,0}(\mathbf{r}), \tag{5.30}$$

where Z_A is the atomic number of atom *A*. The scaled Hirshfeld method reduces to the traditional Hirshfeld partitioning method when $c^0_A = 1$ for all the atoms.[21] Unlike the

traditional Hirshfeld method, however, the promolecule in the scaled Hirshfeld method will have the same charge as the molecule. Therefore, unlike the traditional Hirshfeld method, the scaled Hirshfeld method is equally appropriate for neutral and charged molecules. The scaled Hirshfeld method is an appropriate initial guess since it is a special case of *both* the additive and multiplicative pro-atom models. Specifically, the additive pro-atom model corresponds to $c_{A,n,k} = c_A^0 \delta_{nZ_A} \delta_{k0}$; the multiplicative pro-atom model corresponds to $c_{0,0,0} = c_A^0$ and $c_{A,n,k} = \delta_{nZ_A} \delta_{k0}$. To this initial guess, we add a small positive noise to select atomic states; this ensures that the initial optimization point is away from the boundary of the feasible region, $c_{A,n,k} \geq 0$. After optimization has concluded, we test the gradient of the objective function with respect to all the parameters that are zero, $c_{A,n,k} = 0$, to ensure that the objective function could not be lowered by increasing the values of these parameters. The initial guess for the Lagrange multiplier is computed using Eq. (5.27) (additive pro-atom model) or Eq. (5.29) (multiplicative pro-atom model).

5.2.6 Special Case of Extended Kullback-Leibler

In the previous chapter, we noted that this variational approach to atoms in molecules is only size-consistent if the Lagrange multiplier is a constant, independent of zero. For the additive model, the extended Kullback-Leibler divergence had the advantage of being size consistent. We now show that this is also true for the scaled Hirshfeld charges, the multiplicative model, and indeed any pro-atom model for which each pro-atom density is scaled by a multiplicative constant. This can be seen as a small

generalization of the computational framework associated with minimal basis iterative stockholder (MBIS) partitioning.[82]

Consider a promolecular density with the form

$$\rho^0_{\text{mol}}\left(\{c_A\},\{\boldsymbol{\alpha}_A\},\mathbf{r}\right)=\sum_{A=1}^{N_{\text{atoms}}} c_A\rho^0_A\left(\boldsymbol{\alpha}_A;\mathbf{r}\right) \tag{5.31}$$

The parameters $\boldsymbol{\alpha}_A$ are internal degrees of freedom in the pro-atom model; there are no such internal degrees of freedom for the scaled Hirshfeld pro-atom (cf. Eq. (5.30)) and in the multiplicative model each atom has as internal degrees of freedom the exponents $c_{A,n,k}$ (cf. Eq. (5.15)). The additive model already has the form of Eq. (5.31) with no free parameters, but instead of pro-atom densities it uses the densities of individual charge/excitation states of the pro-atoms.

We rewrite the variational principle as a nested variational principle,

$$\min_{\{c_A\geq 0\}}\min_{\{\boldsymbol{\alpha}_A\}}\int \rho_{\text{mol}}\left(\mathbf{r}\right)\ln\left(\frac{\rho_{\text{mol}}\left(\mathbf{r}\right)}{\rho^0_{\text{mol}}\left(\{c_A\},\{\boldsymbol{\alpha}_A\},\mathbf{r}\right)}\right)+\rho^0_{\text{mol}}\left(\{c_A\},\{\boldsymbol{\alpha}_A\},\mathbf{r}\right)-\rho_{\text{mol}}\left(\mathbf{r}\right)d\mathbf{r} \tag{5.32}$$

Denote the parameters that solve the inner minimization as $\{\boldsymbol{\alpha}^{\min}_A\}$. The minimizing values for the multiplicative scaling factors, $\{c_A\}$, can then be obtained by solving the equations

$$0=\int \rho^0_A\left(\boldsymbol{\alpha}^{\min}_A;\mathbf{r}\right)-\frac{\rho_{\text{mol}}\left(\mathbf{r}\right)}{\rho^0_{\text{mol}}\left(\{c_A\},\{\boldsymbol{\alpha}_A\},\mathbf{r}\right)}\rho^0_A\left(\boldsymbol{\alpha}^{\min}_A\right)d\mathbf{r} \tag{5.33}$$

Recall that the second term is the density of the atom-in-a-molecule, Eq. (5.7). This then implies that the AIM and the pro-atom have the same normalization,

$$N_A^0\left(\boldsymbol{\alpha}_A^{\min}\right)=\int\rho_A^0\left(\boldsymbol{\alpha}_A^{\min};\mathbf{r}\right)d\mathbf{r}=\int\rho_A\left(c_A,\boldsymbol{\alpha}_A^{\min};\mathbf{r}\right)d\mathbf{r}=N_A \tag{5.34}$$

If one sums both sides of this equation over all the atoms, then it becomes clear that molecule and promolecule are normalized to the same number of electrons for the extended Kullback-Leibler information, and so the constraint in Eq. (5.17) does not need to be imposed. Because there is no need for a constraint that couples together the atoms on different molecular fragments, the method is size consistent. That is, for molecular fragments, F and G, which are so well-separated that their fragment molecular densities and their reference pro-atom densities do not overlap, it is equivalent to determine the reference pro-atom atoms either separately or together,

$$\begin{aligned}
&\underbrace{\min}_{\{c_A\geq 0\}}\underbrace{\min}_{\{\boldsymbol{\alpha}_A\}}\int\rho_{\mathrm{F}\cdots\mathrm{G}}\left(\mathbf{r}\right)\ln\left(\frac{\rho_{\mathrm{F}\cdots\mathrm{G}}\left(\mathbf{r}\right)}{\rho^0_{\mathrm{F}\cdots\mathrm{G}}\left(\left\{c_A\right\},\left\{\boldsymbol{\alpha}_A\right\},\mathbf{r}\right)}\right)+\rho^0_{\mathrm{F}\cdots\mathrm{G}}\left(\left\{c_A\right\},\left\{\boldsymbol{\alpha}_A\right\},\mathbf{r}\right)-\rho_{\mathrm{F}\cdots\mathrm{G}}\left(\mathbf{r}\right)d\mathbf{r}\\
&=\underbrace{\min}_{\{c_A\geq 0|A\in\mathrm{F}\}}\underbrace{\min}_{\{\boldsymbol{\alpha}_A|A\in\mathrm{F}\}}\int\rho_{\mathrm{F}}\left(\mathbf{r}\right)\ln\left(\frac{\rho_{\mathrm{F}}\left(\mathbf{r}\right)}{\rho^0_{\mathrm{F}}\left(\left\{c_A\right\},\left\{\boldsymbol{\alpha}_A\right\},\mathbf{r}\right)}\right)+\rho^0_{\mathrm{F}}\left(\left\{c_A\right\},\left\{\boldsymbol{\alpha}_A\right\},\mathbf{r}\right)-\rho_{\mathrm{F}}\left(\mathbf{r}\right)d\mathbf{r}\\
&+\underbrace{\min}_{\{c_A\geq 0|A\in\mathrm{G}\}}\underbrace{\min}_{\{\boldsymbol{\alpha}_A|A\in\mathrm{G}\}}\int\rho_{\mathrm{G}}\left(\mathbf{r}\right)\ln\left(\frac{\rho_{\mathrm{G}}\left(\mathbf{r}\right)}{\rho^0_{\mathrm{G}}\left(\left\{c_A\right\},\left\{\boldsymbol{\alpha}_A\right\},\mathbf{r}\right)}\right)+\rho^0_{\mathrm{G}}\left(\left\{c_A\right\},\left\{\boldsymbol{\alpha}_A\right\},\mathbf{r}\right)-\rho_{\mathrm{G}}\left(\mathbf{r}\right)d\mathbf{r}
\end{aligned} \tag{5.35}$$

where the promolecular density of a fragment is defined by summing over the pro-atom densities in that fragment,

$$\rho^0_{\mathrm{F}}\left(\left\{c_A\right\},\left\{\boldsymbol{\alpha}_A\right\},\mathbf{r}\right)=\sum_{A\in\mathrm{F}}c_A\rho^0_A\left(\boldsymbol{\alpha}_A;\mathbf{r}\right). \tag{5.36}$$

Notice that size-consistency property and the equality of the AIM and pro-atom populations are true even when constraints are imposed on the inner minimization in Eq. (5.32). Given the somewhat contrived form of the multiplicative pro-atom model, Eq.

(5.15), this motivates us to explore what sorts of constraints we could impose to make the multiplicative pro-atomic densities more realistic.

5.2.7 Constraints on the Multiplicative Pro-atom Model

The additive pro-atom model always gives pro-atom densities that satisfy the cusp constraint,[47-48]

$$Z_A = -\frac{1}{2\rho_A^0(\mathbf{R}_A)}\left(\frac{\partial\rho_A^0(\mathbf{r})}{\partial|\mathbf{r}-\mathbf{R}_A|}\right)_{\mathbf{r}=\mathbf{R}_A}. \tag{5.37}$$

Satisfying this constraint for the multiplicative pro-atom model forces one to satisfy the additional constraints

$$\left\{1 = \sum_{n=1}^{N_{\max,A}}\sum_{k=0}^{N_{\text{excite},A,n}} c_{A,n,k}\right\}_{A=1}^{N_{\text{atoms}}}. \tag{5.38}$$

The atoms-in-molecule from the additive pro-atom model have different asymptotic decays. Within the multiplicative pro-atom model, however, we can constrain all the AIM to have the same asymptotic decay by forcing all the pro-atoms to have the same asymptotic decay,

$$\left\{\sum_{n=1}^{N_{\max,A}}\sum_{k=0}^{N_{\text{excite},A,n}} c_{A,n,k}\sqrt{8\text{IP}_{A,n,k}} = \sum_{n=1}^{N_{\max,A}}\sum_{k=0}^{N_{\text{excite},A,n}} c_{1,n,k}\sqrt{8\text{IP}_{1,n,k}}\right\}_{A=2}^{N_{\text{atoms}}}. \tag{5.39}$$

This constraint is suggested by the electronegativity equalization principle.[50-51] Notice, however, that because of the inherent non-locality of the asymptotic density decay,[52] imposing the asymptotic constraint in Eq. (5.39) destroys the size-consistency of the partitioning method. We could furthermore force all the AIM densities to decay at the rate of the molecular density,

$$\left\{ \sum_{n=1}^{N_{\max,A}} \sum_{k=0}^{N_{\text{excite},A,n}} c_{A,n,k} \sqrt{8\text{IP}_{A,n,k}} = \sqrt{8\text{IP}_{\text{mol}}} \right\}_{A=1}^{N_{\text{atoms}}} \tag{5.40}$$

Both the cusp and the asymptotic constraints only affect the outer minimization in Eq. (5.32); this means that even when these constraints are imposed, the AIM and pro-atom charges are the same when the extended Kullback-Leibler information is used.

Since it is difficult to determine the molecular ionization potential directly from the asymptotic decay of the density,[182] constraint in Eq. (5.40) would require additional knowledge about the molecule that might not always be available. In addition, some molecules have ionization potentials that are smaller than the ionization potential of the least-bound charge state of a system. For example, for any molecule containing nitrogen with an ionization potential less than 14.5 eV, it is impossible to match the asymptotic decay of the nitrogen pro-atom to the molecular asymptotic decay. Similarly, for any molecule containing hydrogen with an ionization potential greater than 13.6 eV, it is impossible to match the asymptotic decay of the hydrogen pro-atom to the molecular asymptotic decay.[183] It is not even always possible to satisfy Eq. (5.39): for any molecule containing both nitrogen and hydrogen atoms, the slowest possible asymptotic decay of a nitrogen pro-atom density (IP = 14.5 eV) is faster than the fastest possible decay of a hydrogen pro-atom density (IP = 13.6 eV), and so Eq. (5.39) cannot be satisfied.

In the next section we will perform computational tests on the additive pro-atom model and the multiplicative pro-atom model. We will also consider the utility of constraints on the multiplicative pro-atoms like the nuclear cusp constraint (Eq. (5.38)) and the asymptotic decay constraint (Eq. (5.39)).

5.3 Numerical Assessment

5.3.1 Computational Procedure

All quantum chemistry calculations were performed using *Gaussian09* (version C.01 for the calculations in sections 5.3.2 to 5.3.5; version D.01 for the calculations in section 5.3.6)[184] using the stable=opt keyword to ensure that a local minimum of the energy with respect to orbital rotations was located. Kohn-sham density-functional theory (DFT) calculations were performed employing ultrafine integration grids. For the molecules in sections 5.3.2 to 5.3.5, the geometries were optimized at UωB97XD/cc-pVTZ level of theory, the molecules in section 5.3.6 are a subset of a larger database we are building, and were optimized at the UB3LYP/Def2-TZVPD level.[185] The population analysis was performed based on single point calculations at UHF, UB3LYP,[108, 186-187] and UωB97XD[111] levels of theory with Dunning's (d-aug-)cc-pVXZ (X=D, T, Q, 5) correlation consistent basis set series.[112-114]

To understand how our methods compare to more traditional methods, we also computed atomic populations using natural population analysis (NPA)[14-15] and molecular electrostatic potential fitting (via the Hu-Lu-Yang method[106]; ESP). Among information-theoretic methods, we decided to compare our results to the conventional Hirshfeld method (H), the iterative Hirshfeld method (Hirshfeld-I; HI), iterative stockholder analysis (ISA), and the minimal basis iterative stockholder approach (MBIS). Recall that MBIS differs from the scaled Hirshfeld method (Eq. (5.30), SH), the additive variational Hirshfeld method (AVH), and the multiplicative variational Hirshfeld method (MVH)

only because it uses *s*-type Slater orbitals to construct the pro-atoms. In this chapter, the pro-atoms were constructed using the spherically-averaged densities of isolated atoms and atomic ions, at the same level of theory and basis set used for the molecule being partitioned. Unless otherwise noted, all AVH and MVH calculations are performed using the ground-state densities of neutral and charged atoms that are bound at that level of theory. For example, if an atomic anion is not bound at the Hartree-Fock level for a given basis set, then that anion's electron density is not included in the pro-atom database for calculations using that method. All information-theoretic partitioning was performed using an in-house version of HORTON.[188] For the results presented here, the minimization of the objective function for the AVH, MVH, and SH methods was performed using the Sequential Least SQuares Programming (SLSQP) method, as implemented in Python library SciPy.

5.3.2 Test Case: Lithium Chloride

As a first example, we will consider the lithium chloride molecule. Figure 5.1 shows the charges obtained from various quantum chemistry methods, basis sets, and population analysis approaches. Ionic molecules like LiCl are prototypical failures of the conventional Hirshfeld method: because the Hirshfeld AIM diverge minimally from the neutral pro-atom densities, in ionic molecules the Hirshfeld AIM are too close to neutral, and the charges are too small. The scaled Hirshfeld (SH) method is not much better. All of the other information-theoretic methods give similar charges, with the methods that use basis sets to construct the pro-atom densities (Hirshfeld-I, AVH, and MVH) showing similar basis-set dependence.

Table 5.1 shows the optimized charges and pro-atom parameters of AVH for lithium chloride. Notice that the lithium anion is not bound at the Hartree-Fock level for any of the basis sets considered, and so it is not available to construct the lithium pro-atom. Most of the basis-set dependence of the AVH charges is related to the need for diffuse functions in the basis so that the chlorine anion's density is well-described. Once diffuse functions are included, there is consensus between the various quantum chemistry methods and aug-cc-pVXZ basis sets that the charge on the lithium atom is about +0.97, in accord with our chemical expectations that this molecule is ionic. It is also reassuring that the dominant contribution to the lithium pro-atom density is from the cation, with a very small contribution from the neutral atom and negligible contributions from the other charge states.

Table 5.2 shows the optimized charges and pro-atom parameters of MVH for lithium chloride, with the cusp constraint in Eq. (5.38) imposed. Again, most of the basis-set dependence in MVH is due to the inaccuracy of the reference pro-atom densities when the basis set does not include diffuse functions. For the aug-cc-pVXZ basis sets, the MVH charges are tightly clustered, and the different quantum chemistry techniques give a consensus charge on the Lithium atom of about +0.94. This is slightly smaller than for AVH but nonetheless wholly consistent with our chemical intuition.

The pro-atom parameters in MVH, however, are very inconsistent with our chemical intuition. These parameters are very dependent on the method and basis set, and sometimes one sees significant contributions from the electron density of the lithium anion or the lithium dication. It is mildly reassuring that at least for the largest basis sets

(aug-ccpVXZ, X=T,Q,5) the lithium pro-atom is composed almost exclusively from Li^0 and Li^+, which contribute in roughly equal portions. Remarkably, the charges from the MVH model do not seem to be especially sensitive to the parameters that define the composition of the pro-atom.

Figure 5.1 Charge of lithium atom in lithium chloride computed with various partitioning schemes at various levels of theory. For each scheme, the three columns plot the charges computed using eight Dunning basis sets, i.e. (aug-)ccpVXZ with X=D, T, Q, 5 basis functions, at UHF, UB3LYP, and UωB97XD levels of theory, respectively. The absolute range of the atomic charges obtained using various basis sets at each level of theory is summarized on the *x*-axis alongside the name of partitioning method. The methods used are Hirshfeld (H), Iterative Hirshfeld (HI), Iterative Stockholder Analysis (ISA), Minimal Basis Stockholder Analysis (MBIS), Scaled Hirshfeld (SH), Additive Variational Hirshfeld (AVH), Multiplicative Variational Hirshfeld with the cusp constraint in Eq. (5.38) (MVH), Hu-Lu-Yang electrostatic fitted charges (ESP), and Natural Population Analysis (NPA).

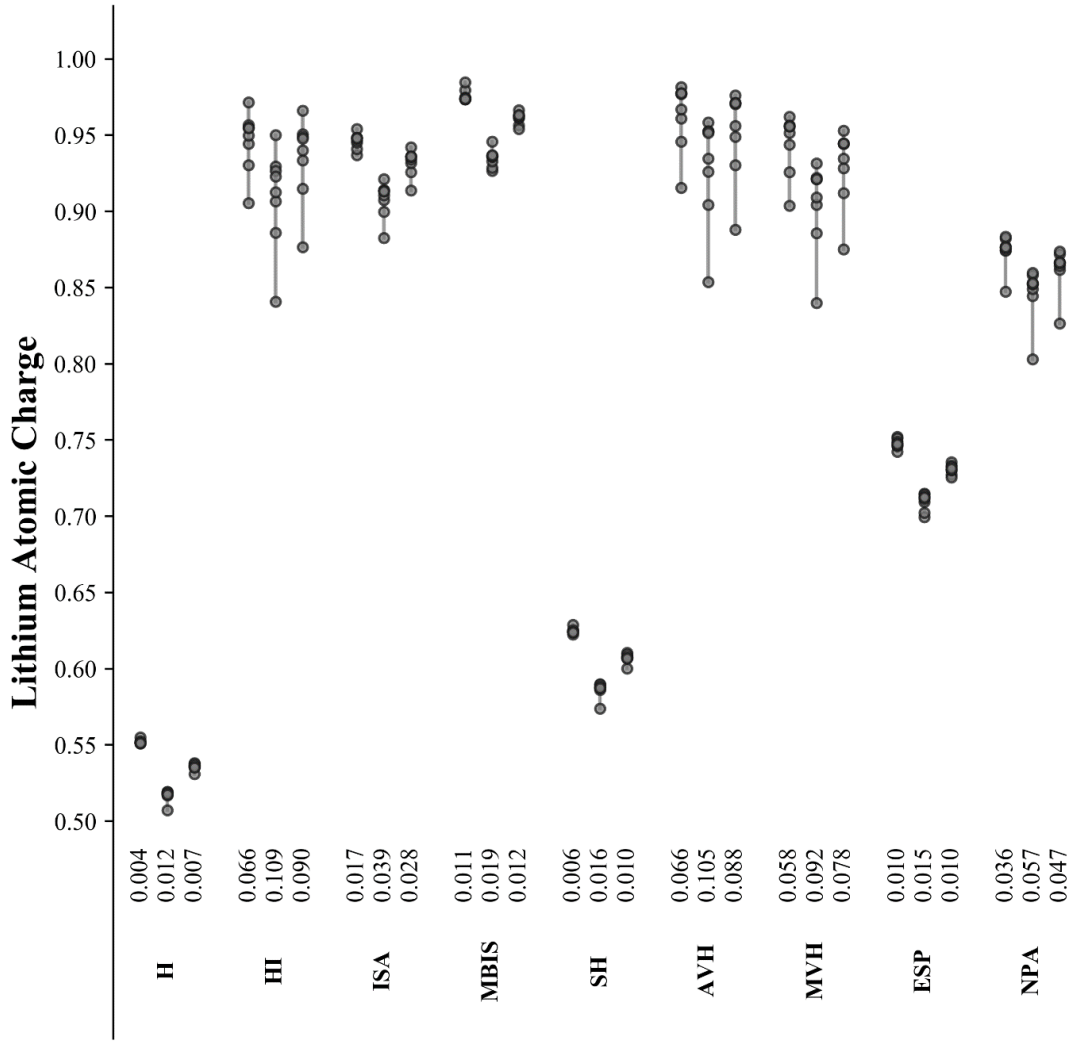


Table 5.1 Additive Variational Hirshfeld (AVH) charge q_{Li} and basis set coefficients $\left\{c_{\mathrm{Li},N,0}\right\}_{i=1}^{4}$ of lithium atom in lithium chloride at various levels of theory. The linear coefficients, from left to right, correspond to spherically averaged ground-state $\rho_{\mathrm{Li}^-}(\mathbf{r})$, $\rho_{\mathrm{Li}}(\mathbf{r})$, $\rho_{\mathrm{Li}^+}(\mathbf{r})$ and $\rho_{\mathrm{Li}^{+2}}(\mathbf{r})$ densities, respectively. The --- indicates that the atomic density was unbound and thus not included in the pro-atom expansion.

Level of Theory	q_{Li}	$c_{\mathrm{Li},4,0}$	$c_{\mathrm{Li},3,0}$	$c_{\mathrm{Li},2,0}$	$c_{\mathrm{Li},1,0}$
UHF/cc-pVDZ	0.9155	---	0.0776	0.9259	0.0000
UHF/cc-pVTZ	0.9457	---	0.0395	0.9679	0.0000
UHF/cc-pVQZ	0.9609	---	0.0224	0.9859	0.0000
UHF/cc-pV5Z	0.9670	---	0.0155	0.9933	0.0000
UHF/aug-cc-pVDZ	0.9817	---	0.0071	0.9985	0.0000
UHF/aug-cc-pVTZ	0.9771	---	0.0041	1.0054	0.0000
UHF/aug-cc-pVQZ	0.9776	---	0.0027	1.0072	0.0000
UHF/aug-cc-pV5Z	0.9776	---	0.0025	1.0074	0.0000
UB3LYP/cc-pVDZ	0.8536	0.0000	0.1376	0.8668	0.0000
UB3LYP/cc-pVTZ	0.9043	0.0000	0.0761	0.9337	0.0000
UB3LYP/cc-pVQZ	0.9261	0.0000	0.0515	0.9596	0.0000
UB3LYP/cc-pV5Z	0.9346	0.0000	0.0403	0.9723	0.0000
UB3LYP/aug-cc-pVDZ	0.9585	0.0004	0.0255	0.9817	0.0000
UB3LYP/aug-cc-pVTZ	0.9524	0.0000	0.0216	0.9914	0.0000
UB3LYP/aug-cc-pVQZ	0.9527	0.0000	0.0195	0.9945	0.0000
UB3LYP/aug-cc-pV5Z	0.9514	0.0000	0.0190	0.9958	0.0000
UωB97XD/cc-pVDZ	0.8879	0.0000	0.1044	0.8995	0.0000
UωB97XD/cc-pVTZ	0.9303	0.0000	0.0538	0.9542	0.0000
UωB97XD/cc-pVQZ	0.9489	0.0000	0.0332	0.9758	0.0000
UωB97XD/cc-pV5Z	0.9561	0.0000	0.0246	0.9850	0.0000
UωB97XD/aug-cc-pVDZ	0.9762	0.0004	0.0101	0.9960	0.0000
UωB97XD/aug-cc-pVTZ	0.9705	0.0000	0.0086	1.0019	0.0000
UωB97XD/aug-cc-pVQZ	0.9712	0.0000	0.0067	1.0043	0.0000
UωB97XD/aug-cc-pV5Z	0.9709	0.0000	0.0066	1.0047	0.0000

Table 5.2 Multiplicative Variational Hirshfeld (MVH) charge q_e and basis set coefficients of the lithium atom in lithium chloride at various levels of theory. These charges were computed considering the cusp constraint and promolecule charge constraint. The nonlinear basis function coefficients $c_{\mathrm{Li},4,0}$, $c_{\mathrm{Li},3,0}$, $c_{\mathrm{Li},2,0}$ and $c_{\mathrm{Li},1,0}$, correspond to the exponents of the spherically averaged ground-state $\rho_{\mathrm{Li}^-}(\mathbf{r})$, $\rho_{\mathrm{Li}}(\mathbf{r})$, $\rho_{\mathrm{Li}^+}(\mathbf{r})$ and $\rho_{\mathrm{Li}^{+2}}(\mathbf{r})$ densities, respectively. (See Eq. (5.15).) The last column presents the pre-factor of lithium pro-atom expansion. The --- indicates that the atomic density was unbound and thus not included in the pro-atom expansion.

Level of Theory	q_{Li}	$c_{\mathrm{Li},4,0}$	$c_{\mathrm{Li},3,0}$	$c_{\mathrm{Li},2,0}$	$c_{\mathrm{Li},1,0}$	$c_{\mathrm{Li},0,0}$
UHF/cc-pVDZ	0.9038	---	0.6221	0.0000	0.3779	1.2848
UHF/cc-pVTZ	0.9258	---	0.6880	0.1022	0.2099	1.1464
UHF/cc-pVQZ	0.9437	---	0.5694	0.2973	0.1333	1.0920
UHF/cc-pV5Z	0.9515	---	0.4892	0.4276	0.0833	1.0574
UHF/aug-cc-pVDZ	0.9621	---	0.4437	0.2668	0.2896	1.2160
UHF/aug-cc-pVTZ	0.9565	---	0.3909	0.6040	0.0051	1.0038
UHF/aug-cc-pVQZ	0.9557	---	0.3853	0.6147	0.0000	1.0009
UHF/aug-cc-pV5Z	0.9557	---	0.3858	0.6119	0.0023	1.0025
UB3LYP/cc-pVDZ	0.8399	0.3180	0.3714	0.0686	0.2420	1.1659
UB3LYP/cc-pVTZ	0.8858	0.5769	0.0753	0.1923	0.1554	1.1061
UB3LYP/cc-pVQZ	0.9042	0.0511	0.5755	0.2530	0.1204	1.0819
UB3LYP/cc-pV5Z	0.9093	0.6148	0.0000	0.2317	0.1535	1.1073
UB3LYP/aug-cc-pVDZ	0.9315	0.1091	0.4131	0.3570	0.1208	1.0813
UB3LYP/aug-cc-pVTZ	0.9213	0.0000	0.5036	0.4951	0.0013	1.0006
UB3LYP/aug-cc-pVQZ	0.9221	0.0000	0.4918	0.5082	0.0000	1.0007
UB3LYP/aug-cc-pV5Z	0.9213	0.0010	0.5200	0.4565	0.0224	1.0162
UωB97XD/cc-pVDZ	0.8750	0.3444	0.3045	0.0605	0.2906	1.2067
UωB97XD/cc-pVTZ	0.9120	0.6249	0.0000	0.2230	0.1521	1.1040
UωB97XD/cc-pVQZ	0.9284	0.0558	0.5416	0.2727	0.1299	1.0893
UωB97XD/cc-pV5Z	0.9346	0.0122	0.5609	0.3070	0.1199	1.0836
UωB97XD/aug-cc-pVDZ	0.9530	0.3221	0.1314	0.3901	0.1564	1.1086
UωB97XD/aug-cc-pVTZ	0.9443	0.0017	0.4425	0.5557	0.0000	0.9998
UωB97XD/aug-cc-pVQZ	0.9444	0.0000	0.4380	0.5620	0.0000	1.0007
UωB97XD/aug-cc-pV5Z	0.9445	0.0020	0.4373	0.5607	0.0000	1.0010

5.3.3 Sensitivity to Divergence Measure

In the previous section we minimized the extended Kullback-Leibler divergence between the molecular and promolecular densities, cf. Eq. (5.4). As discussed in section 5.2.3 however, the method we present works for every *f*-divergence. Some interesting *f*-divergences are listed in the appendix. (In the appendix, we also mention that the result extends even somewhat beyond the class of *f*-divergences.[189]) Here we will focus only on the extended α-divergence,[28, 138-140, 145, 158]

$$D_{\alpha,\mathrm{ext}}\left[\rho_A\middle|\rho_A^0\right]=\frac{1}{\alpha(\alpha-1)}\int\rho_A(\mathbf{r})\left(\frac{\rho_A^0(\mathbf{r})}{\rho_A(\mathbf{r})}\right)^{\alpha}-\rho_A(\mathbf{r})\,d\mathbf{r}\,. \tag{5.41}$$

This generalizes the extended Kullback-Leibler divergence, to which it reduces in the limit $\alpha\to 0$. We will also consider the symmetrized α-divergence, where the densities of the AIM and the pro-atom appear symmetrically,

$$D_{\alpha,\mathrm{sym}}\left[\rho_A\middle|\rho_A^0\right]=\frac{1}{2}\left(D_{\alpha,\mathrm{ext}}\left[\rho_A\middle|\rho_A^0\right]+D_{\alpha,\mathrm{ext}}\left[\rho_A^0\middle|\rho_A\right]\right) \tag{5.42}$$

The symmetrized α-divergence reduces to the symmetrized-Kullback-Leibler divergence (cf. Eq. (5.5)) when $\alpha\to 0$ or $\alpha\to 1$. Clearly the symmetrized α-divergence gives the same answer for $\alpha=\frac{1}{2}\pm\beta$, so we will only consider $\alpha\geq\frac{1}{2}$.

Figure 5.2 shows the dependence of atomic charges for the oxygen, nitrogen, and carbon AIM in formamide, $HCONH_2$, on the value of α in the extended α-divergence. The dependence on the basis set is unremarkable, and the charges of the oxygen and carbon atoms are rather insensitive to the value of α. However, the charges of the nitrogen atom vary by about $\sim\pm 0.15$, depending on the value of α. The α-dependence is mostly

eliminated when one uses the symmetrized α-divergence as presented in Figure 5.3: now the carbon and oxygen atoms are even more insensitive to the choice of α, and the nitrogen atoms charge varies by only $\sim \pm 0.03$. Moreover, the charges for the symmetrized α-divergence are very close to the charges associated with the Hellinger distance ($\alpha = \frac{1}{2}$), and relatively close to the charges associated with the extended Kullback-Leibler divergence. Since the extended Kullback-Leibler ($\alpha \to 0$) and symmetrized α-divergence give similar answers, but only the extended Kullback-Leibler divergence is size-consistent (cf. Eq. (5.35)) and gives atoms and pro-atoms with the same charges (cf. Eq. (5.34)). As we have found no compelling reason to use any divergence other than the extended Kullback-Leibler divergence in our numerical explorations, we will henceforth use only that measure. However, it is worth noting that the symmetrized α-divergence (and we believe, symmetrized *f*-divergences in general) will generally be less sensitive to the choice of internal parameters. The symmetrized formulas also seem to be slightly less sensitive to basis set.

Figure 5.2 Dependence of Additive Variational Hirshfeld (AVH) atomic charges of oxygen, nitrogen and carbon in formamide on the α value when the extended α-divergence measure, (5.41), was used for optimizing the pro-atoms. The molecular and pro-atom densities were computed with unrestricted Hartree-Fock calculations using Dunning basis sets. Similar results were obtained for UB3LYP and UωB97XD levels of theory using Dunning basis sets.

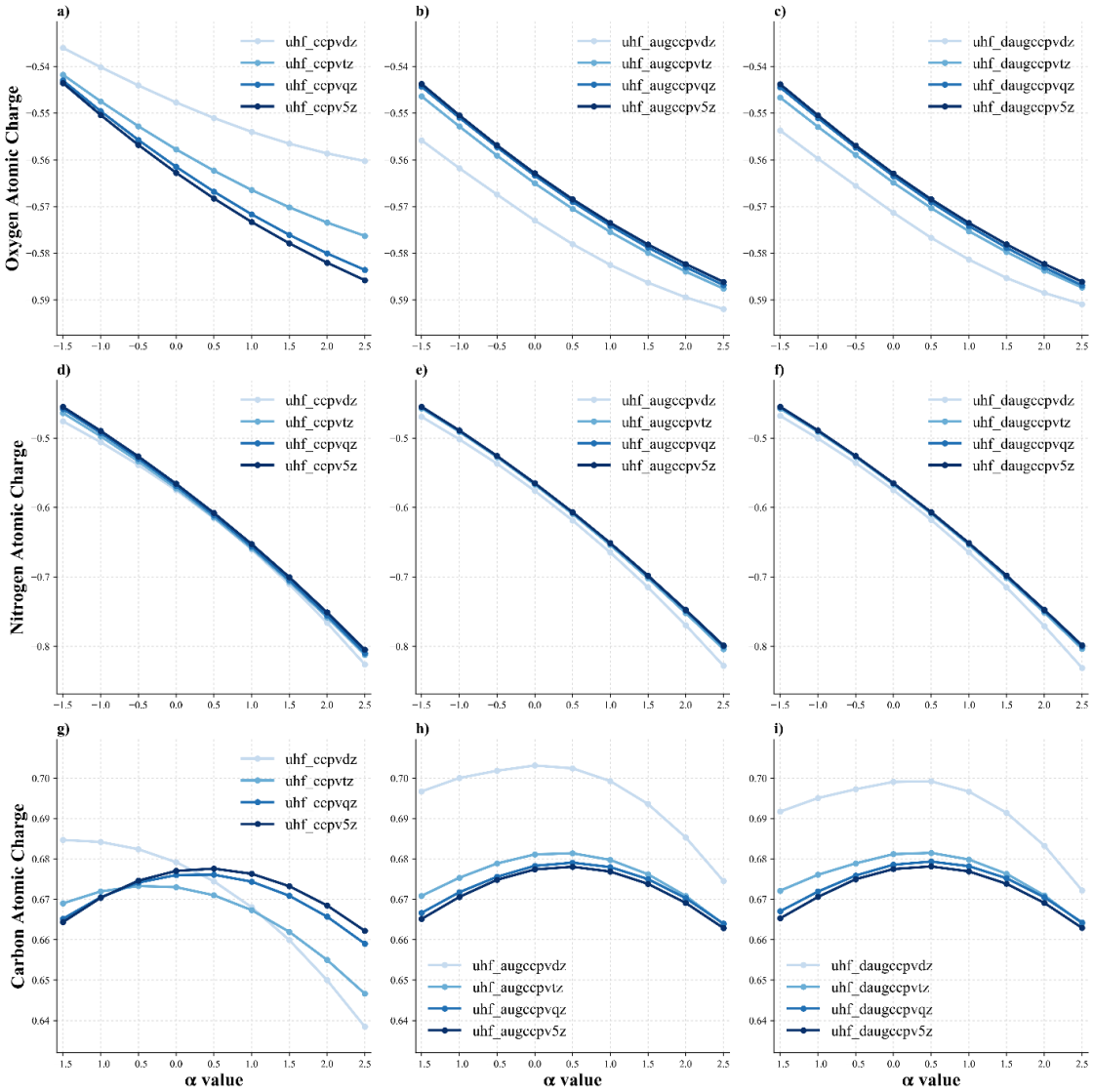
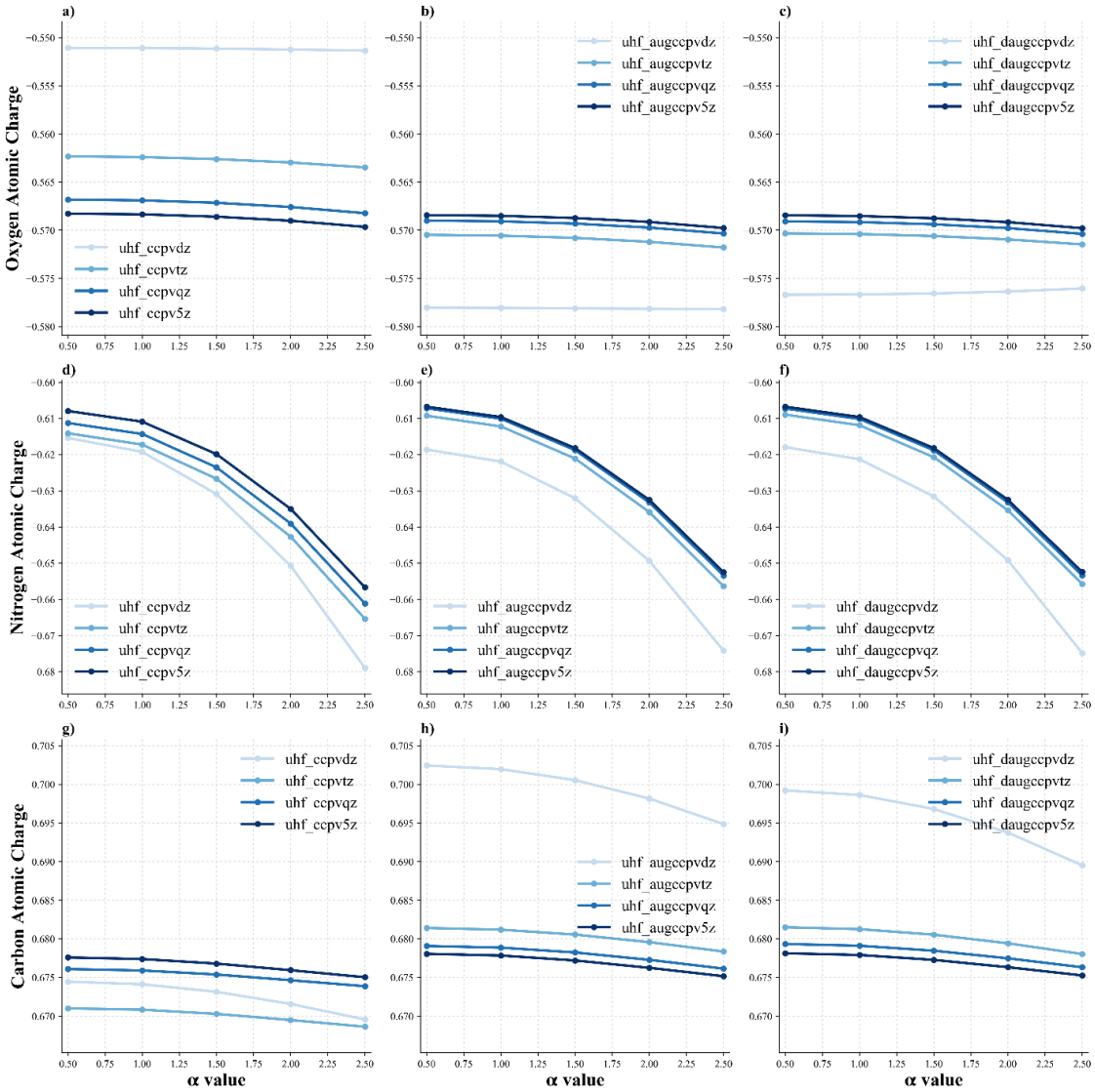


Figure 5.3 Dependence of Additive Variational Hirshfeld (AVH) atomic charges of oxygen, nitrogen and carbon in formamide on the α value when the symmetric α-divergence measure, (5.42), was used for optimizing the pro-atoms. The molecular and pro-atom densities were computed with unrestricted Hartree-Fock calculations using Dunning basis sets. Similar results were obtained for UB3LYP and UωB97XD levels of theory using Dunning basis sets.



5.3.4 Cusp and Ionization Potential Constraints

As discussed in section 5.2.7, one can force the pro-atomic densities in the multiplicative model to have appealing mathematical properties by imposing constraints. Forcing the pro-atom densities to satisfy the correct nuclear cusp constraint, Eq. (5.38), enables the promolecular density to closely fit the molecular density in the vicinity of the nucleus. When we implemented the cusp constraint, we observed that the charges on heavy atoms did not change dramatically. However, negatively-charged hydrogen atoms tended to become less negative. For example, in ammonia, the charge on the nitrogen atom in a d-aug-cc-pV5Z calculation is -0.927 without the cusp constraint and -0.573 with the cusp constraint. (The overall trends are very similar, with nitrogen charges in NH_3 tightly clustered around -0.9 without the cusp constraint around -0.59 with the cusp constraint, regardless of the basis set (d-aug-)cc-pVXZ (X=D, T, Q, 5) or method (UHF, UB3LYP, UωB97XD).) Comparing to the results from electrostatic potential (ESP) fitting, -0.866, one might believe that the cusp constraint is detrimental to the MVH method. However, as we shall discuss in the next section, the ESP charges on the nitrogen atom in ammonia are generally too negative, so the fact the unconstrained MVH gives even more negative charges is unfavorable. The MVH with the cusp constraint gives results very similar to AVH. (For example, the population on the nitrogen in NH_3 at the UB3LYP/d-aug-cc-pV5Z level from AVH differs from that of cusp-constrained MVH by just 0.0006 electrons).

As discussed in section 5.2.7, one can also require that the pro-atoms all have the same asymptotic decay. This constraint is conceptually problematic as the resulting

method is not size consistent, since in a molecule consisting of two well-separated fragments, F and G, the asymptotic decays of the pro-atoms for fragment F are nonetheless constrained based on the asymptotic decays of the pro-atoms for fragment G. This is in accord with the electronegativity equalization principle and the paradoxes associated thereto,[51-52] but it is computationally and intuitively problematic. It is even more problematic to note that the constraint that all the pro-atoms have the same asymptotic decay, Eq. (5.39), cannot even be satisfied in many cases (e.g, any molecule containing both nitrogen and hydrogen atoms). Nonetheless, in cases where we were able to impose the asymptotic decay constraint, the results obtained from it were acceptable, although perhaps not ideal. For example, for magnesium oxide computed at the UB3LYP/aug-cc-pVTZ level, the charge on magnesium changes from +0.824 to +0.713 when adding the asymptotic constraint. Arguably the charges without the asymptotic decay constraint are more reasonable since they are closer to the expected formal charge (Mg^{+2}), but perhaps not enough to dismiss the asymptotic constraint as unworthy of further study. Based on our investigations, the asymptotic constraint has minimal influence in many cases, but in highly ionic systems like MgO, LiCl, etc., the magnitude of the charges decreases upon imposition of the asymptotic constraint.

As we discussed in section 5.3.2, the parameters in MVH are difficult to interpret in general. The MgO is no exception. Without the asymptotic constraints, the exponents for the Mg^0 and Mg^{+1} electron densities are $c_{\mathrm{Mg},12,0} = 0.55$ and $c_{\mathrm{Mg},11,0} = 0.42$ respectively. (All other states, including the intuitive Mg^{+2} state, have exponents less that 0.04.) Similarly, the exponents for the O^- and O^0 electron densities are $c_{\mathrm{O},9,0} = 0.60$ and

$c_{\mathrm{O},8,0} = 0.40$, respectively. Imposing the asymptotic constraint on the pro-atom densities forces the Mg pro-atom to decay more slowly and the O pro-atom to decay more quickly ($c_{\mathrm{Mg},12,0} = 0.86;\ c_{\mathrm{Mg},11,0} = 0.09;\ c_{\mathrm{O},9,0} = 0.22;\ c_{\mathrm{O},8,0} = 0.78$). This is in accord with chemical intuition and the trend for the asymptotic constraint to diminish the magnitude of the charges in small molecules. It also agrees with the observation in section 5.3.2 that large changes in the exponents in the multiplicative pro-atom model often have surprisingly little influence on the computed atomic charges.

5.3.5 Basis Set and Method Dependence

In **Error! Reference source not found.**, the charges of the carbon, nitrogen, and xygen AIM in the formamide molecule, computed at various levels of theory for a variety of population analysis methods, are presented. As before the MVH results include the cusp constraint, Eq. (5.38), but not the ionization potential constraint, Eq. (5.39), which is inapplicable since formamide has both nitrogen and hydrogen atoms. The Hirshfeld and scaled-Hirshfeld, Eq. (5.30), charges are smaller in magnitude than charges from the other methods. Hirshfeld-I charges are very sensitive to basis set, which is typical for Nitrogen-containing elements. Unfortunately, the AVH and MVH methods are also unreasonably sensitive to basis set. Examining the data, we realized this occurred because for some choices of method and basis set, some atomic anions were not bound. This suggested that we should restrict to only allow the electron densities of atomic states that are actually physically bound, thereby eliminating the erroneous electron densities associated with "computationally bound" systems like nitrogen anions. The AVH-PHYS and MVH-

PHYS methods do exactly this. Using only the atomic densities of the physically bound atomic ions to compose the pro-atom densities removes most of the basis-set dependence. To determine which atomic charge states were stable, we used the atoms with stable anions tabulated in ref [190].

It is also apparent that the nitrogen atom in the AVH and MVH methods is less negative than it is in the other methods (except, obviously, for Hirshfeld and scaled-Hirshfeld partitioning). This motivated us to explore a broader set of nitrogen-containing molecules (including molecular ions), which we shall do in section 5.3.6)

To compare the performance of the AVH and MVH methods to more established approaches, we selected a set of small molecules (CH_3^+, CH_4, CH_3^-, NH_4^+, NH_3, NH_2^-, H_3O^+, H_2O, OH^-) for investigating the sensitivity of these methods to the one-electron basis set and the type of electronic structure theory method used, and also for assessing how well different partitioning methods recapture chemical trends. The results are presented in Figure 5.5. As before only allowing the pro-atoms to be constructed using physically bound atomic densities removes most of the basis-set dependence. The resulting methods, labeled AVH-PHYS and MVH-PHYS in Figure 5.5, show good stability with respect to method and basis, and appear to obey the main chemical trends, namely that as the molecular charge increases, the atomic charge of the central non-hydrogen atom should also increase. AVH-PHYS and MVH-PHYS give very similar results, with perhaps slightly better basis-set-insensitivity from the multiplicative pro-atom model (MVH). Notice that AVH and MVH cure the erratic behavior for molecular anions of the closely related MBIS method.

AVH-PHYS and MVH-PHYS give less negative (more positive) charges for the central non-hydrogen atoms in these molecules. This may be desirable. For example, in charges based on fitting the molecular electrostatic potential, the charge on the nitrogen atom in ammonia, NH_3, is generally believed to be too negative, because the ESP-fitting charges are based on putting a charge at the position of the nitrogen nucleus, while the locus of negative charge on the nitrogen atom is associated with the lone pair. It is the favorable that the AVH-PHYS and MVH-PHYS charges on the nitrogen in ammonia are less negative than the charges from electrostatic fitting. It is disconcerting that MBIS, Hirshfeld-I, NPA, and ISA predict charges on the ammonia atom that are even more negative than those predicted by electrostatic fitting.

The lithium nitride molecule in Figure 5.6 shows the same trends as the ammonia molecule, but with greater severity. It is extremely clear that it is critical to avoid atomic densities that are computationally but not physically bound in the AVH and MVH methods. Moreover, the pathologies of the Hirshfeld-I are apparent, not only from the extreme basis-set dependence, but also because our software crashed in some cases because once the nitrogen atom has a charge smaller than -3, one needs the electron density of N^{-4}, which is not available in our database of pro-atoms. MBIS and ISA give charges even more negative than the (presumably already too negative) charges from ESP fitting. The charges on the nitrogen atoms predicted by AVH and MVH, by contrast, seem plausible (though slightly more negative charges would also be reasonable).

An unpleasant feature of AVH and MVH is the prediction of a small positive charge for the central carbon in methane in Figure 5.5. This is an outlier among population

analysis methods, and we feel it may be possible to remove this behavior by including excited-state atomic densities in the pro-atom basis. In fact, the realization that a contribution from the sp^3 configuration of the carbon atom should be included in the carbon pro-atom of aliphatic hydrocarbons was the original motivation for the excited-state AVH method.

Figure 5.4 Atomic charges of oxygen, nitrogen and carbon in formamide computed with various partitioning schemes at various levels of theory. For each scheme, the three columns plot the charges computed using twelve Dunning basis sets, i.e. (d-aug-)cc-pVXZ with X=D, T, Q, 5 basis functions, at UHF, UB3LYP, and UωB97XD levels of theory, respectively. The absolute range of the atomic charges obtained using various basis sets at each level of theory is summarized on the *x*-axis alongside the name of partitioning method. The methods used are Hirshfeld (H), Iterative Hirshfeld (HI), Iterative Stockholder Analysis (ISA), Minimal Basis Stockholder Analysis (MBIS), Scaled Hirshfeld (SH), Additive Variational Hirshfeld with computationally bound proatom basis (AVH), Additive Variational Hirshfeld with physically bound proatom basis (AVH-PHYS), Multiplicative Variational Hirshfeld with computationally bound proatom basis (MVH), Multiplicative Variational Hirshfeld with physically bound proatom basis (MVH-PHYS), Hu-Lu-Yang electrostatic fitted charges (ESP), and Natural Population Analysis (NPA).

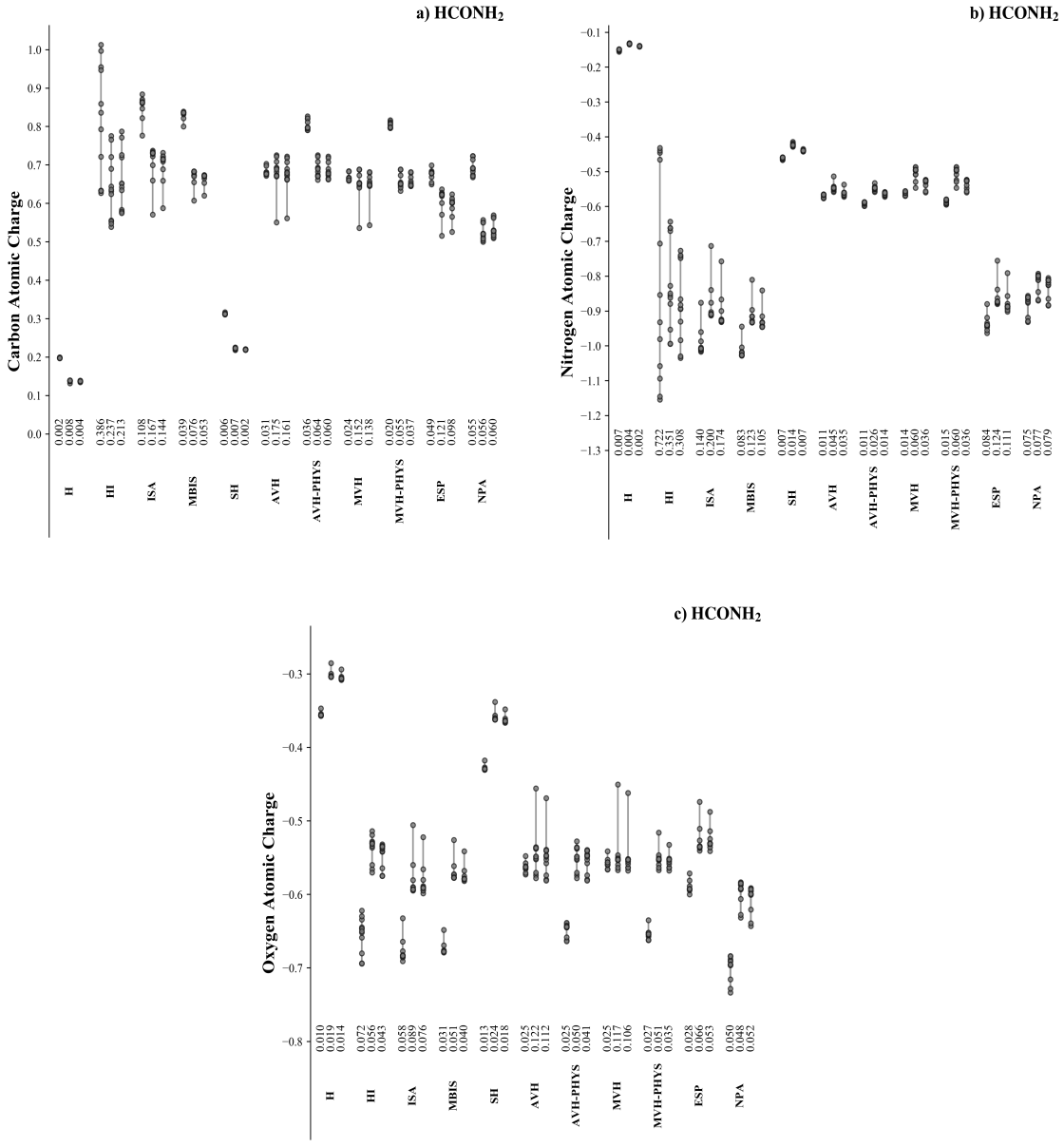


Figure 5.5 Comparison between partitioning schemes at different levels of theory. For each scheme, the three columns plot the charges computed using twelve Dunning basis sets, i.e. (d-aug-)cc-pVXZ with X=D, T, Q, 5 basis functions, at UHF, UB3LYP, and UωB97XD levels of theory, respectively. The absolute range of the atomic charges obtained using various basis sets at each level of theory is summarized on the *x*-axis alongside the name of partitioning method. The methods used are Hirshfeld (H), Iterative Hirshfeld (HI), Iterative Stockholder Analysis (ISA), Minimal Basis Stockholder Analysis (MBIS), Scaled Hirshfeld (SH), Additive Variational Hirshfeld with computationally bound pro-atom basis (AVH), Additive Variational Hirshfeld with physically bound pro-atom basis (AVH-PHYS), Multiplicative Variational Hirshfeld with computationally bound pro-atom basis (MVH), Multiplicative Variational Hirshfeld with physically bound pro-atom basis (MVH-PHYS), Hu-Lu-Yang electrostatic fitted charges (ESP), and Natural Population Analysis (NPA).

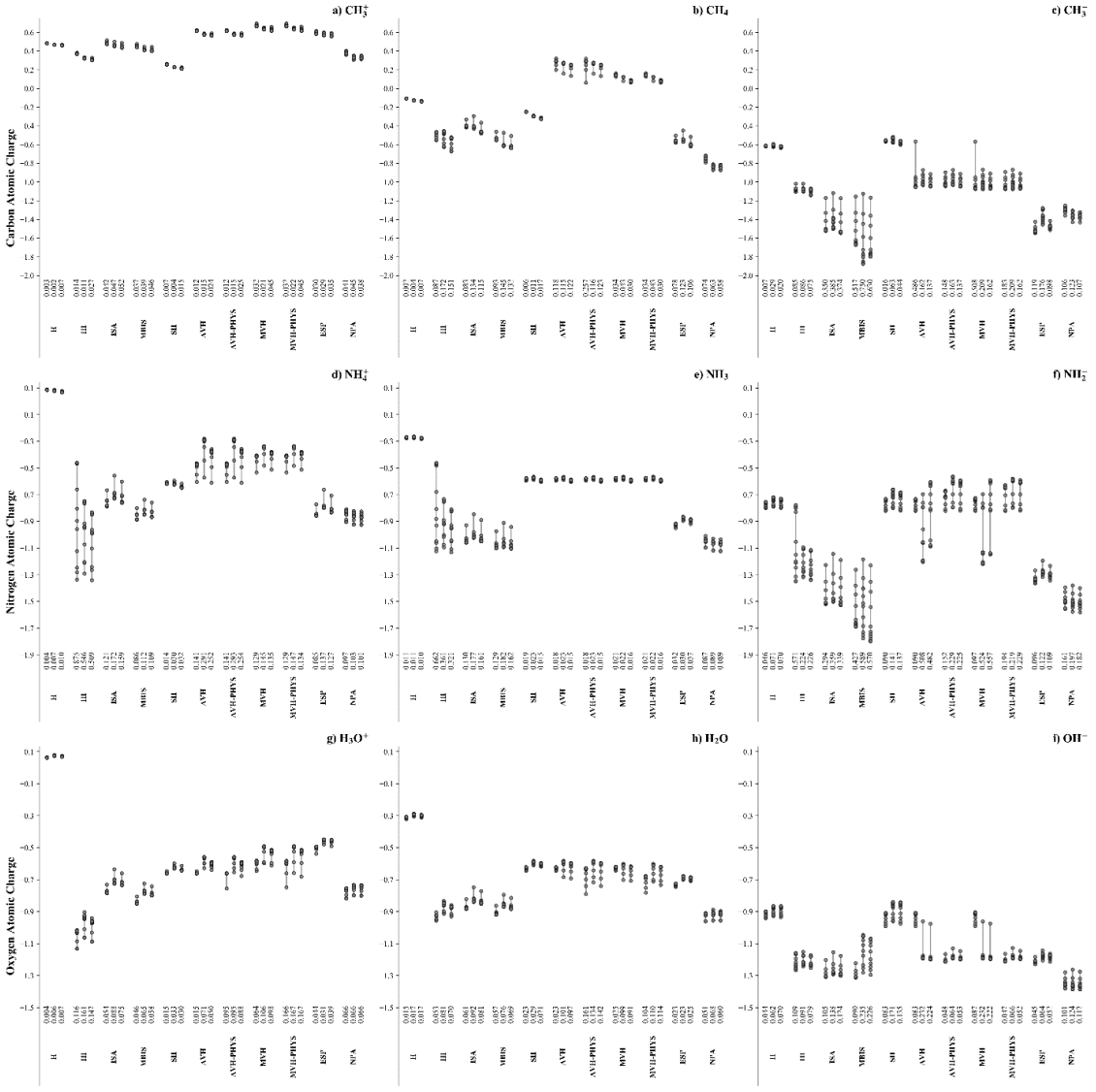
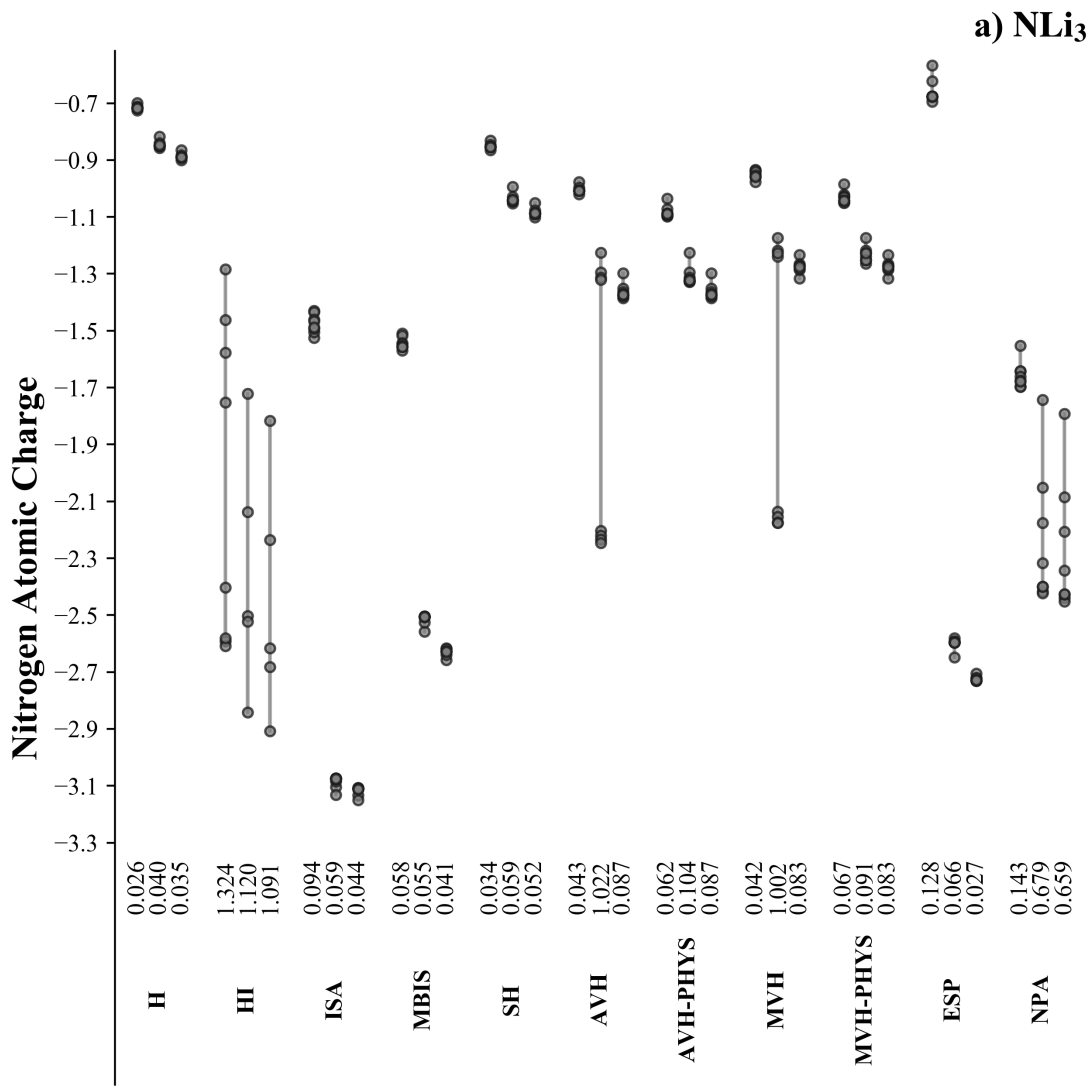


Figure 5.6 Atomic charge of the nitrogen atom in lithium nitride computed with various partitioning schemes at various levels of theory. For each scheme, the three columns plot the charges computed using twelve Dunning basis sets, i.e. (d-aug-)cc-pVXZ with X=D, T, Q, 5 basis functions, at UHF, UB3LYP, and UωB97XD levels of theory, respectively. The absolute range of the atomic charges obtained using various basis sets at each level of theory is summarized on the *x*-axis alongside the name of partitioning method. The methods used are Hirshfeld (H), Iterative Hirshfeld (HI), Iterative Stockholder Analysis (ISA), Minimal Basis Stockholder Analysis (MBIS), Scaled Hirshfeld (SH), Additive Variational Hirshfeld with computationally bound pro-atom basis (AVH), Additive Variational Hirshfeld with physically bound pro-atom basis (AVH-PHYS), Multiplicative Variational Hirshfeld with computationally bound pro-atom basis (MVH), Multiplicative Variational Hirshfeld with physically bound pro-atom basis (MVH-PHYS), Hu-Lu-Yang electrostatic fitted charges (ESP), and Natural Population Analysis (NPA).



5.3.6 Comparison to Conventional Population Analysis Methods

To explore the trends in the AVH and MVH charges relative to other methods, we considered a set of 41 nitrogen-containing molecules listed in Table 5.3. Figure 5.7 to Figure 5.14 compare the trends between various population analysis methods for the atoms that appear in these molecules, with the charges computed using three different quantum chemistry methods (UHF, UB3LYP, UωB97XD) and up to twelve different basis sets (d-aug-, aug-)cc-pVXZ (X=D,T,Q,5), depending on the availability of the basis sets for the elements in each molecule. Based on the results from the previous section, from now on we only use physically bound atomic densities when composing the pro-atoms. We use the cusp constraint for the multiplicative pro-atom model, Eq. (5.38).

Looking at Figure 5.7, it is remarkable how close the MVH and AVH models are. (A corollary to this finding would be that an even more general method, using the *p*-mean pro-atom formula in Eq. (5.12), is unlikely to significantly affect the molecular populations.) Since the AVH method is a convex optimization, and is therefore more robust and easier computationally, it seems preferable to the MVH method.

Figure 5.8 and Figure 5.9 compare AVH atomic charges to the conventional Hirshfeld and scaled Hirshfeld charges. The (scaled-)Hirshfeld method is based on neutral pro-atom densities, and tends to underestimate the magnitude of the atomic charges. The AVH method correlates well with the Hirshfeld and scaled-Hirshfeld charges, but the AVH charges tend to be larger.

Figure 5.10 compares charges from AVH and from Hirshfeld-I charges. Hirshfeld-I charges are usually excellent when they are relatively small, but because the reference

atomic densities of dianions (and even trianions) are unphysical, Hirshfeld-I charges are erratic for molecules containing very negatively charged atoms. For the nitrogen atom, AVH charges closely correlate with Hirshfeld-I charges when the nitrogen atom has a charge greater than −1, but do not decrease to arguably absurd levels thereafter. In this set of molecules, Hirshfeld-I also sometimes gives charges on carbon atoms and lithium atoms that are greater than +1. This seems questionable, especially for lithium atoms. It is reassuring that AVH never gives charges of lithium or carbon greater than +1 for this molecule set.

Figure 5.11 compares charges from AVH and from iterative stockholder analysis (ISA). As with Hirshfeld-I charges, ISA charges tend to be reliable when the magnitude of the charges is relatively small, and then unreliable when the charges become extreme. It is therefore reassuring that AVH charges are very similar to ISA charges for charges between +1 and −1, but less extreme outside this interval. Note that the ISA charges give some absurd results: nitrogen atoms are even more negative than −3, lithium atoms that have charges greater than +1, carbon atoms with charges greater than +1. None of these questionable behaviors is present for AVH. ISA seems to almost always assign a charge near +1 to Lithium atoms, regardless of the chemical context. AVH seems to be more nuanced in this respect (recall the data for NLi_3 in Figure 5.6).

Figure 5.12 compare charges from AVH and from minimal basis iterative stockholder analysis (MBIS). The charges correlate closely, as might be expected since MBIS is also a variational Hirshfeld method. As with Hirshfeld-I and ISA, the charges correlated well when the magnitude of the charges is not too large, but the extreme

negative nitrogen atom charges are not present (although in general more negative AVH nitrogen atoms correspond to more negative MBIS nitrogen atoms). MBIS has a few highly positive carbon atoms, which AVH does not show.

Figure 5.13 compares charge from AVH and from natural population analysis (NPA). The trends are clearly similar, but the correlation is weak. The trends of AVH charges compared to the charges from electrostatic potential fitting (ESP) are similar as plotted in Figure 5.14. In both cases AVH give fewer extremely negative nitrogen atoms, more negative hydrogen atoms, and a greater range of possible lithium charges.

Overall, these figures demonstrate that the variational Hirshfeld models presented in this thesis give reasonable results. Their results are quite similar to those of other information-theoretic partitioning methods, but improve over scaled-Hirshfeld and conventional Hirshfeld methods by having charges of higher magnitude, improve over Hirshfeld-I, ISA, and MBIS by having moderating extremely negative charges (for nitrogen) and extremely positive charges (for lithium).

Table 5.3 The chemical formula, PubChem Compound Identifier (CID), and International Union of Pure and Applied Chemistry (IUPAC) name of the nitrogen-containing molecules explored in section 5.3.6.

	Formula	CID	IUPAC Name
1	H_3N	222	Azane
2	H_4N^+	223	Azanium
3	H_2N^-	2826723	Azanide
4	ClH_2N	25423	Chloramine
5	Cl_2HN	76939	Dichloroamine
6	Cl_3N	61437	Nitrogen Trichloride
7	FH_2N	139987	Monofluoroamine
8	F_2HN	25242	Difluoroamine
9	F_3N	24553	Nitrogen Trifluoride
10	F_4N^+	---	Perfluoroammonium Cation
11	H_2LiN	24532	Lithium;Azanide
12	HLi_2N	---	Lithium Imide
13	Li_3N	---	Lithium Nitride
14	H_3NO	787	Hydroxylamine
15	H_2N_2	123195	Diazene
16	H_4N_2	9321	Hydrazine
17	CHNO	521293	Formonitrile Oxide
18	CHN	768	Formonitrile
19	CHN	6432654	Methanidylidyneazanium
20	CN^-	5975	Cyanide
21	CNO^-	105034	Cyanate
22	$CNO^\bullet$	140912	λ^2-azanylidenemethanone
23	CHNO	6347	Isocyanic Acid
24	CNO^-	12360	Oxidoazaniumylidynemethane
25	CHNO	62317	Hydroxyazaniumylidynemethane
26	ClNO	17601	Nitrosyl Chloride
27	$ClNO_2$	---	Chloro(oxo)azane Oxide
28	HNO	945	Nitroxyl
29	NO+	84878	Azanylidyneoxidanium
30	$NO^\bullet$	145068	Nitric Oxide
31	NO^-	3001380	Nitroxyl Anion
32	NO_2^+	3609161	Nitronium
33	$NO_2^\bullet$	3032552	Nitrogen Peroxide
34	NO_2^-	946	Nitrite
35	HNO_2	24529	Nitrous Acid
36	FNO_2	66203	Nitryl Fluoride
37	HNO_3	944	Nitric Acid
38	$LiNO_3$	10129889	Lithium;Nitrate
39	HNO_3	123349	Hydroxy Nitrite
40	NO_3^-	104806	Oxido Nitrite
41	N_2O	948	Nitrous Oxide

Figure 5.7 For the molecules from Table 5.3, the atomic charges computed using the additive variational Hirshfeld method with physically bound pro-atom basis (AVH-PHYS) are plotted versus the charges computed using the multiplicative variational Hirshfeld method with physically bound pro-atoms (MVH-PHYS). The charges were computed from molecular densities obtained from 36 calculations, from the twelve Dunning basis sets—(d-aug-)cc-pVXZ; with X=D, T, Q, 5—at the UHF, UB3LYP, and UωB97XD levels of theory.

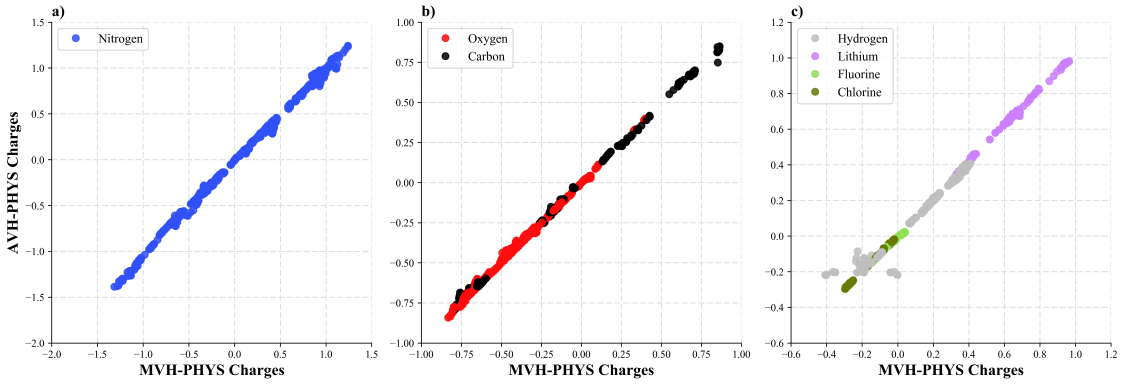


Figure 5.8 For the molecules from Table 5.3, the atomic charges computed using the additive variational Hirshfeld method with physically bound pro-atom basis (AVH-PHYS) are plotted versus the charges computed using the conventional Hirshfeld method with neutral pro-atoms (H). The charges were computed from molecular densities obtained from 36 calculations, from the twelve Dunning basis sets—(d-aug-)cc-pVXZ; with X=D, T, Q, 5—at the UHF, UB3LYP, and UωB97XD levels of theory.

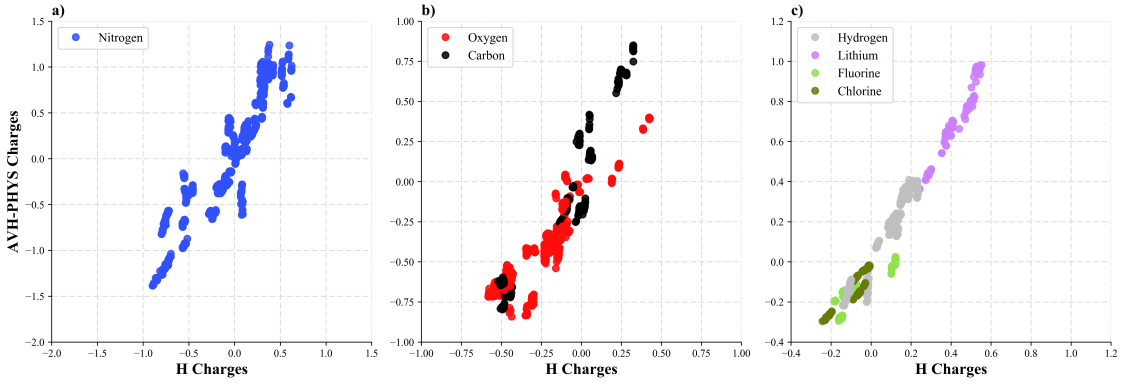


Figure 5.9 For the molecules from Table 5.3, the atomic charges computed using the additive variational Hirshfeld method with physically bound pro-atom basis (AVH-PHYS) are plotted versus the charges computed using the scaled Hirshfeld method with scaled neutral pro-atoms (SH). The charges were computed from molecular densities obtained from 36 calculations, from the twelve Dunning basis sets—(d-aug-)cc-pVXZ; with X=D, T, Q, 5—at the UHF, UB3LYP, and UωB97XD levels of theory.

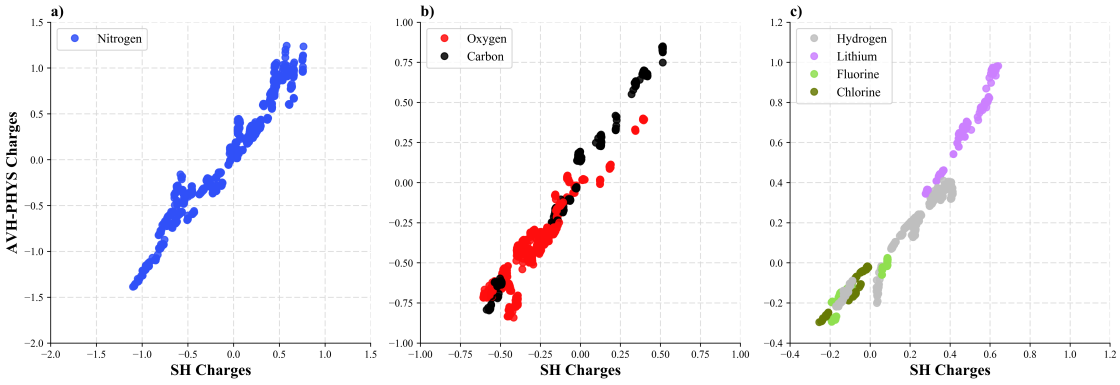


Figure 5.10 For the molecules from Table 5.3, the atomic charges computed using the additive variational Hirshfeld method with physically bound pro-atom basis (AVH-PHYS) are plotted versus iterative Hirshfeld charges (HI). The charges were computed from molecular densities obtained from 36 calculations, from the twelve Dunning basis sets—(d-aug-)cc-pVXZ; with X=D, T, Q, 5—at the UHF, UB3LYP, and UωB97XD levels of theory.

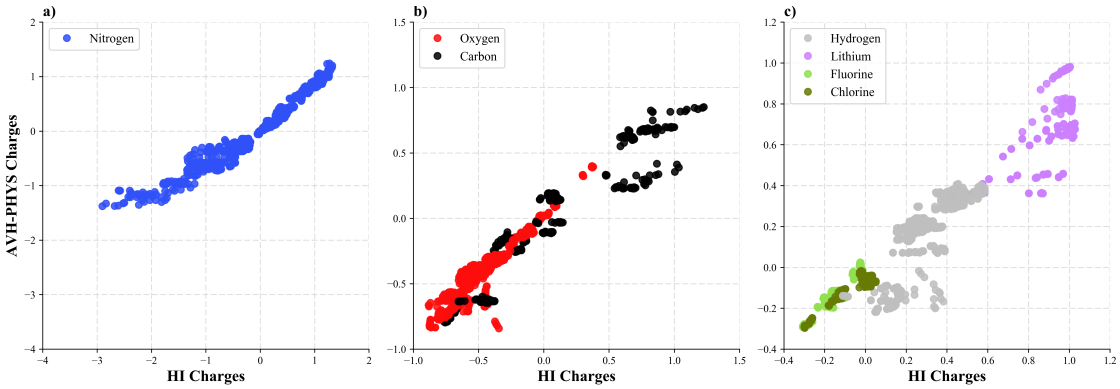


Figure 5.11 For the molecules from Table 5.3, the atomic charges computed using the additive variational Hirshfeld method with physically bound pro-atom basis (AVH-PHYS) are plotted versus charges computed using iterative stockholder analysis (ISA). The charges were computed from molecular densities obtained from 36 calculations, from the twelve Dunning basis sets—(d-aug-)cc-pVXZ; with X=D, T, Q, 5—at the UHF, UB3LYP, and UωB97XD levels of theory.

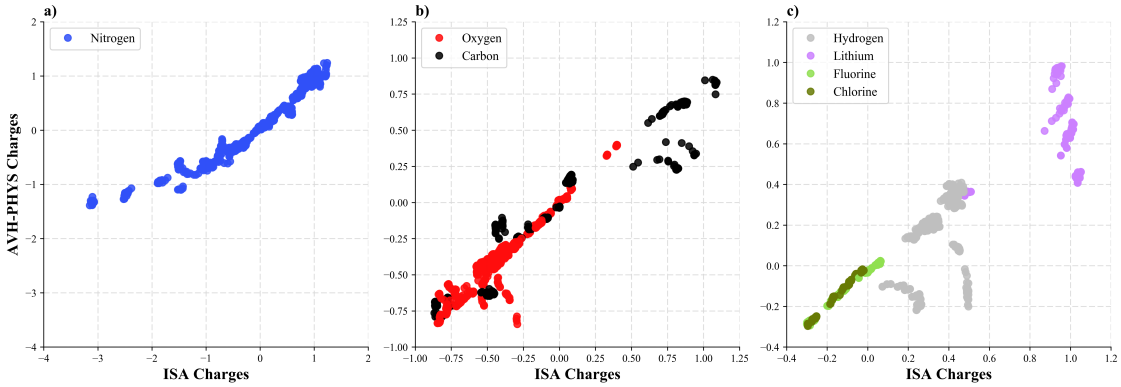


Figure 5.12 For the molecules from Table 5.3, the atomic charges computed using the additive variational Hirshfeld method with physically bound pro-atom basis (AVH-PHYS) are plotted versus minimal basis iterative stockholder charges (MBIS). The charges were computed from molecular densities obtained from 36 calculations, from the twelve Dunning basis sets—(d-aug-)cc-pVXZ; with X=D, T, Q, 5—at the UHF, UB3LYP, and UωB97XD levels of theory.

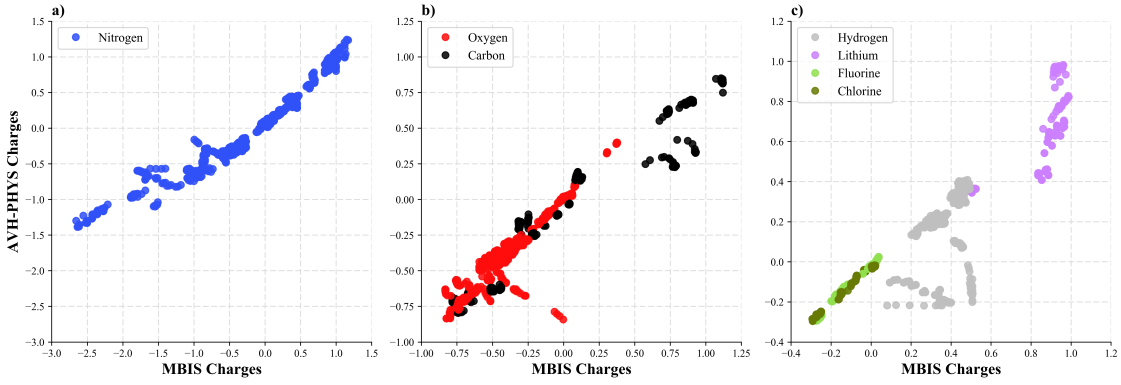


Figure 5.13 For the molecules from Table 5.3, the atomic charges computed using the additive variational Hirshfeld method with physically bound pro-atom basis (AVH-PHYS) are plotted versus the atomic charges from natural population analysis (NPA). The charges were computed from molecular densities obtained from 36 calculations, from the twelve Dunning basis sets—(d-aug-)cc-pVXZ; with X=D, T, Q, 5—at the UHF, UB3LYP, and UωB97XD levels of theory.

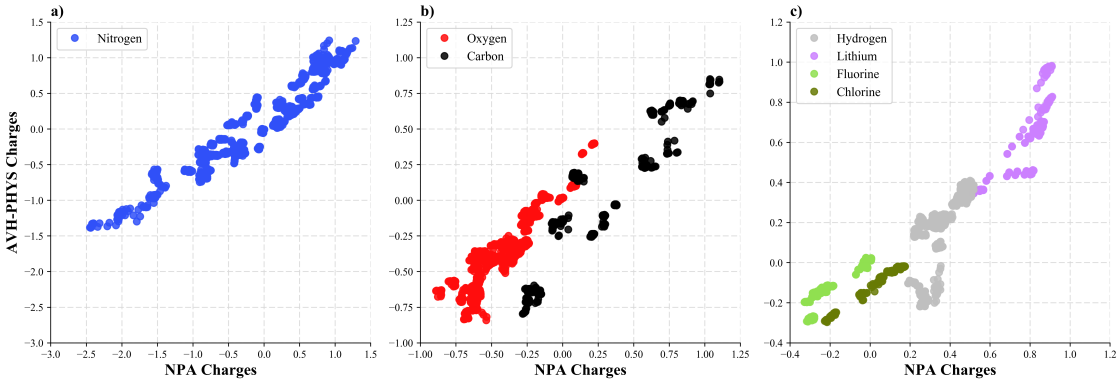
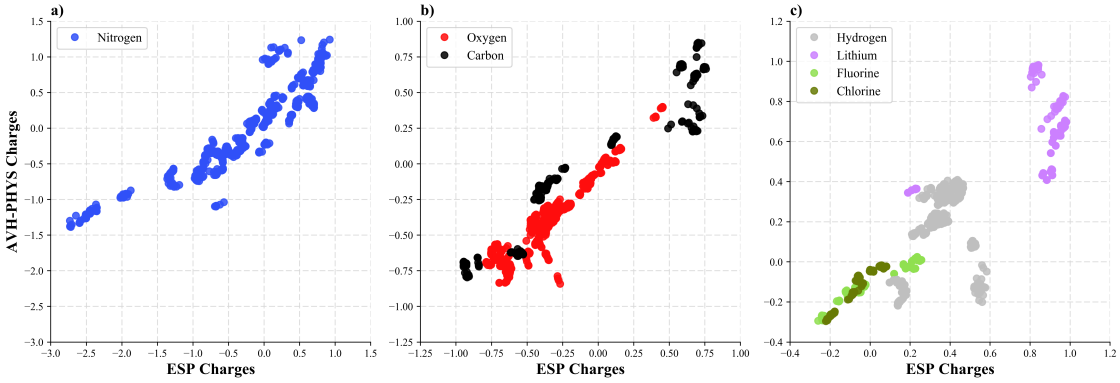


Figure 5.14 For the molecules from Table 5.3, the atomic charges computed using the additive variational Hirshfeld method with physically bound pro-atom basis (AVH-PHYS) are plotted versus the atomic charges obtained by fitting the molecular electrostatic potential using the Hu-Lu-Yang method (ESP). The charges were computed from molecular densities obtained from 36 calculations, from the twelve Dunning basis sets—(d-aug-)cc-pVXZ; with X=D, T, Q, 5—at the UHF, UB3LYP, and UωB97XD levels of theory.



5.4 Conclusion

In this chapter, we presented a variational Hirshfeld method, where one minimizes the *f*-divergence between the molecular density and its approximate, called the promolecular density. The form of the promolecular density is flexible, but we have proposed that it be written as a sum of atomic contributions, Eq. (5.8), where the atomic contributions are composed from the spherically-averaged ground state and/or excited state densities of the isolated atoms and atomic ions. We observed that including only the densities that correspond to physically bound ground states made these methods relatively insensitive to the choice of method and basis set. We presented three ways of composing the pro-atom densities:

(a) A weighted arithmetic average, Eq. (5.11), we call this the additive pro-atom model and call the resulting partitioning method additive variational Hirshfeld (AVH).

(b) A weighted geometric average, Eq. (5.15), we call this the multiplicative pro-atom model and call the resulting partitioning method multiplicative variational Hirshfeld (MVH).

(c) A scaled Hirshfeld method where only the contribution from the neutral atom's density is used, Eq. (5.30).

Because all three methods generally employ the electron densities of atomic anions, we observed that good results generally require diffuse functions and a triple-zeta basis set. We therefore recommend using basis sets no worse than aug-cc-pVTZ, Def2-TZVPD, or 6-311++G(2df,2pd).

We discussed how one can use all three of these methods for any *f*-divergence. If, however, one uses the extended Kullback-Leibler divergence, Eq. (5.4), these three methods are all size consistent (cf. Eq. (5.35)) and have equal atomic and pro-atom charges, Eq. (5.34). Other choices of the *f*-divergence do not have these appealing properties, and often give rather similar atomic charges.

The scaled Hirshfeld method is interesting primarily as an initial guess for the AVH and MVH methods, but it also is arguably the most straightforward way to extend the conventional Hirshfeld method to molecular ions. Like conventional Hirshfeld charges, the scaled Hirshfeld charges seem to be systematically too small in size. The AVH and MVH methods give results that agree, broadly, with chemical intuition, chemical trends, and the charges obtained by other, more established, methods.

Unlike the AVH method, which is a simple convex optimization, the objective function in MVH is generally nonconvex. The AVH coefficients often lend themselves to a chemical interpretation, but the MVH coefficients are relatively sensitive to method and basis set, and are difficult, if not impossible, to interpret. However, the MVH (with the additional constraint that the pro-atom densities have the correct electron-nuclear cusp) and AVH give almost the same results, especially for sufficiently large basis sets. This suggests that the AVH method, using the extended Kullback-Leibler divergence, is a method with promising mathematical and computational properties, worthy of further investigation.

6 Conclusion and Outlook

In the language of chemistry, molecules are built from atoms and functional groups.[1, 3, 30, 191] Although the atoms and functional groups are deformed (or "promoted") when they are combined, they nonetheless maintain their quiddity. This is why, for example, the periodic table, along with tables of atomic properties (like the electronegativity, hardness, polarizability, etc.)[38, 190] are essential to practicing chemists. This motivated the strategy first proposed by Nalewajski, Parr, and Ayers: define the electron density of an atom in a molecule to maximize its resemblance to the electron densities of the isolated reference atoms and atomic ions enshrined in the periodic table.[24-27] The measure of "resemblance" between the atom-in-molecule's density, $\rho_A(\mathbf{r})$, and the reference isolated pro-atom's density, $\rho_A^0(\mathbf{r})$, was originally taken to be the Kullback-Leibler directed divergence.[26] The electron density was also chosen as the fundamental descriptor of atoms because of the Hohenberg-Kohn theorem,[37, 192] and inspired by the pioneering work of Richard Bader.[22, 30, 44, 191]

This sets the theoretical framework within which most Hirshfeld partitioning methods have been developed. In this framework, one exhaustively partitions the molecular density into its atomic contributions by minimizing the divergence of the atom-in-molecule densities from their corresponding reference pro-atomic densities, subject to the constraint that the sum of the atom-in-molecule densities is the total molecular

density. There is enormous freedom in this philosophy, notably in the way one measures the resemblance between the atoms and pro-atoms and in how one selects the pro-atoms. In this thesis we have explored both of these degrees of freedom and have not only established the boundaries of the theoretical framework of the Hirshfeld-partitioning, but also proposed a variational approach for selecting the reference pro-atoms, which has elegant mathematical and chemical properties.

In the first part of this thesis, we scrutinized various classes of divergence measures to characterize the pervasiveness of the Hirshfeld partitioning scheme. Specifically, we show that for any given set of pro-atomic densities, the only *local* density functionals that lead to the popular Hirshfeld partitioning are *f*-divergences. Other local divergence measures do not give the Hirshfeld partitioning, but every *f*-divergence suffices to obtain the Hirshfeld atoms-in-molecule. This was generalized even more by exploring divergences that are *nonlocal* density functionals but which also give the Hirshfeld partitioning. Therefore, the first part of this thesis establishes a solid mathematical framework for Hirshfeld partitioning approaches, and subsumes previous haphazard explorations and extensions of the Hirshfeld partitioning method.[28, 189, 193]

In the second part of this thesis, we explore novel representations of reference pro-atom densities. Hirshfeld schemes mainly differ in how they select the reference pro-atoms, which directly affects the quality of the atomic charges obtained. Specially, we proposed additive and multiplicative pro-piece density models to represent the reference pro-atom density. These provide very flexible pro-atom density models and allow inclusion of ground and excited states of neutral and charged atomic species as well as

any other type of basis functions. The parameters in these models were variationally optimized so that the pro-molecular density resembles the molecular density as accurately as possible when measured by an *f*-divergence subject to the constraint that the molecule and pro-molecule have the same number of electrons. The choice of *f*-divergence does, however, affect the optimal choice of pro-atoms to some extent. To circumvent the many possibilities provided by the family of *f*-divergences, we appeal to the extended Kullback-Liebler divergence measure, which has unique mathematical and chemical characteristics that differentiate it from other *f*-divergence measure.

Our endeavors culminate in presenting the unrivaled Additive Variational Hirshfeld (AVH) partitioning scheme. This novel scheme represents each pro-atom as a linear combination of non-negative basis functions, like spherically averaged densities of isolated (neutral and charged) atomic species corresponding to ground and/or excited states. The contribution of each basis function is determined through a variational principle, by minimizing the extended Kullback-Liebler divergence between the molecular and pro-molecular density. This specific divergence measure automatically fulfills the mathematical and chemical requirement that the molecule and pro-molecule to have the same number of electrons. In addition, it guarantees that each atom and pro-atom have the same number of electrons, and guarantees that the partitioning is size consistent. This results in unique and variationally determined pro-atom densities for the Hirshfeld partitioning. Any specific constraint on atoms can also be added to the variational formulation of AVH in a well-defined and straightforward manner.

We believe that the *f*-divergence family provides a unified theoretical framework for all Hirshfeld-like partitioning methods. In combination with the additive pro-piece model, this mathematical framework was leveraged to propose the unrivaled Additive Variational Hirshfeld (AVH) partitioning. Considering the unique mathematical features of the AVH scheme and the promising numerical results we have obtained, we believe that it has the potential to supplant other Hirshfeld partitioning schemes in near future.

7 Appendix

7.1 Detailed Derivation of the Hirshfeld Atom in a Molecule

For all of the Hirshfeld partitioning methods considered, the method of derivation of AIM density is basically similar. To demonstrate the procedure, we here provide a detailed derivation of the Hirshfeld AIM.

According to Nalewajski and Parr, the density that minimizes the Kullback-Leibler directed divergence between the density of the AIM and the density of the neutral atom, subject to the constraint that the sum of AIM densities is equal to the total molecular density, leads to the Hirshfeld AIM. This corresponds to optimizing the Lagrangian,

$$\Lambda\left[\{\rho_A\}\right] = \sum_{A=1}^{N_{\text{atoms}}} \int \rho_A(\mathbf{r}) \ln\left(\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})}\right) d\mathbf{r} + \int \lambda(\mathbf{r}) \left[\rho_{\text{mol}}(\mathbf{r}) - \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r})\right] d\mathbf{r} \tag{7.1}$$

where $\lambda(\mathbf{r})$ is the Lagrange multiplier function that forces the sum of the AIM densities to equal the molecular density at every point in space. The Lagrangian is stationary when the following equations are satisfied,

$$\begin{aligned} 0 &= \frac{\delta\Lambda\left[\{\rho_A\}\right]}{\delta\lambda(\mathbf{r})} = \rho_{\text{mol}}(\mathbf{r}) - \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r}) \\ \left\{0 = \frac{\delta\Lambda\left[\{\rho_A\}\right]}{\delta\rho_A(\mathbf{r})} = \ln\left(\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})}\right) + 1 - \lambda(\mathbf{r})\right\}_{A=1}^{N_{\text{atoms}}} & \end{aligned} \tag{7.2}$$

The second set of equations can be rearranged as

$$\ln\left(\frac{\rho_{\mathrm{A}}(\mathbf{r})}{\rho_{\mathrm{A}}^{0}(\mathbf{r})}\right)+1=\lambda(\mathbf{r})=\ln\left(\frac{\rho_{\mathrm{B}}(\mathbf{r})}{\rho_{\mathrm{B}}^{0}(\mathbf{r})}\right)+1 \tag{7.3}$$

implying that the ratio between the atomic density and the pro-atom density is the same for all atoms (because $\ln(x)+1$ is a monotonic function),

$$\frac{\rho_{\mathrm{A}}(\mathbf{r})}{\rho_{\mathrm{A}}^{0}(\mathbf{r})}=\frac{\rho_{\mathrm{B}}(\mathbf{r})}{\rho_{\mathrm{B}}^{0}(\mathbf{r})} \tag{7.4}$$

Rearranging this equation and summing over *B* gives

$$\frac{\rho_{\mathrm{A}}(\mathbf{r})}{\rho_{\mathrm{A}}^{0}(\mathbf{r})}\sum_{\mathrm{B}=1}^{N_{\mathrm{atoms}}}\rho_{\mathrm{B}}^{0}(\mathbf{r})=\sum_{\mathrm{B}=1}^{N_{\mathrm{atoms}}}\rho_{\mathrm{B}}(\mathbf{r}) \tag{7.5}$$

Then, using the constraint on the sum of AIM densities, the first equation in Eq. (7.2), and rearranging the resulting equation, one obtains the density of the Hirshfeld AIM,

$$\rho_{\mathrm{A}}(\mathbf{r})=\rho_{\mathrm{A}}^{0}(\mathbf{r})\frac{\rho_{\mathrm{mol}}(\mathbf{r})}{\sum_{\mathrm{B}=1}^{N_{\mathrm{atoms}}}\rho_{\mathrm{B}}^{0}(\mathbf{r})}. \tag{7.6}$$

To establish that the sum of the divergences of the AIM from the pro-atoms is equal to the divergence of the molecule from the promolecule, insert the definition of the Hirshfeld AIM, Eq. (7.6), back into the objective function for the variational principle. Rearranging then gives,

$$\sum_{A=1}^{N_{\text{atoms}}} \int \rho_A(\mathbf{r}) \ln\left(\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})}\right) d\mathbf{r} = \int \sum_{A=1}^{N_{\text{atoms}}} \left[\rho_{\text{A}}^0(\mathbf{r}) \frac{\rho_{\text{mol}}(\mathbf{r})}{\rho_{\text{mol}}^0(\mathbf{r})}\right] \ln\left(\frac{\rho_{\text{A}}^0(\mathbf{r})\rho_{\text{mol}}(\mathbf{r})}{\rho_A^0(\mathbf{r})\rho_{\text{mol}}^0(\mathbf{r})}\right) d\mathbf{r}$$
$$= \int \frac{\rho_{\text{mol}}(\mathbf{r})}{\rho_{\text{mol}}^0(\mathbf{r})} \ln\left(\frac{\rho_{\text{mol}}(\mathbf{r})}{\rho_{\text{mol}}^0(\mathbf{r})}\right) \sum_{A=1}^{N_{\text{atoms}}} \left[\rho_{\text{A}}^0(\mathbf{r})\right] d\mathbf{r} \tag{7.7}$$
$$= \int \rho_{\text{mol}}(\mathbf{r}) \ln\left(\frac{\rho_{\text{mol}}(\mathbf{r})}{\rho_{\text{mol}}^0(\mathbf{r})}\right) d\mathbf{r}$$

where in the last line we have used the definition of the promolecular density,

$$\rho_{\text{mol}}^0(\mathbf{r}) = \sum_{A=1}^{N_{\text{atoms}}} \rho_A^0(\mathbf{r}) \tag{7.8}$$

7.2 Perspective on Information-Theoretic Measures

A reader familiar with information theory will notice that the Nalewajski-Parr approach violates the spirit of information theory insofar as the atomic and pro-atom densities are not necessarily normalized to the same number of electrons. This leads to nonintuitive results, chief among them the fact that loss of information that occurs when the pro-atom distorts to the atom,

$$S_A\left[\rho_A, \rho_A^0\right] = \int \rho_A(\mathbf{r}) \ln\left(\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})}\right) d\mathbf{r} \tag{7.9}$$

is frequently negative when the atom has fewer electrons than the pro-atom. While this is mitigated in the Hirshfeld-I family of methods upon convergence, it is desirable to resolve this conundrum in the elementary Hirshfeld method.

To resolve the problem, we note that information theory is usually applied to probability distribution functions that are normalized to one, and speculate that we can

use the (pro)atom shape function, or density per particle, instead of the electron density,

$$\sigma_A(\mathbf{r}) = \frac{\rho_A(\mathbf{r})}{N_A} = \frac{\rho_A(\mathbf{r})}{\int \rho_A(\mathbf{r})\, d\mathbf{r}}. \tag{7.10}$$

Substituting the shape function into Eq. (1.8)

$$\underbrace{\min}_{\left\{N_A;\sigma_A(\mathbf{r}) \middle| \sum_{A=1}^{N_{\text{atoms}}} N_A\sigma_A(\mathbf{r})=\rho_{\text{mol}}(\mathbf{r})\right\}} \sum_{A=1}^{N_{\text{atoms}}} \int N_A\sigma_A(\mathbf{r}) \ln\left(\frac{N_A\sigma_A(\mathbf{r})}{N_A^0\sigma_A^0(\mathbf{r})}\right) d\mathbf{r} \tag{7.11}$$

and simplifying gives the expression,

$$\underbrace{\min}_{\left\{N_A;\sigma_A(\mathbf{r}) \middle| \sum_{A=1}^{N_{\text{atoms}}} N_A\sigma_A(\mathbf{r})=\rho_{\text{mol}}(\mathbf{r})\right\}} \left(\sum_{A=1}^{N_{\text{atoms}}} N_A \int \sigma_A(\mathbf{r}) \ln\left(\frac{\sigma_A(\mathbf{r})}{\sigma_A^0(\mathbf{r})}\right) d\mathbf{r} + \sum_{A=1}^{N_{\text{atoms}}} N_A \ln\left(\frac{N_A}{N_A^0}\right) \right) \tag{7.12}$$

Introducing the number of electrons in the molecule and the promolecule, N_{mol} and N_{mol}^0 respectively, and defining the fraction of electrons in the (pro)molecule as

$$x_A = \frac{N_A}{N_{mol}} \tag{7.13}$$

this can be rewritten as

$$\underbrace{\min}_{\left\{N_A;\sigma_A(\mathbf{r}) \middle| \sum_{A=1}^{N_{\text{atoms}}} N_A\sigma_A(\mathbf{r})=\rho_{\text{mol}}(\mathbf{r})\right\}} \left(\begin{array}{l} N_{mol}\left(\sum_{A=1}^{N_{\text{atoms}}} x_A \int \sigma_A(\mathbf{r}) \ln\left(\frac{\sigma_A(\mathbf{r})}{\sigma_A^0(\mathbf{r})}\right) d\mathbf{r} \right) \\ + N_{mol}\left(\sum_{A=1}^{N_{\text{atoms}}} x_A \ln\left(\frac{x_A}{x_A^0}\right) \right) + N_{mol} \ln\left(\frac{N_{mol}}{N_{mol}^0}\right) \end{array} \right) \tag{7.14}$$

The last term is a constant, and reflects the fact that the information loss can decrease to $-\infty$ if the pro-atom densities are chosen to have very many electrons. (This could already have been inferred directly from Eq. (1.8).) It is therefore only sensible to consider

information-theoretic partitionings where the promolecule and the molecule possess the same number of electrons.

The first two terms in Eq. (7.14) are nonnegative and have direct physical interpretations. The first sum, of which every term is nonnegative, can be viewed as the entropy-of-polarization, since it measures the way the shape of the pro-atoms' electron distributions deform upon molecule formation. The second sum strongly resembles the entropy of mixing in classical thermodynamics: it measures the effects of electron transfer between atoms. In Hirshfeld-I the second is zero at convergence, and in Hirshfeld-Iλ, and Hirshfeld-E this term is chosen to be zero by added constraints. In variational Hirshfeld-I, the entropy of mixing is not zero.

There has been significant interest in generalizations of the Hirshfeld partitioning to other measures of the distance between distributions.[28-29, 118, 189,194] For example, one can replace the Shannon entropy (and the Kullback-Leibler divergence) with analogous nonextensive entropies by Tsallis, Reyni, and others.[28, 118, 189] Remarkably, these generalized entropies generally lead back to the Hirshfeld definition of the AIM, as encapsulated by the Eq. (1.9). However, the simplification from Eq. (7.11) to Eq. (7.12) makes essential use of the properties of the logarithm, and so the rigor of these approaches may be questioned. Certainly their interpretative power is weakened by the absence of a decomposition into entropy-of-polarization and entropy-of-mixing contributions.

All is not lost, however, as long as the distributive rule, Eq. (1.3), holds. If the distributive rule holds, and the promolecule and molecule are constrained to have the

same number of electrons, then nonextensive entropies and other reasonable measures of the "distance" between the molecule and the promolecule will be nonnegative $D\left[\rho_{\text{mol}};\rho^0_{\text{mol}}\right]\geq 0$. It is still true that individual atomic contributions, $D\left[\rho_A;\rho^0_A\right]$, can be negative, but as we have indicated, this was true even for the venerable Hirshfeld method. This does not mean that methods based on more general measures of the deviation between electron densities should not be used, but merely that for information measures other than Kullback-Leibler divergence, only $D\left[\rho_{\text{mol}};\rho^0_{\text{mol}}\right]$, and not the individual atomic contributions to it, should be used in interpretation.

There has been interest in generalizing the approach of Nalewajski and Parr in Eq. (1.8) to other measures for the deviation between two electron densities. Generally we can write,

$$\underbrace{\min}_{\left\{\rho_A(\mathbf{r})\middle|\sum_{A=1}^{N_{\text{atoms}}}\rho_A(\mathbf{r})=\rho_{\text{mol}}(\mathbf{r})\right\}}\ \sum_{A=1}^{N_{\text{atoms}}} D\left[\rho_A(\mathbf{r});\rho^0_A(\mathbf{r})\right] \qquad (7.15)$$

$$\text{constraints}$$

As mentioned, this definition should be viewed skeptically unless (1) the distributive rule in Eq. (1.3) holds and (2) the molecule and promolecule have the same number of electrons. Restricting ourselves to only variational methods, the only degrees of freedom are:

- The functional used to measure the deviation between electron densities.
- The definition of the pro-atom.
- Whether any constraints are imposed on the minimization. For example, we may wish to constrain the molecule and promolecule to have the same charge or, more

stringently, to constrain the AIM and its pro-atom to have the same population. Here, we will overview some of the choices for the deviation-functional and the pro-atom-definition that have been considered in the literature, or in unpublished research by us. We will pay particular attention to whether these functionals satisfy the distributive rule and whether they lead to localized AIM.

I. Measures of the Deviation between Electron Density

a. Distance Metrics

A measure of the dissimilarity, $D\left[\rho;\rho^0\right]$, between any two nonnegative integrable functions is said to be a distance if it satisfies the following requirements,

(i) $D\left[\rho;\rho^0\right]\geq 0$ (nonnegativity; separation of points)

(ii) $D\left[\rho;\rho^0\right]=0\leftrightarrow\rho(\mathbf{r})=\rho^0(\mathbf{r})$. (distance between equivalent functions is zero)

(iii) $D\left[\rho;\rho^0\right]=D\left[\rho^0;\rho\right]$. (symmetry)

(iv) $D\left[\rho_1;\rho^0\right]+D\left[\rho_2;\rho^0\right]\geq D\left[\rho_1;\rho_2\right]$. (triangle inequality)

In the context of divergence measures, often the first requirement is relaxed because it is assumed that all probability distributions are normalized to one, giving,

(i') $D\left[\rho;\rho^0\right]\geq 0$ if $\int\rho(\mathbf{r})d\mathbf{r}=\int\rho^0(\mathbf{r})d\mathbf{r}$.

Any functional that satisfies **(i')** and **(ii)** is said to be a divergence measure. Functionals that satisfy **(i')**, **(ii)**, and **(iii)** are symmetric divergence measures. Divergence measures

that satisfy requirements **(i)** and **(ii)** are called extended because they can be applied to non-normalized probability distribution functions.

b. Kullback-Leibler Directed Divergence and its Generalizations

Most Hirshfeld-related techniques use the Kullback-Leibler directed divergence in Eq. (1.8). The resulting partitioning methods satisfy the separated-atom limit and the distributive property in Eq. (1.3). However, the Kullback-Leibler directed divergence does not measure the *distance* between two electron densities because it does not satisfy the triangle inequality and it is not symmetric. That is,

$$D_{\mathrm{KL}}\left[\rho;\rho^0\right]=\int\rho(\mathbf{r})\ln\left(\frac{\rho(\mathbf{r})}{\rho^0(\mathbf{r})}\right)d\mathbf{r} \tag{7.16}$$

only satisfies the **(i')** and **(ii)** properties of a distance metric. The symmetric (undirected) Kullback-Leibler divergence treats the density and prodensity equivalently,

$$\begin{aligned} D_{\mathrm{symKL}}\left[\rho;\rho^0\right] &= \int\left(\rho(\mathbf{r})-\rho^0(\mathbf{r})\right)\ln\left(\frac{\rho(\mathbf{r})}{\rho^0(\mathbf{r})}\right)d\mathbf{r} \\ &= D_{\mathrm{KL}}\left[\rho;\rho^0\right]+D_{\mathrm{KL}}\left[\rho^0;\rho\right] \end{aligned} \tag{7.17}$$

$D_{\mathrm{symKL}}\left[\rho;\rho^0\right]$ satisfies properties **(i)**, **(ii)** and **(iii)** of a distance metric. Using the symmetric Kullback-Leibler divergence or any of the forms that lie between Eqs. (7.16) and (7.17),

$$D_{\mathrm{KL}}^{\alpha,\beta}\left[\rho;\rho^0\right]=\alpha D_{\mathrm{KL}}\left[\rho;\rho^0\right]+\beta D_{\mathrm{KL}}\left[\rho^0;\rho\right] \qquad \alpha,\beta\geq 0 \tag{7.18}$$

gives back the Hirshfeld partitioning.[27] One can also symmetrize the "reference" densities, obtaining the Jensen-Shannon divergence,

$$D_{\mathrm{JS}}\left[\rho;\rho^{0}\right]=\tfrac{1}{2}D_{\mathrm{KL}}\left[\rho;\tfrac{1}{2}\left(\rho^{0}+\rho\right)\right]+\tfrac{1}{2}D_{\mathrm{KL}}\left[\rho^{0};\tfrac{1}{2}\left(\rho^{0}+\rho\right)\right] \qquad (7.19)$$

$D_{\mathrm{JS}}\left[\rho;\rho^{0}\right]$ satisfies properties **(i)**, **(ii)** and **(iii)**. Moreover, it is "almost" a distance since $\left(D_{\mathrm{JS}}\left(\rho;\rho^{0}\right)\right)^{1/2}$ satisfies the triangle inequality.[195] Finally, we mention the extended Kullback-Leibler divergence, which satisfies properties **(i)** and **(ii)**,[160-162]

$$D_{\mathrm{KL}}\left[\rho;\rho^{0}\right]=\int\left(\rho(\mathbf{r})\ln\left(\frac{\rho(\mathbf{r})}{\rho^{0}(\mathbf{r})}\right)+\rho^{0}(\mathbf{r})-\rho(\mathbf{r})\right)d\mathbf{r}. \qquad (7.20)$$

All of these divergence measures recover the Hirshfeld partitioning, satisfy the separated-atom limit, and fulfill the distributive property in Eq. (1.3).

c. (Generalized) Hellinger-Bhattacharya Distance

If one wishes to satisfy all four properties **(i)** to **(iv)** of a metric one can use the (generalized) Hellinger-Bhattacharya distance,

$$D_{\mathrm{BH}}^{\nu}\left[\rho;\rho^{0}\right]=\int\left(\left(\rho(\mathbf{r})\right)^{1/\nu}-\left(\rho^{0}(\mathbf{r})\right)^{1/\nu}\right)^{\nu}d\mathbf{r} \qquad \nu\neq\{0,1\} \qquad (7.21)$$

where ν is any real number except zero or one. This is the popular Hellinger-Bhattacharya distance when ν = 2. This family of divergences in Eq. (7.21) gives back the Hirshfeld definition, and it satisfies the separated atom limit and the distributive rule.[29] In addition, because this measure satisfies property **(i)**, it does not require any constraint on the promolecule to have the same number of electrons as the molecule to be mathematically valid.

d. α-divergence; Tsallis and Reyni forms of the entropy

The Kullback-Leibler approach is based on the Shannon entropy. One can generalize to nonextensive entropies like the Tsallis entropy

$$D^{\alpha}_{\text{Tsallis}}\left[\rho;\rho^{0}\right]=\frac{1}{\alpha-1}\int\left[\left(\rho(\mathbf{r})\right)^{\alpha}\left(\rho^{0}(\mathbf{r})\right)^{1-\alpha}-\rho(\mathbf{r})\right]d\mathbf{r} \qquad \alpha\in \mathrm{R} \qquad (7.22)$$

or the rescaling of the Tsallis entropy into the α-divergence,

$$D^{\alpha}\left[\rho;\rho^{0}\right]=\frac{1}{\alpha(\alpha-1)}\int\left[\left(\rho(\mathbf{r})\right)^{1-\alpha}\left(\rho^{0}(\mathbf{r})\right)^{\alpha}-\rho(\mathbf{r})\right]d\mathbf{r} \qquad \alpha\in \mathrm{R} \qquad (7.23)$$

Replacing the Kullback-Leibler divergence in Eq. (1.8) with these alternatives always recovers the Hirshfeld partitioning; the resulting methods satisfy the separated-atom limit and the distributive rule.

e. L1-norm

The L1 distance between electron densities,

$$D^{1}\left[\rho;\rho^{0}\right]=\int\left|\rho(\mathbf{r})-\rho^{0}(\mathbf{r})\right|d\mathbf{r} \qquad (7.24)$$

is a distance metric. It it *consistent* with the Hirshfeld partitioning, but it does not determine the AIM density uniquely.[196]

f. *f*-divergence.

All of the divergence functionals in **(b)-(e)** are special cases of the *f*-divergence,

$$D_{f}\left[\rho;\rho^{0}\right]=\int\rho(\mathbf{r})f\left(\frac{\rho^{0}(\mathbf{r})}{\rho(\mathbf{r})}\right)d\mathbf{r} \qquad \begin{matrix} f(1)=0 \\ f(x)\ \text{is convex} \end{matrix} \qquad (7.25)$$

where $f(x)$ is *any* convex function with $f(1) = 0$. Indeed, one can show that for any divergence measure that is a local functional, the Hirshfeld partitioning is obtained if and only if the deviation between the atomic and pro-atomic densities is measured with an *f*-divergence.[28] One can also show that every *f*-divergence satisfies the separated atom limit and the distributive rule. There are many forms of divergence where it is difficult to determine what the correct function $f(x)$ is, but because we know these forms give back the Hirshfeld AIM, we know such an $f(x)$ must exist and that the separated atom limit and distributive rule will be satisfied. For example, the average-density-weighted-density-distance,

$$D_{\text{avg}}^{n}\left[\rho;\rho^{0}\right]=\int\frac{\left(\rho(\mathbf{r})-\rho^{0}(\mathbf{r})\right)^{2n}}{\left[\frac{1}{2}\left(\rho(\mathbf{r})+\rho^{0}(\mathbf{r})\right)\right]^{2n-1}}d\mathbf{r} \qquad n>0 \tag{7.26}$$

recovers the Hirshfeld AIM, and must therefore be an *f*-divergence.[28]

All *f*-divergences satisfy properties **(i')** and **(ii)**. Additional properties can be satisfied by imposing additional restrictions on the form of the function $f(x)$. For example, extended *f*-divergences satisfying properties **(i)** and **(ii)** are obtained by requiring that $f'(1)=0$ (where we are using $f'(1)$ to denote the derivative of $f(x)$ evaluated at 1). Any *f*-divergence can be converted into an extended *f*-divergence by defining $f_{\text{ext}}(x)=f(x)-(x-1)f'(1)$. Symmetrized *f*-divergences with the form

$$D_{f_{\text{sym}}}\left[\rho;\rho^{0}\right]=\frac{1}{2}\int\rho(\mathbf{r})f\left(\frac{\rho^{0}(\mathbf{r})}{\rho(\mathbf{r})}\right)+\rho^{0}(\mathbf{r})f\left(\frac{\rho(\mathbf{r})}{\rho^{0}(\mathbf{r})}\right)d\mathbf{r} \tag{7.27}$$

are also *f*-divergences. These divergences satisfy requirements **(i)**, **(ii)**, and **(iii)**. Any *f*-divergence can be converted into a symmetrized *f*-divergence by defining

$f_{\text{sym}}(x) = \frac{1}{2}\left(f(x) + xf\left(x^{-1}\right)\right)$. It is important to note that every symmetric *f*-divergence is also an extended *f*-divergence.

It is sometimes useful to consider the average of the density and the prodensity to be the reference, leading to divergences with the form,

$$D_{f_{\text{symref}}}\left[\rho;\rho^0\right] = \int \rho(\mathbf{r}) f\left(\frac{\frac{1}{2}\rho^0(\mathbf{r}) + \frac{1}{2}\rho(\mathbf{r})}{\rho(\mathbf{r})}\right) d\mathbf{r} \tag{7.28}$$

This is also an *f*-divergence, and in this case the function *f*(x) satisfies the equation $f(x) = f\left(\frac{1}{2} + \frac{1}{2}x\right)$ and the resulting divergence satisfies properties **(i')** and **(ii)**. Any *f*-divergence can be converted to an *f*-divergence with a symmetric reference density by defining $f_{\text{symref}}(x) = f\left(\frac{1}{2} + \frac{1}{2}x\right)$. One can satisfy properties **(i)**, **(ii)** and **(iii)** by going to the generalized Jensen-Shannon form,

$$D_{f_{\text{genJS}}}\left[\rho;\rho^0\right] = \int \rho(\mathbf{r}) f\left(\frac{\frac{1}{2}\rho^0(\mathbf{r}) + \frac{1}{2}\rho(\mathbf{r})}{\rho(\mathbf{r})}\right) d\mathbf{r} + \int \rho^0(\mathbf{r}) f\left(\frac{\frac{1}{2}\rho^0(\mathbf{r}) + \frac{1}{2}\rho(\mathbf{r})}{\rho^0(\mathbf{r})}\right) d\mathbf{r} \tag{7.29}$$

which corresponds to the choice $f_{\text{genJS}}(x) = \frac{1}{2}\left(f\left(\frac{1}{2} + \frac{1}{2}x\right) + xf\left(\frac{1}{2} + \frac{1}{2}x^{-1}\right)\right)$. For the class of α-divergences, Eq. (7.23), the square root of the *f*-divergence based on $f_{\text{genJS}}(x)$ satisfies the triangle inequality, and is therefore a distance metric.[141-142]

The L1 norm corresponds to the limiting case where *f*(*x*) is no longer convex, but merely nonconcave, specifically $f(x) = |x-1|$. This is consistent with the observation that the L1 norm **(e)** is consistent with the Hirshfeld partitioning but is not uniquely associated to the Hirshfeld partitioning.

g. Nonextensive Entropy Measures

While the family of *f*-divergences are the only *local* functionals that give the Hirshfeld partitioning, there are also nonlocal functionals that give the Hirshfeld partitioning.[189] Many of these functionals, like the Réyni divergence,

$$D_R^\alpha\left[\{\rho_A\};\{\rho_A^0\}\right]=\frac{1}{\alpha-1}\ln\left(\frac{\int\sum_A^{N_{\text{atoms}}}\left(\rho_A(\mathbf{r})\left(\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})}\right)^{\alpha-1}\right)d\mathbf{r}}{\int\sum_A^{N_{\text{atoms}}}\rho_A(\mathbf{r})\,d\mathbf{r}}\right) \tag{7.30}$$

are closely related to the α-divergences **(d)**. Other choices can be considered generalizations of the *f*-divergence. For example, the *H*-divergences defined by

$$H_{h,\varphi_1,\varphi_2}\left[\{\rho_A\};\{\rho_A^0\}\right]=h\left(\frac{\int\sum_{A=1}^{N_{\text{atoms}}}\rho_A(\mathbf{r})\varphi_1\left(\frac{\rho_A^0(\mathbf{r})}{\rho_A(\mathbf{r})}\right)d\mathbf{r}}{\int\sum_{A=1}^{N_{\text{atoms}}}\rho_A(\mathbf{r})\varphi_2\left(\frac{\rho_A^0(\mathbf{r})}{\rho_A(\mathbf{r})}\right)d\mathbf{r}}\right) \tag{7.31}$$

give the Hirshfeld partitioning if (1) $h(x)$ is monotonically increasing; (2) $h(0)=0$; (3) $\varphi_1(x)$ is convex; (4) $\varphi_1(1)=0$; (5) $\varphi_2(x)$ is nonconvex, and (6) $\varphi_2(x)>0$.

h. Kernel norm/Mahalonobis distance

Suppose that $K(\mathbf{r},\mathbf{r}')$ is a positive definite integral kernel, meaning that

$$\iint g(\mathbf{r})K(\mathbf{r},\mathbf{r}')g(\mathbf{r}')\,d\mathbf{r}\,d\mathbf{r}'>0 \tag{7.32}$$

for all $g(\mathbf{r})\neq 0$. The associated kernel divergence,

$$D_K\left[\rho;\rho^0\right]=\iint\left(\rho(\mathbf{r})-\rho^0(\mathbf{r})\right)K(\mathbf{r},\mathbf{r}')\left(\rho(\mathbf{r}')-\rho^0(\mathbf{r}')\right)d\mathbf{r}d\mathbf{r}' \tag{7.33}$$

satisfies properties **(i)**, **(ii)**, and **(iii).** Similar to the Jensen-Shannon divergence, the square root of the kernel divergence satisfies the triangle inequality, **(iv)**. The square root of Eq. (7.33) is usually called the kernel norm or the Mahalanobis distance. Popular choices for the integral kernel are the Dirac delta function kernel, $K(\mathbf{r},\mathbf{r}')=\delta(\mathbf{r}-\mathbf{r}')$ (giving the L2-norm)[28]

$$\left\|\rho-\rho^0\right\|_2=\int\left(\rho(\mathbf{r})-\rho^0(\mathbf{r})\right)^2 d\mathbf{r} \tag{7.34}$$

and the Coulomb kernel, $K(\mathbf{r},\mathbf{r}')=\left|\mathbf{r}-\mathbf{r}'\right|^{-1}$,[123, 197-199]

$$D_K\left[\rho;\rho^0\right]=\iint\frac{\left(\rho(\mathbf{r})-\rho^0(\mathbf{r})\right)\left(\rho(\mathbf{r}')-\rho^0(\mathbf{r}')\right)}{\left|\mathbf{r}-\mathbf{r}'\right|}d\mathbf{r}d\mathbf{r}'. \tag{7.35}$$

Unfortunately, when this divergence measure is used in Eq. (1.8) one does not recover the Hirshfeld partitioning, but instead a partitioning with extremely nonlocal AIM densities, because the deformation density—the difference between the molecular density and the promolecular density—is divided equally between all the AIM,

$$\rho_A(\mathbf{r})=\rho_A^0(\mathbf{r})+\frac{\rho_{\text{mol}}(\mathbf{r})-\rho_{\text{mol}}^0(\mathbf{r})}{N_{\text{atoms}}}. \tag{7.36}$$

The distributive property, Eq. (1.3), is satisfied for the kernel divergence.

i. Bregman divergence

The Bregman divergence of the electron density, $\rho(\mathbf{r})$, from a reference pro-density, $\rho^0(\mathbf{r})$, is defined as

$$D_{\text{Bregman},C}\left[\rho;\rho^0\right]=C[\rho]-C\left[\rho^0\right]-\int\frac{\delta C\left[\rho^0\right]}{\delta\rho(\mathbf{r})}\left(\rho(\mathbf{r})-\rho^0(\mathbf{r})\right)d\mathbf{r} \quad (7.37)$$

where $C[\rho]$ is a differentiable and convex density functional. Unfortunately, when the Bregman divergence measure is used in Eq. (1.8), in general one does not recover the Hirshfeld partitioning. (An exception is the Kullback-Leibler divergence, which can be expressed as a Bregman divergence.) In general, it is difficult to express the AIM obtained from the Bregman divergence, but if one assumes that $C[\rho]$ is twice differentiable, then the following density-dependent integral kernel exists and is invertible,

$$K_{\rho_A,\rho_A^0}(\mathbf{r},\mathbf{r}')=\overline{\frac{\delta^2 C\left[t\rho_A+(1-t)\rho_A^0\right]}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')}}=\int_0^1\frac{\delta^2 C\left[t\rho_A+(1-t)\rho_A^0\right]}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')}dt \quad (7.38)$$

This allows us to generalizes the Hirshfeld expression for an atom in a molecule to the Bregman divergence:

$$\rho_A(\mathbf{r})=\rho_A^0(\mathbf{r})+\int K^{-1}_{\rho_A,\rho_A^0}(\mathbf{r},\mathbf{r}')\left(\sum_{B=1}^{N_{\text{atoms}}}K^{-1}_{\rho_B,\rho_B^0}(\mathbf{r}',\mathbf{r}'')\right)^{-1}\left(\rho_{\text{mol}}(\mathbf{r}'')-\rho^0_{\text{mol}}(\mathbf{r}'')\right)d\mathbf{r}'d\mathbf{r}''. \quad (7.39)$$

In general, the AIM obtained from the Bregman divergence are unreasonably delocalized. This is clear when one considers that the kernel norm **(g)** corresponds to the special case of the Bregman divergence where

$$C[\rho]=\iint\rho(\mathbf{r})K(\mathbf{r},\mathbf{r}')\rho(\mathbf{r}')d\mathbf{r}d\mathbf{r}' \quad (7.40)$$

and the kernel in Eq. (7.38), $K_{\rho_A,\rho_A^0}(\mathbf{r},\mathbf{r}')=K(\mathbf{r},\mathbf{r}')$, is therefore density independent.

II. Possible pro-atom density definitions

a. Spherical averaged neutral atoms

In the original Hirshfeld partitioning, the pro-atom densities are chosen as the spherically averaged neutral atom densities. Traditionally the pro-atom densities are evaluated using the same quantum chemistry method and the same basis set that was used to evaluate the molecular density. Constructing a reference database of spherically-averaged atomic densities for neutral atoms (and, depending on the partitioning approach taken, also atomic ions and possibly even atomic excited states) is conceivable. There are potential pitfalls here, however, because describing the AIM and pro-atom densities at different levels of theory will generally cause the divergence between the atomic densities, $D\left[\rho_A;\rho_A^0\right]$, to be bigger.

It is also possible to forgo spherical averaging of the atomic densities, so that the atomic densities retain directionality. For open-shell atoms one then must determine the appropriate atomic density amongst the infinite number of possible degenerate densities. (This includes not only the need to select the correct orientation of the atomic density, but also the need to select the appropriate representation for the atomic density. For example, the atomic density for a *p*-block atom (groups 13-17 in the periodic table) can be built from Cartesian *p*-orbitals, spherical harmonic *p*-orbitals, *etc.*).) Minimization of $D\left[\rho_{\text{mol}},\sum_{A=1}^{N_{\text{atoms}}}\rho_A^0\right]$ with respect to the non-spherically-averaged pro-atom densities is a nonconvex optimization problem with many local minima. The choice of spherically averaged pro-atom densities, therefore, is numerically motivated, and potentially

compromises chemical properties. Note, however, that when the pro-atoms are spherically symmetric, minimization of $\sum_{A=1}^{N_{\text{atoms}}} D\left[\rho_A;\rho_A^0\right]$ favors AIM that are also nearly spherically symmetric. This leads to AIM with smaller higher-order multipoles, leading to more chemically intuitive atomic partial charges. Neutral pro-atoms seem to be poor choices for molecules containing atoms with sizable partial charges. Choosing neutral pro-atoms also implies that the promolecule will be neutral, which is inappropriate for molecular ions.

b. Atomic densities with fractional charge

In the Hirshfeld-I, Hirshfeld-Iλ, and variational Hirshfeld-I methods, atomic densities with fractional charge are used. The mathematically correct way to do this is given by the zero temperature limit of the grand canonical ensemble, cf. Eq. (1.10).[61-64] One could also determine the electron density for the fractionally-populated isolated atoms directly, but this is more expensive (because it requires that one calculate the isolated atomic densities with the appropriate charge at each iteration). This also compromises the accuracy of the method insofar as Hartree-Fock, Kohn-Sham density-functional theory, and some *ab initio* wavefunction methods are much less accurate for systems with fractional electron number.[200-204]

Because of the derivative discontinuity of the density at integer population, using fractionally charged isolated atoms as pro-atoms leads to inherently discontinuous optimization methods for the AIM. This can make it difficult to determine the AIM, and also makes it difficult to exclude the possibility of finding a suboptimal solution.

c. Basis sets for expanding the pro-atomic density

In the Hirshfeld-E method, one decides instead to approximate the electron density as a linear combination of the atomic Fukui functions (Eq. (1.17)). Similarly, in the Gaussian iterative stockhold analysis (GISA), one approximates the electron density as a linear combination of *s*-type Gaussians. In the minimal basis iterative stockholder method (MBIS), the electron density is approximated as a sum of *s*-type Slater functions (Eq. (1.25)). In general, one forces the coefficients in the expansion to be nonnegative, as this guarantees the pro-atom densities to be nonnegative. (For Hirshfeld-E, ensuring nonnegative pro-atoms is more subtle, as the Fukui function can be negative.[3, 168, 205-209])

The Hirshfeld-E method is inspired by the representation of the *N*-electron density as a sum of successive Fukui functions. One could instead be inspired by the representation of the *N*-electron density as the sum of the squares of the atomic natural orbitals, multiplied by the appropriate occupation numbers,[210]

$$\rho^0_{A;N}(\mathbf{r}) = \sum_{k=1}^{N_{\text{orbitals}}} n_{A;k} \left|\phi_{A;k}(\mathbf{r})\right|^2 . \tag{7.41}$$

The analogue of Eq. (1.17) would be to use the spherically-averaged atomic natural orbitals as a basis set for expanding the pro-atom densities,

$$\rho^0_{A}(\mathbf{r}) = \sum_{k=1}^{N_{\text{orbitals}}} c_k \left|\phi_{A;k}(r)\right|^2 . \tag{7.42}$$

The advantage of the orbital-driven approach is that it precludes the need to perform separate calculations of the electron density of all possible atomic ions. The disadvantage of the orbital-driven approach is that in the basis-set limit, the expansion basis allows one

to mimic the (undesirable) features of ISA, e.g., slowly-decaying and nonmonotonic pro-atom densities.

In general, basis-set expansions that are too short will not give accurate pro-atoms, leading to problems like one observes for MBIS for molecular anions. Basis-set expansions that are too long will lead to pro-atoms that are too delocalized, leading to excessive conformation dependence and unphysical charges for atoms surrounded by a (nearly) spherical shell of other atoms like one observes for ISA and GISA. One therefore needs a natural way to truncate the expansion. One way to do this is to consider the spherically-averaged densities of the physically bound atomic ions as a basis,

$$\rho_A^0(\mathbf{r}) = \sum_{N=1}^{N_{\max}} c_N \rho_{A;N}(r) \qquad c_N \geq 0 \tag{7.43}$$

This naturally prevents some of the problems that afflict Hirshfeld-I, as the densities of unbound atomic (di)anions are no longer needed.

Because the basis-set expansions we consider are restricted to nonnegative expansion coefficients, $c_k \geq 0$, they can be interpreted as a weighted average of the basis functions. Clearly one could consider other ways of averaging the basis functions using, e.g., the power mean. Of these choices, the geometric mean is particularly appealing, as it allows one to control the asymptotic decay of the pro-atom density, so that all the pro-atom densities might decay at the same rate. (This would ensure that the AIM densities decay at the same asymptotic rate, which is one of the desiderata listed in section II.B.) For example, instead of an additive combination of atomic density basis functions like Eq. (7.43), one could consider the multiplicative form:

$$\rho_A^0(\mathbf{r}) = c_0 \prod_{N=1}^{N_{\max}} \left(\rho_{A;N}(r)\right)^{c_N} \qquad c_N \geq 0. \tag{7.44}$$

Our preliminary calculations show that Eqs. (7.43) and (7.44) give promising results, and are therefore a favorable tradeoff between too-restrictive and too-general basis-set expansions.

d. Spherically averaged atomic density

Given the inherent freedom associated with specifying an appropriate atomic basis set for the pro-atom density, it is appealing to allow the AIM to specify its own spherical reference pro-atom, without any restriction in form. That is the idea behind iterative stockholder analysis (ISA), where the spherical average of the AIM density is used as the pro-atom, Eq. (1.19). Alternatively, this corresponds to finding the nonnegative spherical functions, $\{b_A(r)\}$, such that the divergence between the molecular and promolecular densities is minimized,

$$\underbrace{\min}_{\{b_A(r)\geq 0\}_{A=1}^{N_{\text{atoms}}}} D\left[\rho_{\text{mol}}; \sum_{A=1}^{N_{\text{atoms}}} b_A\left(\left|\mathbf{r}-\mathbf{R}_A\right|\right)\right] \tag{7.45}$$

ISA is therefore equivalent to basis-set expansion in an infinite basis set, which is why the pro-atom density from ISA is often chemically nonsensical.

7.3 Plots of Atomic Basis Functions

Neutral and charged atomic species are frequently used in most Hirshfeld partitioning schemes. Here we provide (spherically averaged) density plots of ground-state species of neutral/charged hydrogen, carbon, nitrogen, oxygen and fluorine atoms to show their dependence on the level of theory and basis set used. Also, we tabulate the ionization potential of various atomic anions to highlight the unbound and/or basis-set bound negatively charges species at various levels of theory and basis set.

Table 7.1 Ionization potential (IP) of negatively charged atomic species, $IP\left(X^{-}\right)=E_{X}\left(N\right)-E_{X^{-}}\left(N+1\right)$, at various levels of theory and basis set. Unbound species (bold in red) have a negative IP.

Level of Theory	IP(H^-)	IP(C^-)	IP(N^-)	IP(O^-)	IP(F^-)
UHF/cc-pVDZ	**-0.0505**	**-0.0070**	**-0.1218**	**-0.0710**	**-0.0093**
UHF/cc-pVTZ	**-0.0331**	0.0046	**-0.0969**	**-0.0456**	0.0188
UHF/cc-pVQZ	**-0.0265**	0.0104	**-0.0860**	**-0.0344**	0.0309
UHF/cc-pV5Z	**-0.0194**	0.0142	**-0.0770**	**-0.0260**	0.0394
UHF/aug-cc-pVDZ	**-0.0122**	0.0172	**-0.0683**	**-0.0192**	0.0472
UHF/aug-cc-pVTZ	**-0.0112**	0.0166	**-0.0666**	**-0.0208**	0.0439
UHF/aug-cc-pVQZ	**-0.0107**	0.0166	**-0.0648**	**-0.0210**	0.0434
UHF/aug-cc-pV5Z	**-0.0099**	0.0166	**-0.0603**	**-0.0211**	0.0432
UHF/d-aug-cc-pVDZ	**-0.0051**	0.0172	**-0.0310**	**-0.0191**	0.0472
UHF/d-aug-cc-pVTZ	**-0.0045**	0.0166	**-0.0305**	**-0.0207**	0.0439
UHF/d-aug-cc-pVQZ	**-0.0044**	0.0166	**-0.0293**	**-0.0210**	0.0434
UHF/d-aug-cc-pV5Z	**-0.0042**	0.0166	**-0.0271**	**-0.0211**	0.0432
UB3LYP/cc-pVDZ	**-0.0126**	0.0121	**-0.0635**	**-0.0152**	0.0432
UB3LYP/cc-pVTZ	0.0077	0.0301	**-0.0295**	0.0227	0.0873
UB3LYP/cc-pVQZ	0.0152	0.0379	**-0.0157**	0.0383	0.1053
UB3LYP/cc-pV5Z	0.0233	0.0447	**-0.0031**	0.0519	0.1203
UB3LYP/aug-cc-pVDZ	0.0326	0.0503	0.0063	0.0616	0.1309
UB3LYP/aug-cc-pVTZ	0.0334	0.0504	0.0079	0.0616	0.1297
UB3LYP/aug-cc-pVQZ	0.0336	0.0505	0.0085	0.0618	0.1296
UB3LYP/aug-cc-pV5Z	0.0337	0.0507	0.0093	0.0620	0.1296
UB3LYP/d-aug-cc-pVDZ	0.0336	0.0512	0.0124	0.0628	0.1314
UB3LYP/d-aug-cc-pVTZ	0.0342	0.0512	0.0135	0.0627	0.1301
UB3LYP/d-aug-cc-pVQZ	0.0342	0.0511	0.0137	0.0626	0.1299
UB3LYP/d-aug-cc-pV5Z	0.0342	0.0512	0.0143	0.0625	0.1298
UωB97XD/cc-pVDZ	**-0.0153**	0.0154	**-0.0674**	**-0.0146**	0.0470
UωB97XD/cc-pVTZ	0.0051	0.0301	**-0.0374**	0.0187	0.0856
UωB97XD/cc-pVQZ	0.0128	0.0376	**-0.0244**	0.0332	0.1021
UωB97XD/cc-pV5Z	0.0209	0.0437	**-0.0126**	0.0455	0.1151
UωB97XD/aug-cc-pVDZ	0.0291	0.0485	**-0.0035**	0.0553	0.1263
UωB97XD/aug-cc-pVTZ	0.0298	0.0481	**-0.0030**	0.0540	0.1236
UωB97XD/aug-cc-pVQZ	0.0299	0.0482	**-0.0028**	0.0537	0.1227
UωB97XD/aug-cc-pV5Z	0.0300	0.0484	**-0.0021**	0.0539	0.1228
UωB97XD/d-aug-cc-pVDZ	0.0294	0.0490	0.0004	0.0561	0.1267
UωB97XD/d-aug-cc-pVTZ	0.0299	0.0484	0.0003	0.0546	0.1239
UωB97XD/d-aug-cc-pVQZ	0.0300	0.0484	0.0002	0.0541	0.1229
UωB97XD/d-aug-cc-pV5Z	0.0300	0.0485	0.0005	0.0542	0.1229

Figure 7.1 Log plot of spherically averaged density of hydrogen atomic species at various levels of theory and basis set. Unbound species are denoted with dashed lines. The energy values in plot labels are in atomic units.

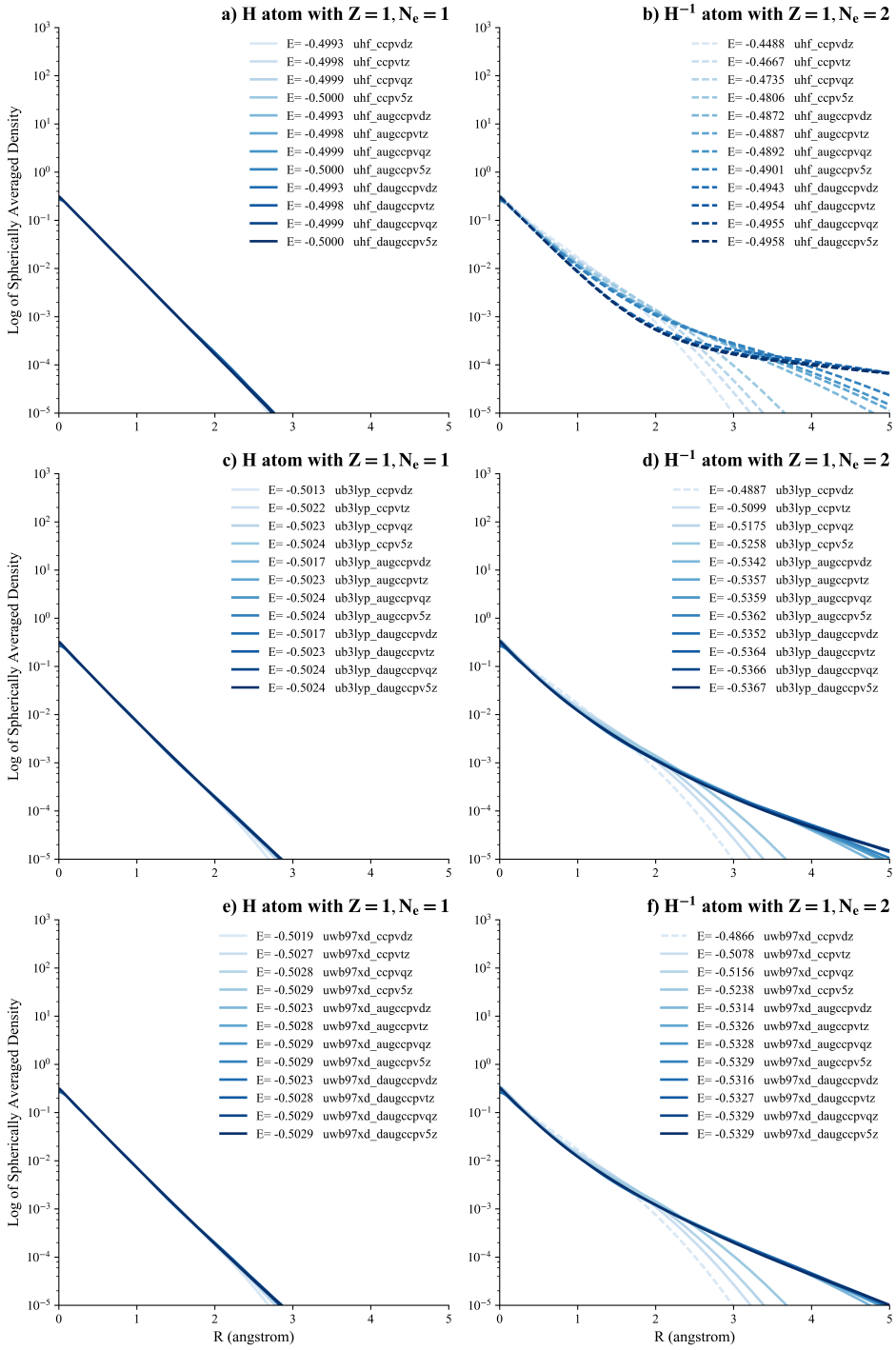


Figure 7.2 Log plot of spherically averaged density of carbon atomic species at various levels of theory and basis set. Unbound species are denoted with dashed lines. The energy values in plot labels are in atomic units.

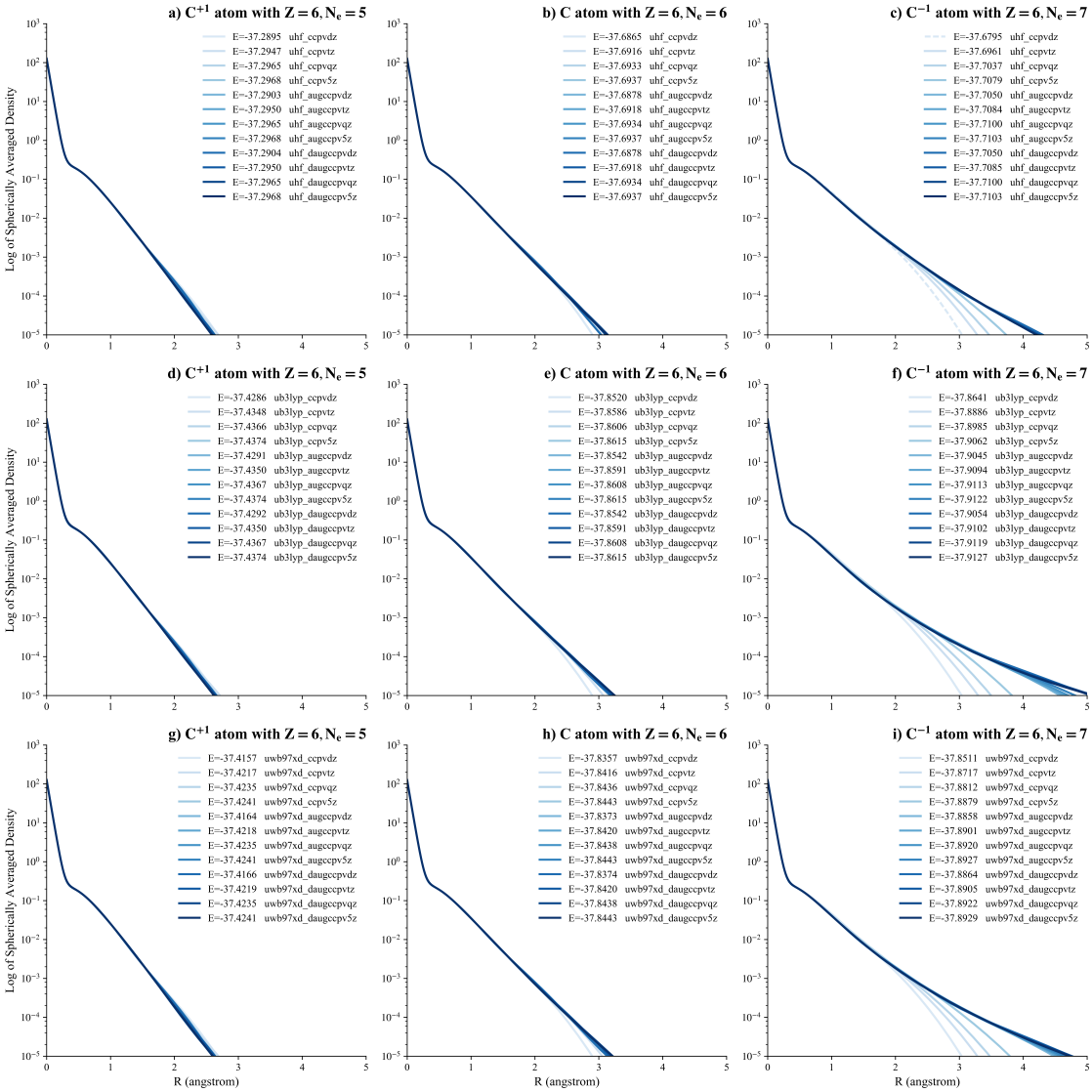


Figure 7.3 Log plot of spherically averaged density of nitrogen atomic species at various levels of theory and basis set. Unbound species are denoted with dashed lines. The energy values in plot labels are in atomic units.

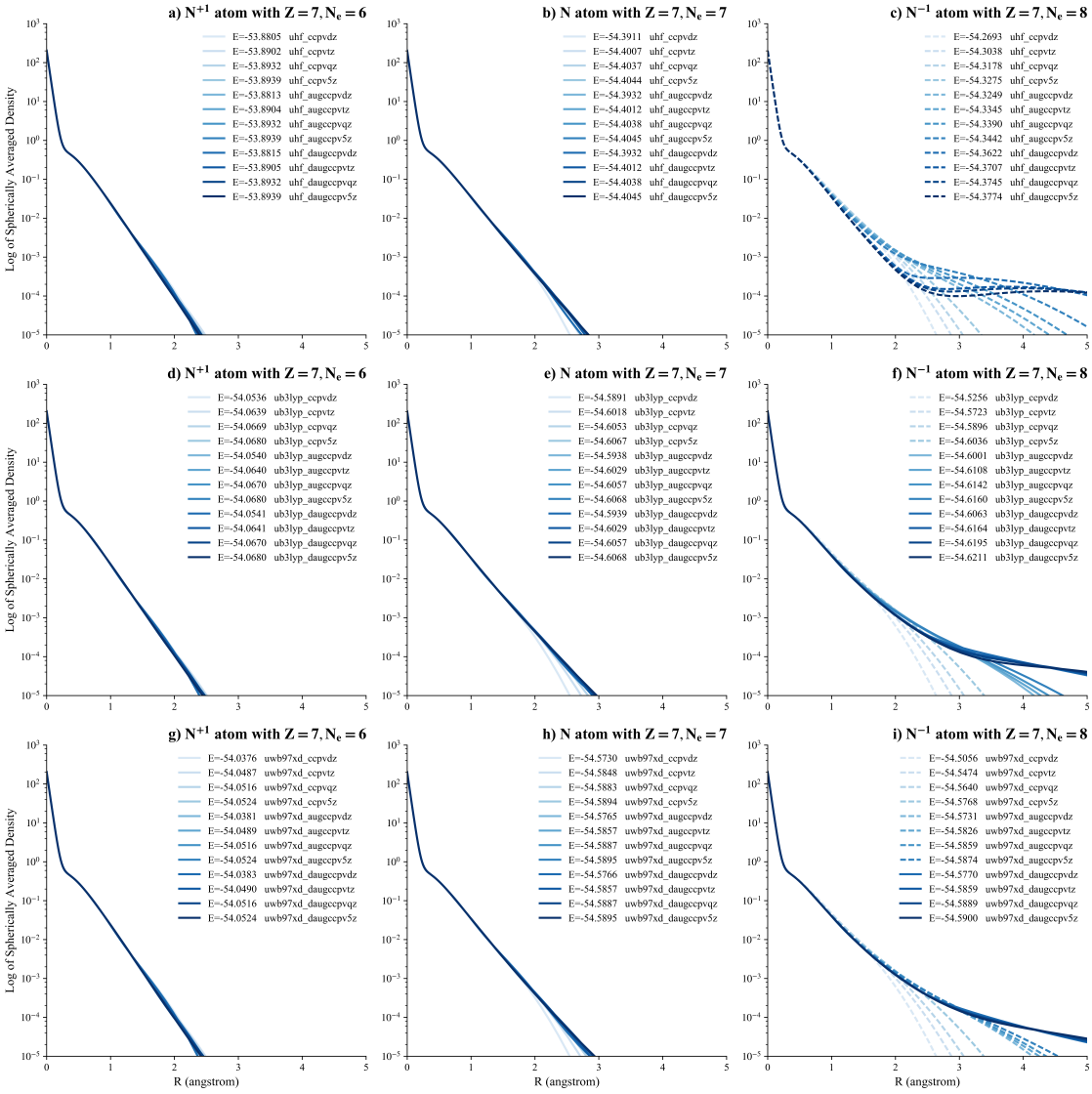


Figure 7.4 Log plot of spherically averaged density of oxygen atomic species at various levels of theory and basis set. Unbound species are denoted with dashed lines. The energy values in plot labels are in atomic units.

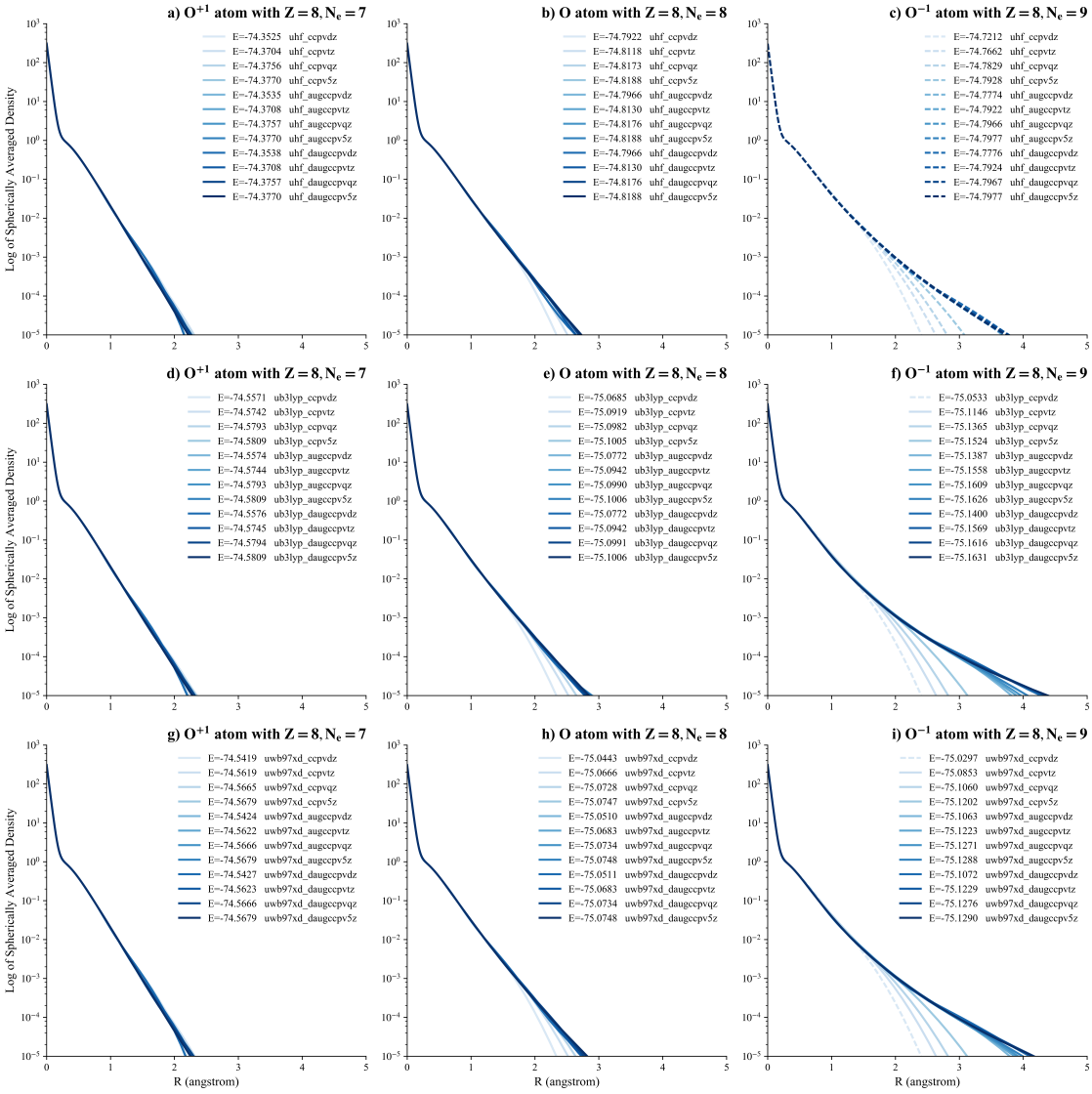
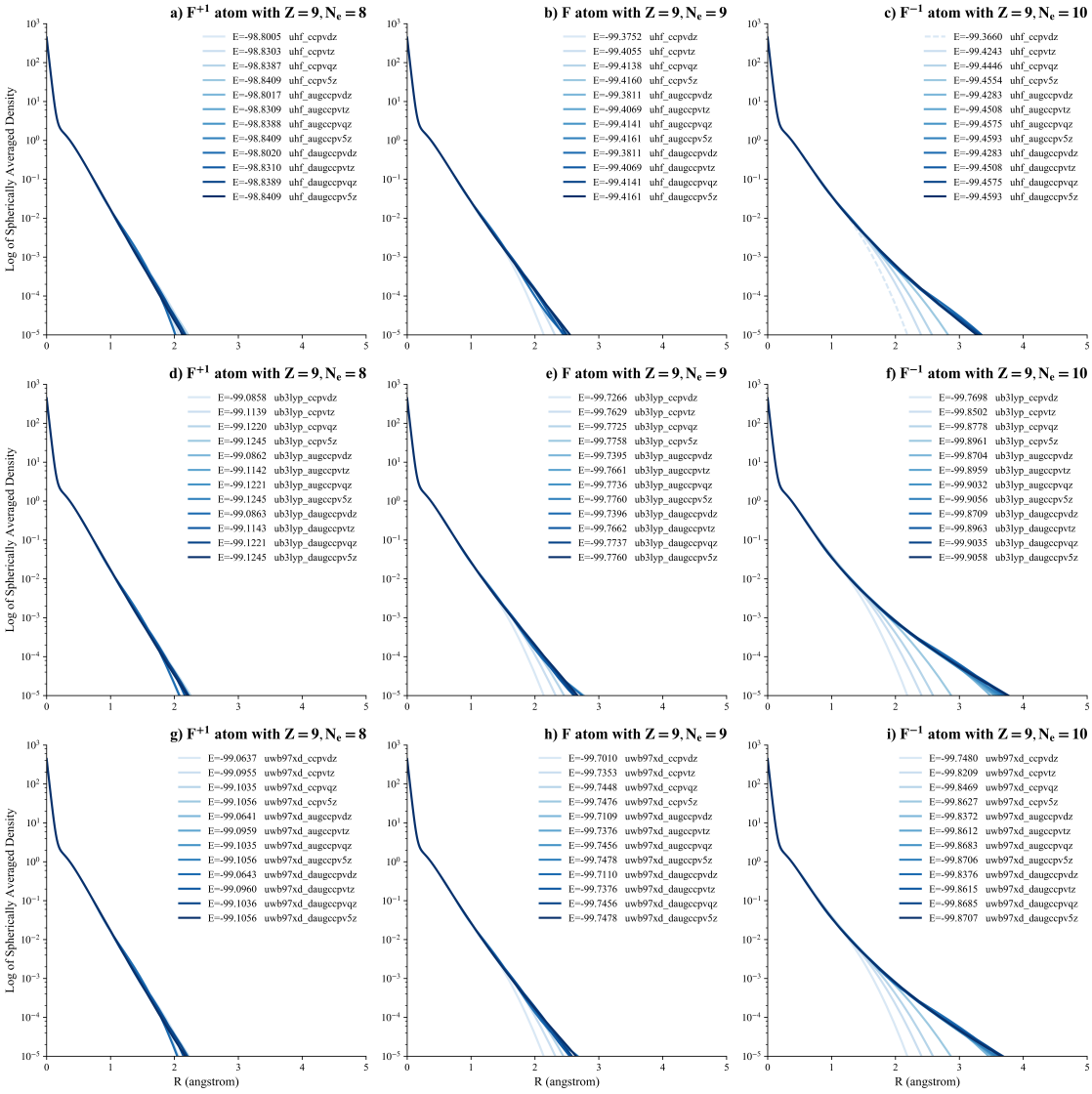


Figure 7.5 Log plot of spherically averaged density of fluorine atomic species at various levels of theory and basis set. Unbound species are denoted with dashed lines. The energy values in plot labels are in atomic units.



8 Bibliography

1. Parr, R. G.; Ayers, P. W.; Nalewajski, R. F., What Is an Atom in a Molecule? *The Journal of Physical Chemistry A* **2005,** *109* (17), 3957-3959.
2. Frenking, G.; Krapp, A., Unicorns in the world of chemical bonding models. *J. Comput. Chem.* **2006,** *28* (1), 15-24.
3. Matta, C. F.; Bader, R. F. W., An experimentalist's reply to "What is an atom in a molecule?". *J. Phys. Chem. A* **2006,** *110* (19), 6365-6371.
4. Ayers, P. L.; Boyd, R. J.; Bultinck, P.; Caffarel, M.; Carbó-Dorca, R.; Causá, M.; Cioslowski, J.; Contreras-Garcia, J.; Cooper, D. L.; Coppens, P.; Gatti, C.; Grabowsky, S.; Lazzeretti, P.; Macchi, P.; Martín Pendás, Á.; Popelier, P. L. A.; Ruedenberg, K.; Rzepa, H.; Savin, A.; Sax, A.; Schwarz, W. H. E.; Shahbazian, S.; Silvi, B.; Solà, M.; Tsirelson, V., Six questions on topology in theoretical chemistry. *Computational and Theoretical Chemistry* **2015,** *1053*, 2-16.
5. Gonthier, J. F.; Steinmann, S. N.; Wodrich, M. D.; Corminboeuf, C., Quantification of "fuzzy" chemical concepts: a computational perspective. *Chem. Soc. Rev.* **2012,** *41* (13), 4671-4687.
6. Mulliken, R. S., Electronic Population Analysis on LCAO–MO Molecular Wave Functions. I. *The Journal of Chemical Physics* **1955,** *23* (10), 1833-1840.

7. Mulliken, R. S., Electronic Population Analysis on LCAO–MO Molecular Wave Functions. II. Overlap Populations, Bond Orders, and Covalent Bond Energies. *The Journal of Chemical Physics* **1955,** *23* (10), 1841-1846.

8. Mulliken, R. S., Electronic Population Analysis on LCAO‑MO Molecular Wave Functions. III. Effects of Hybridization on Overlap and Gross AO Populations. *The Journal of Chemical Physics* **1955,** *23* (12), 2338-2342.

9. Mulliken, R. S., Electronic Population Analysis on LCAO‑MO Molecular Wave Functions. IV. Bonding and Antibonding in LCAO and Valence‑Bond Theories. *The Journal of Chemical Physics* **1955,** *23* (12), 2343-2346.

10. Löwdin, P. O., On the Non‑Orthogonality Problem Connected with the Use of Atomic Wave Functions in the Theory of Molecules and Crystals. *The Journal of Chemical Physics* **1950,** *18* (3), 365-375.

11. Löwdin, P.-O., On the Nonorthogonality Problem. In *Advances in Quantum Chemistry Volume 5*, Elsevier: 1970; pp 185-199.

12. Davidson, E. R., Electronic Population Analysis of Molecular Wavefunctions. *The Journal of Chemical Physics* **1967,** *46* (9), 3320-3324.

13. Moffitt, W., Atoms in Molecules and Crystals. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **1951,** *210* (1101), 245-268.

14. Reed, A. E.; Weinstock, R. B.; Weinhold, F., Natural population analysis. *The Journal of Chemical Physics* **1985,** *83* (2), 735-746.

15. Weinhold, F.; Landis, C. R., *Valency and Bonding*. Cambridge University Press: 2005.

16. Lu, W. C.; Wang, C. Z.; Schmidt, M. W.; Bytautas, L.; Ho, K. M.; Ruedenberg, K., Molecule intrinsic minimal basis sets. I. Exact resolution of ab initio optimized molecular orbitals in terms of deformed atomic minimal-basis orbitals. *The Journal of Chemical Physics* **2004,** *120* (6), 2629-2637.

17. Lu, W. C.; Wang, C. Z.; Schmidt, M. W.; Bytautas, L.; Ho, K. M.; Ruedenberg, K., Molecule intrinsic minimal basis sets. II. Bonding analyses for Si4H6 and Si2 to Si10. *The Journal of Chemical Physics* **2004,** *120* (6), 2638-2651.

18. West, A. C.; Schmidt, M. W.; Gordon, M. S.; Ruedenberg, K., A comprehensive analysis of molecule-intrinsic quasi-atomic, bonding, and correlating orbitals. I. Hartree-Fock wave functions. *The Journal of Chemical Physics* **2013,** *139* (23), 234107.

19. Knizia, G., Intrinsic Atomic Orbitals: An Unbiased Bridge between Quantum Theory and Chemical Concepts. *Journal of Chemical Theory and Computation* **2013,** *9* (11), 4834-4843.

20. Politzer, P.; Harris, R. R., Electronic density distribution in nitric oxide. *J. Am. Chem. Soc.* **1970,** *92* (7), 1834-1836.

21. Hirshfeld, F. L., Bonded-atom fragments for describing molecular charge densities. *Theor. Chim. Acta* **1977,** *44* (2), 129-138.

22. Eugen Schwarz, W. H., Richard F. Bader: Atoms in Molecules (A Quantum Theory) Clarendon Press 1990, Oxford. ISBN 019-855-1681, 438 pages, Preis:

£50. *Berichte der Bunsengesellschaft für physikalische Chemie* **1991,** *95* (10), 1308-1308.

23. Bader, R. F. W.; Beddall, P. M., Virial Field Relationship for Molecular Charge Distributions and the Spatial Partitioning of Molecular Properties. *The Journal of Chemical Physics* **1972,** *56* (7), 3320-3329.

24. Nalewajski, R. F.; witka, E. b.; Michalak, A., Information distance analysis of molecular electron densities. *Int. J. Quantum Chem* **2002,** *87* (4), 198-213.

25. Nalewajski, R. F.; Parr, R. G., Information Theory Thermodynamics of Molecules and Their Hirshfeld Fragments. *The Journal of Physical Chemistry A* **2001,** *105* (31), 7391-7400.

26. Nalewajski, R. F.; Parr, R. G., Information theory, atoms in molecules, and molecular similarity. *Proceedings of the National Academy of Sciences* **2000,** *97* (16), 8879-8882.

27. Ayers, P. W., Atoms in molecules, an axiomatic approach. I. Maximum transferability. *The Journal of Chemical Physics* **2000,** *113* (24), 10886-10898.

28. Heidar-Zadeh, F.; Ayers, P. W., How pervasive is the Hirshfeld partitioning? *J. Chem. Phys.* **2015,** *142* (4), 044107.

29. Heidar-Zadeh, F.; Ayers, P. W.; Bultinck, P., Deriving the Hirshfeld partitioning using distance metrics. *J. Chem. Phys.* **2014,** *141* (9).

30. Bader, R. F. W.; Nguyen-Dang, T. T., Quantum Theory of Atoms in Molecules–Dalton Revisited. In *Adv. Quantum Chem.*, Elsevier: 1981; pp 63-124.

31. Manz, T. A.; Sholl, D. S., Improved Atoms-in-Molecule Charge Partitioning Functional for Simultaneously Reproducing the Electrostatic Potential and Chemical States in Periodic and Nonperiodic Materials. *Journal of Chemical Theory and Computation* **2012,** *8* (8), 2844-2867.

32. Limas, N. G.; Manz, T. A., Introducing DDEC6 atomic population analysis: part 2. Computed results for a wide range of periodic and nonperiodic materials. *RSC Adv.* **2016,** *6* (51), 45727-45747.

33. Manz, T. A.; Limas, N. G., Introducing DDEC6 atomic population analysis: part 1. Charge partitioning theory and methodology. *RSC Adv.* **2016,** *6* (53), 47771-47801.

34. Goli, M.; Shahbazian, S., Toward the multi-component quantum theory of atoms in molecules: a variational derivation. *Theor. Chem. Acc.* **2013,** *132* (6).

35. Goli, M.; Shahbazian, S., The two-component quantum theory of atoms in molecules (TC-QTAIM): foundations. *Theor. Chem. Acc.* **2012,** *131* (5).

36. Bader, R. F. W., A quantum theory of molecular structure and its applications. *Chem. Rev.* **1991,** *91* (5), 893-928.

37. Hohenberg, P.; Kohn, W., Inhomogeneous Electron Gas. *Phys. Rev.* **1964,** *136* (3B), B864-B871.

38. Calais, J.-L., Density-functional theory of atoms and molecules. R.G. Parr and W. Yang, Oxford University Press, New York, Oxford, 1989. IX + 333 pp. Price £45.00. *Int. J. Quantum Chem* **1993,** *47* (1), 101-101.

39. Kohn, W.; Becke, A. D.; Parr, R. G., Density Functional Theory of Electronic Structure. *The Journal of Physical Chemistry* **1996,** *100* (31), 12974-12980.

40. Capitani, J. F.; Nalewajski, R. F.; Parr, R. G., Non-Born–Oppenheimer density functional theory of molecular systems. *The Journal of Chemical Physics* **1982,** *76* (1), 568-573.

41. Chakraborty, A.; Pak, M. V.; Hammes-Schiffer, S., Development of Electron-Proton Density Functionals for Multicomponent Density Functional Theory. *Phys. Rev. Lett.* **2008,** *101* (15).

42. Chakraborty, A.; Pak, M. V.; Hammes-Schiffer, S., Properties of the exact universal functional in multicomponent density functional theory. *The Journal of Chemical Physics* **2009,** *131* (12), 124115.

43. Ayers, P. W., Density bifunctional theory using the mass density and the charge density. *Theor. Chem. Acc.* **2005,** *115* (4), 253-256.

44. Heidarzadeh, F.; Shahbazian, S., The Quantum Divided Basins: A New Class of Quantum Subsystems. *Int. J. Quantum Chem* **2011,** *111* (12), 2788-2801.

45. Zadeh, F. H.; Shahbazian, S., Toward a fuzzy atom view within the context of the quantum theory of atoms in molecules: quasi-atoms. *Theor. Chem. Acc.* **2011,** *128* (2), 175-181.

46. Bultinck, P.; Ayers, P. W.; Fias, S.; Tiels, K.; Van Alsenoy, C., Uniqueness and basis set dependence of iterative Hirshfeld charges. *Chem. Phys. Lett.* **2007,** *444* (1-3), 205-208.

47. Kato, T., On the eigenfunctions of many-particle systems in quantum mechanics. *Commun. Pure Appl. Math.* **1957,** *10* (2), 151-177.

48. Steiner, E., Charge Densities in Atoms. *The Journal of Chemical Physics* **1963,** *39* (9), 2365-2366.

49. Weinstein, H.; Politzer, P.; Srebrenik, S., A misconception concerning the electronic density distribution of an atom. *Theor. Chim. Acta* **1975,** *38* (2), 159-163.

50. Sanderson, R. T., An Interpretation of Bond Lengths and a Classification of Bonds. *Science* **1951,** *114* (2973), 670-672.

51. Parr, R. G.; Donnelly, R. A.; Levy, M.; Palke, W. E., Electronegativity: The density functional viewpoint. *The Journal of Chemical Physics* **1978,** *68* (8), 3801-3807.

52. Ayers, P. W., On the electronegativity nonlocality paradox. *Theor. Chem. Acc.* **2007,** *118* (2), 371-381.

53. Verstraelen, T.; Vandenbrande, S.; Ayers, P. W., Direct computation of parameters for accurate polarizable force fields. *J. Chem. Phys.* **2014,** *141* (19).

54. Bultinck, P.; Jayatilaka, D.; Cardenas, C., A problematic issue for atoms in molecules: Impact of (quasi-)degenerate states on Quantum Theory Atoms in Molecules and Hirshfeld-I properties. *Computational and Theoretical Chemistry* **2015,** *1053*, 106-111.

55. Bultinck, P.; Cardenas, C.; Fuentealba, P.; Johnson, P. A.; Ayers, P. W., Atomic Charges and the Electrostatic Potential Are Ill-Defined in Degenerate Ground States. *Journal of Chemical Theory and Computation* **2013,** *9* (11), 4779-4788.

56. Cárdenas, C.; Ayers, P. W.; Cedillo, A., Reactivity indicators for degenerate states in the density-functional theoretic chemical reactivity theory. *The Journal of Chemical Physics* **2011,** *134* (17), 174103.

57. Ayers, P. W., Axiomatic formulations of the Hohenberg-Kohn functional. *Phys. Rev. A* **2006,** *73* (1).

58. Manz, T. A.; Sholl, D. S., Chemically Meaningful Atomic Charges That Reproduce the Electrostatic Potential in Periodic and Nonperiodic Materials. *Journal of Chemical Theory and Computation* **2010,** *6* (8), 2455-2468.

59. Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G., Charge Model 5: An Extension of Hirshfeld Population Analysis for the Accurate Description of Molecular Interactions in Gaseous and Condensed Phases. *Journal of Chemical Theory and Computation* **2012,** *8* (2), 527-541.

60. Bultinck, P.; Van Alsenoy, C.; Ayers, P. W.; Carbó-Dorca, R., Critical analysis and extension of the Hirshfeld atoms in molecules. *The Journal of Chemical Physics* **2007,** *126* (14), 144111.

61. Perdew, J. P.; Parr, R. G.; Levy, M.; Balduz, J. L., Density-Functional Theory for Fractional Particle Number: Derivative Discontinuities of the Energy. *Phys. Rev. Lett.* **1982,** *49* (23), 1691-1694.

62. Yang, W.; Zhang, Y.; Ayers, P. W., Degenerate Ground States and a Fractional Number of Electrons in Density and Reduced Density Matrix Functional Theory. *Phys. Rev. Lett.* **2000,** *84* (22), 5172-5175.

63. Ayers, P. W., The dependence on and continuity of the energy and other molecular properties with respect to the number of electrons. *J. Math. Chem.* **2008,** *43* (1), 285-303.

64. Ayers, P. W.; Levy, M., Perspective on "Density functional approach to the frontier-electron theory of chemical reactivity". In *Theor. Chem. Acc.*, Springer Berlin Heidelberg: 2000; pp 353-360.

65. Van Damme, S.; Bultinck, P.; Fias, S., Electrostatic Potentials from Self-Consistent Hirshfeld Atomic Charges. *Journal of Chemical Theory and Computation* **2009,** *5* (2), 334-340.

66. Manz, T. A., Comment on "Extending Hirshfeld-I to Bulk and Periodic Materials". *J. Comput. Chem.* **2013,** *34* (5), 418-421.

67. Vanpoucke, D. E. P.; Bultinck, P.; Van Driessche, I., Extending Hirshfeld-I to bulk and periodic materials. *J. Comput. Chem.* **2013,** *34* (5), 405-417.

68. Vanpoucke, D. E. P.; Van Driessche, I.; Bultinck, P., Reply to 'comment on "extending hirshfeld-I to bulk and periodic materials"'. *J. Comput. Chem.* **2013,** *34* (5), 422-427.

69. Geldof, D.; Krishtal, A.; Blockhuys, F.; Van Alsenoy, C., An Extension of the Hirshfeld Method to Open Shell Systems Using Fractional Occupations. *Journal of Chemical Theory and Computation* **2011,** *7* (5), 1328-1335.

70. Geldof, D.; Krishtal, A.; Blockhuys, F.; Van Alsenoy, C., FOHI-D: An iterative Hirshfeld procedure including atomic dipoles. *J. Chem. Phys.* **2014,** *140* (14).

71. Finzel, K.; Pendas, A. M.; Francisco, E., Efficient algorithms for Hirshfeld-I charges. *J. Chem. Phys.* **2015,** *143* (8).

72. Ghillemijn, D.; Bultinck, P.; Van Neck, D.; Ayers, P. W., A Self-Consistent Hirshfeld Method for the Atom in the Molecule Based on Minimization of Information Loss. *J. Comput. Chem.* **2011,** *32* (8), 1561-1567.

73. Verstraelen, T.; Ayers, P. W.; Van Speybroeck, V.; Waroquier, M., Hirshfeld-E Partitioning: AIM Charges with an Improved Trade-off between Robustness and Accurate Electrostatics. *Journal of Chemical Theory and Computation* **2013,** *9* (5), 2221-2225.

74. Parr, R. G.; Yang, W., Density functional approach to the frontier-electron theory of chemical reactivity. *J. Am. Chem. Soc.* **1984,** *106* (14), 4049-4050.

75. Yang, W.; Parr, R. G.; Pucci, R., Electron density, Kohn–Sham frontier orbitals, and Fukui functions. *The Journal of Chemical Physics* **1984,** *81* (6), 2862-2863.

76. Ayers, P.; Yang, W.; Bartolotti, L., Fukui Function. In *Chemical Reactivity Theory*, CRC Press: 2009.

77. Lillestolen, T. C.; Wheatley, R. J., Atomic charge densities generated using an iterative stockholder procedure. *J. Chem. Phys.* **2009,** *131* (14).

78. Lillestolen, T. C.; Wheatley, R. J., Redefining the atom: atomic charge densities produced by an iterative stockholder approach. *Chem. Commun.* **2008,** (45), 5909.

79. Bultinck, P.; Cooper, D. L.; Van Neck, D., Comparison of the Hirshfeld-I and iterated stockholder atoms in molecules schemes. *PCCP* **2009,** *11* (18), 3424.

80. Verstraelen, T.; Ayers, P. W.; Van Speybroeck, V.; Waroquier, M., The conformational sensitivity of iterative stockholder partitioning schemes. *Chem. Phys. Lett.* **2012,** *545*, 138-143.

81. Lee, L. P.; Limas, N. G.; Cole, D. J.; Payne, M. C.; Skylaris, C.-K.; Manz, T. A., Expanding the Scope of Density Derived Electrostatic and Chemical Charge Partitioning to Thousands of Atoms. *Journal of Chemical Theory and Computation* **2014,** *10* (12), 5377-5390.

82. Verstraelen, T.; Vandenbrande, S.; Heidar-Zadeh, F.; Vanduyfhuys, L.; Van Speybroeck, V.; Waroquier, M.; Ayers, P. W., Minimal Basis Iterative Stockholder: Atoms in Molecules for Force-Field Development. *Journal of Chemical Theory and Computation* **2016,** *12* (8), 3894-3912.

83. Vandenbrande, S.; Waroquier, M.; Speybroeck, V. V.; Verstraelen, T., The Monomer Electron Density Force Field (MEDFF): A Physically Inspired Model for Noncovalent Interactions. *Journal of Chemical Theory and Computation* **2017,** *13* (1), 161-179.

84. Bader, R. F. W.; Stephens, M. E., Spatial localization of the electronic pair and number distributions in molecules. *J. Am. Chem. Soc.* **1975,** *97* (26), 7391-7399.

85. Hansen, N. K.; Coppens, P., Electron population analysis of accurate diffraction data. 6. Testing aspherical atom refinements on small-molecule data sets. *Acta Crystallographica Section A* **1978,** *34* (NOV), 909-921.

86. Srebrenik, S.; Bader, R. F. W.; Nguyen-Dang, T. T., Subspace quantum mechanics and the variational principle. *The Journal of Chemical Physics* **1978,** *68* (8), 3667-3679.

87. Bader, R. F. W.; Srebrenik, S.; Nguyen-Dang, T. T., Subspace quantum dynamics and the quantum action principle. *The Journal of Chemical Physics* **1978,** *68* (8), 3680-3691.

88. Cohen, L., Representable local kinetic energy. *The Journal of Chemical Physics* **1984,** *80* (9), 4277-4279.

89. Bader, R. F. W., Principle of stationary action and the definition of a proper open system. *Physical Review B* **1994,** *49* (19), 13348-13356.

90. Bader, R. F. W., Everyman's Derivation of the Theory of Atoms in Molecules. *The Journal of Physical Chemistry A* **2007,** *111* (32), 7966-7972.

91. Ayers, P. W.; Parr, R. G.; Nagy, A., Local kinetic energy and local temperature in the density-functional theory of electronic structure. *Int. J. Quantum Chem* **2002,** *90* (1), 309-326.

92. Anderson, J. S. M.; Ayers, P. W.; Hernandez, J. I. R., How Ambiguous Is the Local Kinetic Energy?†. *The Journal of Physical Chemistry A* **2010,** *114* (33), 8884-8895.

93. Cao, W. L.; Gatti, C.; MacDougall, P. J.; Bader, R. F. W., On the presence of non-nuclear attractors in the charge distributions of Li and Na clusters. *Chem. Phys. Lett.* **1987,** *141* (5), 380-385.

94. Cioslowski, J., Nonnuclear attractors in the lithium dimeric molecule. *The Journal of Physical Chemistry* **1990,** *94* (14), 5496-5498.

95. Pendás, A. M.; Blanco, M. A.; Costales, A.; Sánchez, P. M.; Luaña, V., Non-nuclear Maxima of the Electron Density. *Phys. Rev. Lett.* **1999,** *83* (10), 1930-1933.

96. Popelier, P. L. A.; Joubert, L.; Kosov, D. S., Convergence of the Electrostatic Interaction Based on Topological Atoms. *The Journal of Physical Chemistry A* **2001,** *105* (35), 8254-8261.

97. Popelier, P. L. A.; Rafat, M., The electrostatic potential generated by topological atoms: a continuous multipole method leading to larger convergence regions. *Chem. Phys. Lett.* **2003,** *376* (1-2), 148-153.

98. Rafat, M.; Popelier, P. L. A., The electrostatic potential generated by topological atoms. II. Inverse multipole moments. *The Journal of Chemical Physics* **2005,** *123* (20), 204103.

99. Rafat, M.; Popelier, P. L. A., Long range behavior of high-rank topological multipole moments. *J. Comput. Chem.* **2007,** *28* (4), 832-838.

100. Momany, F. A., Determination of partial atomic charges from ab initio molecular electrostatic potentials. Application to formamide, methanol, and formic acid. *The Journal of Physical Chemistry* **1978,** *82* (5), 592-601.

101. Cox, S. R.; Williams, D. E., Representation of the molecular electrostatic potential by a net atomic charge model. *J. Comput. Chem.* **1981,** *2* (3), 304-323.

102. Singh, U. C.; Kollman, P. A., An approach to computing electrostatis charges for molecules. *J. Comput. Chem.* **1984,** *5* (2), 129-145.

103. Besler, B. H.; Merz, K. M.; Kollman, P. A., Atomic charges derived from semiempirical methods. *J. Comput. Chem.* **1990,** *11* (4), 431-439.

104. Chirlian, L. E.; Francl, M. M., Atomic charges derived from electrostatic potentials: A detailed study. *J. Comput. Chem.* **1987,** *8* (6), 894-905.

105. Breneman, C. M.; Wiberg, K. B., Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J. Comput. Chem.* **1990,** *11* (3), 361-373.

106. Hu, H.; Lu, Z.; Yang, W., Fitting molecular electrostatic potentials from quantum mechanical calculations. *Journal of Chemical Theory and Computation* **2007,** *3* (3), 1004-1013.

107. Gaussian 09, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K.

Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2016.

108. Becke, A. D., Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988,** *38* (6), 3098-3100.

109. Lee, C.; Yang, W.; Parr, R. G., Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B* **1988,** *37* (2), 785-789.

110. Becke, A. D., Density‐functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics* **1993,** *98* (7), 5648-5652.

111. Chai, J.-D.; Head-Gordon, M., Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *PCCP* **2008,** *10* (44), 6615.

112. Dunning, T. H., Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *The Journal of Chemical Physics* **1989,** *90* (2), 1007-1023.

113. Woon, D. E.; Dunning, T. H., Gaussian basis sets for use in correlated molecular calculations. IV. Calculation of static electrical response properties. *The Journal of Chemical Physics* **1994,** *100* (4), 2975-2988.

114. Kendall, R. A.; Dunning, T. H.; Harrison, R. J., Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *The Journal of Chemical Physics* **1992,** *96* (9), 6796-6806.

115. Toon Verstraelen, Pawel Tecmer, Farnaz Heidar-Zadeh, Katharina Boguslawski, Matthew Chan, Yilin Zhao, Taewon D. Kim, Steven Vandenbrande, Derrick Yang, Cristina E. González-Espinoza, Stijn Fias, Peter A. Limacher, Diego Berrocal, Ali Malek, Paul W. Ayers HORTON 2.0.0, http://theochem.github.com/horton/, 2015

116. F. Heidar-Zadeh and P. W. Ayers "Is There are Variational Principle for Iterative Hirshfeld Charges?" (submitted)

117. Saha, S.; Roy, R. K.; Ayers, P. W., Are the Hirshfeld and Mulliken Population Analysis Schemes Consistent With Chemical Intuition? *Int. J. Quantum Chem* **2009,** *109* (9), 1790-1806.

118. Ayers, P. W., Information Theory, the Shape Function, and the Hirshfeld Atom. *Theor. Chem. Acc.* **2006,** *115* (5), 370-378.

119. Dunlap, B. I.; Rösch, N.; Trickey, S. B., Variational fitting methods for electronic structure calculations. *Mol. Phys.* **2010,** *108* (21-23), 3167-3180.

120. Dunlap, B. I., Robust and variational fitting: Removing the four-center integrals from center stage in quantum chemistry. *Journal of Molecular Structure: THEOCHEM* **2000,** *529* (1-3), 37-40.

121. Dunlap, B. I., Robust variational fitting: Gáspár's variational exchange can accurately be treated analytically. *Journal of Molecular Structure: THEOCHEM* **2000,** *501-502*, 221-228.

122. Dunlap, B. I., Robust and variational fitting. *PCCP* **2000,** *2* (10), 2113-2116.

123. Dunlap, B. I.; Connolly, J. W. D.; Sabin, J. R., On some approximations in applications of Xα theory. *The Journal of Chemical Physics* **1979,** *71* (8), 3396-3402.

124. Parr, R. G.; Bartolotti, L. J., Some remarks on the density functional theory of few-electron systems. *The Journal of Physical Chemistry* **1983,** *87* (15), 2810-2815.

125. Ayers, P.; Cedillo, A., Shape Function. In *Chemical Reactivity Theory*, CRC Press: 2009.

126. De Proft, F.; Ayers, P. W.; Sen, K. D.; Geerlings, P., On the importance of the "density per particle" (shape function) in the density functional theory. *The Journal of Chemical Physics* **2004,** *120* (21), 9969-9973.

127. Rong, C. Y.; Lu, T.; Ayers, P. W.; Chattaraj, P. K.; Liu, S. B., Scaling properties of information-theoretic quantities in density functional reactivity theory. *PCCP* **2015,** *17* (7), 4977-4988.

128. Borgoo, A.; Godefroid, M.; Indelicato, P.; De Proft, F.; Geerlings, P., Quantum similarity study of atomic density functions: Insights from information theory and the role of relativistic effects. *The Journal of Chemical Physics* **2007,** *126* (4), 044102.

129. Sen, K. D.; Proft, F. D.; Borgoo, A.; Geerlings, P., N-derivative of Shannon entropy of shape function for atoms. *Chem. Phys. Lett.* **2005,** *410* (1-3), 70-76.

130. Borgoo, A.; Godefroid, M.; Sen, K. D.; De Proft, F.; Geerlings, P., Quantum similarity of atoms: a numerical Hartree–Fock and Information Theory approach. *Chem. Phys. Lett.* **2004,** *399* (4-6), 363-367.

131. Rychlewski, J.; Parr, R. G., The atom in a molecule: A wave function approach. *The Journal of Chemical Physics* **1986,** *84* (3), 1696-1703.

132. Cohen, M. H.; Wasserman, A., On the Foundations of Chemical Reactivity Theory. *The Journal of Physical Chemistry A* **2007,** *111* (11), 2229-2242.

133. Elliott, P.; Cohen, M. H.; Wasserman, A.; Burke, K., Density Functional Partition Theory with Fractional Occupations. *Journal of Chemical Theory and Computation* **2009,** *5* (4), 827-833.

134. Tang, R.; Nafziger, J.; Wasserman, A., Fragment occupations in partition density functional theory. *PCCP* **2012,** *14* (21), 7780.

135. Zhang, Y.; Wasserman, A., Transferability of Atomic Properties in Molecular Partitioning: A Comparison. *Journal of Chemical Theory and Computation* **2010,** *6* (11), 3312-3318.

136. Kullback, S., *Information theory and statistics*. Dover Publications: Mineola, N.Y., 1997; p xv, 399 p.

137. Kullback, S.; Leibler, R. A., On Information and Sufficiency. *The Annals of Mathematical Statistics* **1951,** *22* (1), 79-86.

138. Ali, S. M.; Silvey, S. D., A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society Series B-Methodological* **1966,** *28*, 131-142.

139. Morimoto, T., Markov Processes and the H-Theorem. *J. Phys. Soc. Jpn.* **1963,** *18* (3), 328-331.

140. Csiszár, I., Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizitat von Markoffschen Ketten. *Magyar. Tud. Akad. Mat. Kutato Int. Kozl* **1963,** *8*, 85-108.

141. Liese, F.; Vajda, I., On Divergences and Informations in Statistics and Information Theory. *IEEE Trans. Inf Theory* **2006,** *52* (10), 4394-4412.

142. Österreicher, F.; Vajda, I., A new class of metric divergences on probability spaces and its applicability in statistics. *Annals of the Institute of Statistical Mathematics* **2003,** *55* (3), 639-653.

143. Deza, M. M.; Deza, E., Encyclopedia of Distances. Springer Berlin Heidelberg: 2013.

144. Tsallis, C., Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.* **1988,** *52* (1-2), 479-487.

145. Nielsen, F.; Nock, R., A closed-form expression for the Sharma–Mittal entropy of exponential families. *Journal of Physics A: Mathematical and Theoretical* **2011,** *45* (3), 032003.

146. Lenzi, E. K.; Mendes, R. S.; da Silva, L. R., Statistical mechanics based on Renyi entropy. *Physica A: Statistical Mechanics and its Applications* **2000,** *280* (3-4), 337-345.

147. Réyni, A. In *On measures of information and entropy*, Procedings of the fourth Berkeley symposium on mathematics, statistics, and probability, 1960; pp 547-561.

148. Nielsen, F.; Nock, R., A closed-form expression for the Sharma-Mittal entropy of exponential families. *Journal of Physics A-Mathematical and Theoretical* **2012,** *45* (3).

149. Gupta, H. C.; Sharma, B. D., On non-additive measures of inaccuracy. *Czechloslovak Mathematics Journal* **1976,** *26*, 584-595.

150. Sharma, B. D.; Taneja, I. J., Entropy of type (alpha,beta) and other generalized measures in information theory. *Metrika* **1975,** *22*, 205-215.

151. Sharma, B. D.; Mittal, D. P., New nonadditive measures of relative information. *Journal of Combinatorics Information and System Sciences* **1977,** *2*, 122-132.

152. Masi, M., A step beyond Tsallis and Rényi entropies. *Phys. Lett. A* **2005,** *338* (3-5), 217-224.

153. Esteban, M. D.; Morales, D., A summary on entropy statistics. *Kybernetika* **1995,** *31* (4), 337-346.

154. Tsekouras, G. A.; Tsallis, C., Generalized entropy arising from a distribution ofqindices. *Physical Review E* **2005,** *71* (4).

155. Liu, M.; Vemuri, B. C.; Amari, S.-I.; Nielsen, F., Total Bregman divergence and its applications to shape retrieval. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE: 2010.

156. Nielsen, F.; Nock, R., Total Jensen divergences: Definition, properties and clustering. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE: 2015.

157. Nock, R.; Nielsen, F.; Amari, S.-I., On Conformal Divergences and Their Population Minimizers. *IEEE Trans. Inf Theory* **2016,** *62* (1), 527-538.

158. Chernoff, H., A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on the sum of Observations. *The Annals of Mathematical Statistics* **1952,** *23* (4), 493-507.

159. Verstraelen, T.; Vandenbrande, S.; Ayers, P. W., Direct computation of parameters for accurate polarizable force fields. *The Journal of Chemical Physics* **2014,** *141* (19), 194114.

160. Skilling, J., The Axioms of Maximum Entropy. In *Maximum-Entropy and Bayesian Methods in Science and Engineering*, Springer Netherlands: 1988; pp 173-187.

161. Skilling, J., Classic Maximum Entropy. In *Maximum Entropy and Bayesian Methods*, Springer Netherlands: 1989; pp 45-52.

162. Zhu, H.; Rohwer, R., Bayesian invariant measurements of generalization. *Neural Processing Letters* **1995,** *2* (6), 28-31.

163. Zhang, J., Divergence Function, Duality, and Convex Analysis. *Neural Computation* **2004,** *16* (1), 159-195.

164. Minka, T. In *Divergence measures and message passing*, (Technical Report) Microsoft Research: 2005.

165. Vöhringer-Martinez, E.; Verstraelen, T.; Ayers, P. W., The Influence of Ser-154, Cys-113, and the Phosphorylated Threonine Residue on the Catalytic Reaction Mechanism of Pin1. *The Journal of Physical Chemistry B* **2014,** *118* (33), 9871-9880.

166. Angulo, J. C.; Dehesa, J. S., Atomic systems with a completely monotonic electron density. *Phys. Rev. A* **1991,** *44* (3), 1516-1522.

167. Angulo, J. C.; Dehesa, J. S., Charge monotonicity of atomic systems and radial expectation values. *Zeitschrift für Physik D Atoms, Molecules and Clusters* **1993,** *25* (4), 287-293.

168. Ayers, P. W.; Morrison, R. C.; Roy, R. K., Variational principles for describing chemical reactions: Condensed reactivity indices. *J. Chem. Phys.* **2002,** *116* (20), 8731-8744.

169. Verstraelen, T.; Ayers, P. W.; Van Speybroeck, V.; Waroquier, M., Extended Hirshfeld: Atomic charges that combine accurate electrostatics with transferability. *Abstr. Pap. Am. Chem. Soc.* **2014,** *247*, 1.

170. F. Heidar-Zadeh, I. Vinogradov, P. W. Ayers, "Hirshfeld Partitioning from Nonextensive Entropies" (submitted).

171. Heidar-Zadeh, F.; Ayers, P. W., Fuzzy atoms in molecules from Bregman divergences. *Theor. Chem. Acc.* **2017,** *136* (8), 92.

172. F. Heidar-Zadeh, P. W. Ayers, T. Verstraelen, I. Vinogradov, E. Vohringer-Martinez, P. Bultinck, Information-Theoretic Approaches to Atoms-in-Molecules: Hirshfeld Family of Partitioning Schemes (submitted)

173. F. Heidar-Zadeh, E. Vohringer-Martinez, and P. W. Ayers, Constrained Iterative Hirshfeld Charges: A Variational Approach (submitted).

174. Gallup, G. A., *Valence bond methods: Theory and applications*. Cambridge UP: Cambridge, 2002.

175. Wu, W.; Su, P. F.; Shaik, S.; Hiberty, P. C., Classical Valence Bond Approach by Modern Methods. *Chem. Rev.* **2011,** *111* (11), 7557-7593.

176. Shaik, S.; Hiberty, P. C., A primer on qualitative valence bond theory - a theory coming of age. *Wiley Interdisciplinary Reviews-Computational Molecular Science* **2011,** *1* (1), 18-29.

177. Shaik, S.; Hiberty, P. C., *A chemist's guide to valence bond theory*. Wiley: Hoboken, 2008.

178. Katriel, J.; Davidson, E. R., Asymptotic behavior of atomic and molecular wave functions. *Proceedings of the National Academy of Sciences* **1980,** *77*, 4403-4406.

179. Morrell, M. M.; Parr, R. G.; Levy, M., Calculation of ionization potentials from density matrixes and natural functions, and the long-range behavior of natural orbitals and electron density. *J. Chem. Phys.* **1975,** *62*, 549-554.

180. Ayers, P. W., Density per particle as a descriptor of coulombic systems. *Proceedings of the National Academy of Sciences* **2000,** *97*, 1959-1964.

181. Ayers, P. W.; Levy, M.; Nagy, A., Time-independent density-functional theory for excited states of Coulomb systems. *Phys. Rev. A* **2012,** *85* (4), 7.

182. Ayers, P. W.; Day, O. W.; Morrison, R. C., Analysis of density functionals and their density tails in H 2 *Int. J. Quantum Chem* **1998,** *69*, 541-550.

183. Similar but less prevalent problems occur for molecules containing (noble gasses, Be, Mg, Mn,..) without bound anions.

184. Gaussian 09, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2016.

185. Weigend, F.; Ahlrichs, R., Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *PCCP* **2005,** *7*, 3297-3305.

186. Lee, C.; Yang, W.; Parr, R. G., Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B* **1988,** *37*, 785-789.

187. Becke, A. D., Density-functional thermochemistry. 3. The role of exact exchange. *J. Chem. Phys.* **1993,** *98*, 5648-5652.

188. Toon Verstraelen, Pawel Tecmer, Farnaz Heidar-Zadeh, Katharina Boguslawski, Matthew Chan, Yilin Zhao, Taewon D. Kim, Steven Vandenbrande, Derrick Yang, Cristina E. González-Espinoza, Stijn Fias, Peter A. Limacher, Diego Berrocal, Ali Malek, Paul W. Ayers HORTON 2.0.0, http://theochem.github.com/horton/, 2015

189. Heidar-Zadeh, F.; Vinogradov, I.; Ayers, P. W., Hirshfeld partitioning from non-extensive entropies. *Theor. Chem. Acc.* **2017,** *136* (4), 54.

190. Cárdenas, C.; Heidar-Zadeh, F.; Ayers, P. W., Benchmark values of chemical potential and chemical hardness for atoms and atomic ions (including unstable anions) from the energies of isoelectronic series. *Phys. Chem. Chem. Phys.* **2016,** *18* (36), 25721-25734.

191. Bader, R. F. W., Atoms in molecules. *Acc. Chem. Res.* **1985,** *18* (1), 9-15.

192. Chen, D. L.; Stern, A. C.; Space, B.; Johnson, J. K., Atomic Charges Derived from Electrostatic Potentials for Molecular and Periodic Systems. *J. Phys. Chem. A* **2010,** *114* (37), 10225-10233.

193. Heidar-Zadeh, F.; Ayers, P. W.; Bultinck, P., Deriving the Hirshfeld partitioning using distance metrics. *The Journal of Chemical Physics* **2014,** *141* (9), 094103.

194. F. Heidar-Zadeh and P. W. Ayers (submitted)

195. Lin, J., Divergence measures based on the Shannon entropy. *IEEE Trans. Inf Theory* **1991,** *37* (1), 145-151.

196. Ayers, P. W., Atoms in molecules, an axiomatic approach. I. Maximum transferability. *J. Chem. Phys.* **2000,** *113*, 10886-10898.

197. Zhao, Q.; Parr, R. G., Constrained-search method to determine electronic wave functions from electronic densities. *J. Chem. Phys.* **1993,** *98* (1), 543-8.

198. Zhao, Q.; Morrison, R. C.; Parr, R. G., From Electron-Densities to Kohn-Sham Kinetic Energies, Orbital Energies, Exchange-Correlation Potentials, and Exchange- Correlation Energies. *Phys. Rev. A* **1994,** *50*, 2138-2142.

199. Roothaan, C. C. J., New developments in molecular orbital theory. *Rev. Mod. Phys.* **1951,** *23* (2), 69-89.

200. Mori-Sanchez, P.; Cohen, A. J.; Yang, W. T., Many-electron self-interaction error in approximate density functionals. *J. Chem. Phys.* **2006,** *125*, 201102.

201. Cohen, A. J.; Mori-Sanchez, P.; Yang, W. T., Second-Order Perturbation Theory with Fractional Charges and Fractional Spins. *Journal of Chemical Theory and Computation* **2009,** *5*, 786-792.

202. Yang, W.; Mori-Sanchez, P.; Cohen, A. J., Extension of many-body theory and approximate density functionals to fractional charges and fractional spins. *J. Chem. Phys.* **2013,** *139* (10).

203. Ruzsinszky, A.; Perdew, J. P.; Csonka, G. I.; Vydrov, O. A.; Scuseria, G. E., Spurious fractional charge on dissociated atoms: Pervasive and resilient self-interaction error of common density functionals. *J. Chem. Phys.* **2006,** *125*.

204. Ruzsinszky, A.; Perdew, J. P.; Csonka, G. I.; Vydrov, O. A.; Scuseria, G. E., Density Functionals that are one- and two- are not always many-electron self-interaction-free, as shown for H2+, He2+, LiH+, and Ne2+. *J. Chem. Phys.* **2007,** *126*, 104102.

205. Melin, J.; Ayers, P. W.; Ortiz, J. V., Removing electrons can increase the electron density: A computational study of negative Fukui functions. *J. Phys. Chem. A* **2007,** *111*, 10017-10019.

206. Bultinck, P.; Carbó-Dorca, R., Negative and infinite Fukui functions: The role of diagonal dominance in the hardness matrix. *J. Math. Chem.* **2003,** *34*, 67-74.

207. Bultinck, P.; Clarisse, D.; Ayers, P. W.; Carbo-Dorca, R., The Fukui matrix: a simple approach to the analysis of the Fukui function and its positive character. *PCCP* **2011,** *13*, 6110-6115.

208. Echegaray, E.; Cardenas, C.; Rabi, S.; Rabi, N.; Lee, S.; Zadeh, F. H.; Toro-Labbe, A.; Anderson, J. S. M.; Ayers, P. W., In pursuit of negative Fukui functions: examples where the highest occupied molecular orbital fails to dominate the chemical reactivity. *J. Mol. Model.* **2013,** *19* (7), 2779-2783.

209. Echegaray, E.; Rabi, S.; Cardenas, C.; Zadeh, F. H.; Rabi, N.; Lee, S.; Anderson, J. S. M.; Toro-Labbe, A.; Ayers, P. W., In pursuit of negative Fukui functions: molecules with very small band gaps. *J. Mol. Model.* **2014,** *20* (3).

210. Lowdin, P. O., Quantum theory of many-particle systems. I. Physical Interpretation by means of density matrices, natural spin-orbitals, and convergence problems in the method of configuration interaction. *Phys.Rev.* **1955,** *97*, 1474-1489.