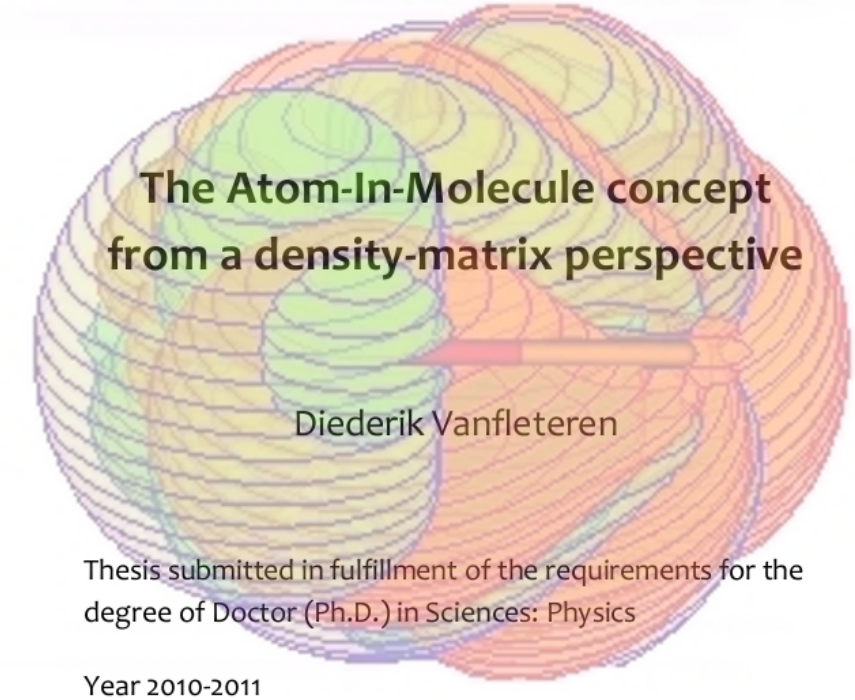
Faculty of Sciences
Departement of Physics and Astronomy



The Atom-In-Molecule concept from a density-matrix perspective

Diederik Vanfleteren

Thesis submitted in fulfillment of the requirements for the degree of Doctor (Ph.D.) in Sciences: Physics

Year 2010-2011

Supervisors: Prof. Dr. Dimitri Van Neck
Prof. Dr. Patrick Bultinck

Faculty of Sciences
Departement of Physics and Astronomy

The Atom-In-Molecule concept from a density-matrix perspective

Diederik Vanfleteren

Thesis submitted in fulfillment of the requirements for the degree of Doctor (Ph.D.) in Sciences: Physics

Year 2010-2011

Supervisors: Prof. Dr. Dimitri Van Neck
Prof. Dr. Patrick Bultinck

Acknowledgement

In the first place, I would like to thank my promotor Prof. Dimitri Van Neck. Without his firm theoretical support, the most important results of this Ph.D would not have been obtained. In particular, I acknowledge his continuous effort and investment of time to help young researchers, irrespective of their experience and capacities. In the second place, I would like to thank my co-promotor Prof. Patrick Bultinck for the conceptual and computational support for this research. We experienced a fruitful collaboration and interesting exchange of ideas in numerous meetings. Furthermore, I acknowledge Prof. Michel Waroquier, who offered me the opportunity to work in a challenging and professional environment. As head of the Centre for Molecular Modeling, he managed to transform the CMM into a large and dynamical group of researchers. Without his support, I would not have had the chance to finish and present this work. Finally, I would like to address special thanks to Prof. Paul W. Ayers from the McMaster university in Canada, who provided enlightening answers during several extended visits to our university. His support in proving special properties must be acknowledged.

Contents

List of Tables

List of Figures

Nederlandstalige Samenvatting

Het concept van atomen en functionele groepen in molecules (AIM) is zeer nuttig gebleken om de chemische eigenschappen en de reactiviteit van molecules te verklaren. Dit concept volgt echter niet rechtstreeks uit de kwantumchemische theorie van moleculaire orbitalen (MO), die ontstaat uit toepassing van de wetten van de kwantummechanica op een systeem van elektronen en kernen. Daarom zijn diverse technieken ontwikkeld om a posteriori kwantumchemishe atomen te extraheren uit MO berekeningen. Mullikens populatie analyse, bijvoorbeeld, is gebaseerd op het toekennen van basisfuncties aan atoomkernen. In de meerderheid van de technieken wordt de elektrondichtheid verdeeld over atomaire domeinen met strakke grenzen (zoals in de Quantum Theory of Atoms In Molecules) of meer diffuse grenzen (zoals in de Hirshfeld techniek, de Iterated Stockholder Approach of de "fuzzy-atom"benadering van Mayer). Hoewel het AIM geen observabele is en er dus geen unieke definitie voor bestaat, hebben deze methoden hun succes reeds bewezen voor het bepalen van partiële ladingen op atomen, valenties op atomen, covalente bindingsordes en lokale descriptoren van de moleculaire reactiviteit. Zij werden ook aangewend om Vanderwaals-interacties en waterstofbinding te onderzoeken. Aangezien niet voor alle moleculaire eigenschappen een expliciete dichtheidsfunctionaal bestaat (bv. de kinetische energie), worden AIM waarden voor dergelijke eigenschappen in de praktijk bepaald door de moleculaire eigenschap over de atomaire domeinen te verdelen. Deze benadering is ambigu en problematisch, in het bijzonder voor overlappende atomen.

Deze thesis handelt over AIM technieken die werden ontwikkeld op basis van een verdeling van de volledige ééndeeltjes dichtheidsmatrix (1DM), en niet enkel de diagonaal hiervan (die de elektronendichtheid voorstelt). Voor alle eigenschappen die afhangen van de coordinaten van een enkel

elektron (zoals de kinetische energie) of voor eigenschappen die afhangen van de coordinaten van twee elektronen (zoals de repulsie-energie tussen elektronen) maar berekend worden binnen een gemiddeld-veld benadering, bestaat er een eenvoudige dichtheidsmatrixfunctionaal. Een verdeling van de 1DM over atomen leidt dan ook tot eenduidig bepaalde AIM-eigenschappen. In een eerste techniek wordt als randvoorwaarde opgelegd dat de verdeling van de 1DM consistent moet zijn met één van de courante methoden om de elektronendichtheid te verdelen over de atomen. Dit garandeert het lokale en positieve karakter van de AIM dichtheid, wat in vele gevallen nodig is om degelijke resultaten te bekomen. Deze niet-triviale randvoorwaarde is echter geen voldoende voorwaarde om de AIM dichtheidsmatrices te bepalen. We identificeerden verschillende AIM dichtheidsmatrixpartitioneringen die voldoen aan deze randvoorwaarde. Zij worden bekomen via verdeling van de 1DM over atomen en bindingen, en een herverdeling van de bindingsbijdragen over de atomen. Hierbij wordt gebruik gemaakt van atomaire gewichtsfuncties. Uit de keuze van het herverdelingsmechanisme en de randvoorwaarde, worden de atomaire gewichtsfuncties voor dit schema geconstrueerd. De geïntroduceerde techniek werd toegepast om atomaire en bindingsenergieën te bepalen voor het Hirshfeld-I AIM model. Het voordeel van deze werkwijze, in vergelijking met de meest courante methoden om atomaire energieen en bindingsenergieën te bepalen, is dat de energietermen volledig consistent zijn met de onderliggende verdeling van de elektronische structuur. Zonder dergelijke consistentie zijn de energietermen dubbelzinnig en mogelijk misleidend. Bovendien kan deze werkwijze efficiënt worden geïmplementeerd in de één-deeltjes Hilbert ruimte, waardoor de lastige numerieke integeraties die de meer courante methoden kenmerken, kunnen worden vermeden.

We stelden vast dat voor het Hirshfeld-I AIM model de verdeling van de dichtheidsmatrix over atomen en bindingen enkele tekortkomingen vertoont: (i) De atomaire gewichtsfuncties vertonen geen selectiviteit voor de valentieorbitalen die betrokken zijn in de binding. Bijdragen van core-elektronen en vrije electronenparen op een bepaald atoom kunnen aldus gedeeltelijk aan de bindingen en de andere atomen worden toegekend. (ii) Wanneer de gewichtsfuncties worden uitgedrukt in de één-electron Hilbert ruimte, dan vertonen zij geen optimale convergentie bij uitbreiding van de basisset. (iii) Een derde tekortkoming is dat de verdeling van de ruimte niet gebeurt aan de hand van projectieoperatoren op orthogonale deelruimtes. Die laatste eigenschap is nodig om de molecuulfragmenten van het schema te kunnen interpreteren als geëxciteerde atomen en elektronische verschuivingen veroorzaakt door het bindingsproces. Om tegemoet te komen aan al deze onvolkomenheden, werd een nieuwe AIM techniek ontwikkeld. Deze techniek maakt gebruik van atomaire gewichtsmatrices in de één-elektron

Hilbert ruimte, in plaats van gewichtsfuncties in de 3D ruimte. De gewichtsmatrices worden geconcipieerd als het aandeel van de individuele 1DMs van geïsoleerde atomen in hun totale som. Een recursief schema vormt deze gewichtsmatrices om tot orthogonale projectieoperatoren op één-elektron deelruimtes. De gewichtsmatrices worden daarna gebruikt om de dichtheidsmatrix te verdelen over atomen en bindingen. Aan de hand van de populaties van de atomaire fragmenten worden dichtheidsmatrices voor "geëxciteerde"geïsoleerde atomen samengesteld, waarna het hele proces wordt hernomen tot convergentie van de populaties van de atomaire fragmenten. Deze techniek voorkomt dat significante bijdragen van core-elektronen en vrije-elektronenparen in de bindingen terechtkomen. Voor diverse eigenschappen die verband houden met het bindingsproces, zoals de indices die de covalente bindingsorde uitdrukken, worden sterk verbeterde waarden bekomen.

AIM technieken die gebaseerd zijn op de volledige ééndeeltjesdichtheidsmatrix zijn aldus nuttig om de beschrijving van atomen in bindingen in de molecule te verbeteren zodat bekomen AIM waarden meer accuraat zijn. Ze vermijden ook de omslachtige numerieke integraties die nodig zijn in de conventionele AIM methoden. Daarnaast laten ze toe om waarden te berekenen voor AIM grootheden die verband houden met belangrijke concepten in de chemie, maar die niet kunnen worden bekomen met de conventionele methoden. Immers, niet alle eigenschappen van "molecuulfragmenten"kunnen worden afgeleid uit AIM dichtheden of uit een fragmentering van moleculaire eigenschappen. Een eenvoudig voorbeeld is de elektronegativiteit van het Interagerend Kwantum Atoom (IQA). Het IQA is een combinatie van een AIM fragment van de elektronische structuur met een enkele kern. De berekening van de elektronegativiteit van dit "geëxciteerd" en "vervormd" atoom vereist de kennis van IQA orbitaalenergieën en dus van de IQA dichtheidsmatrix (op het gemiddeld-veld niveau). We hebben aangetoond dat er een verband bestaat tussen de elektronegativiteit van de IQA en de inductieve effecten van atomen en functionele groepen in molecules. Dit illustreert dat belangrijke informatie in verband met de bindingsmechanismen in molecules, die niet kan worden bekomen met de conventionele AIM technieken, beschikbaar wordt via AIM technieken gebaseerd op de dichtheidsmatrices.

Summary

In traditional chemistry, the concept of atoms and functional groups in molecules has been applied successfully to explain the properties and reactivity of molecules. This concept does not naturally show up in the ab initio theoretical approaches, e.g. molecular orbital theory, in which the molecule is considered as a quantum mechanical system of electrons and nuclei. Many techniques have been developed to define atomic regions within these quantum mechanical objects and to extract atom-in-molecule (AIM) properties from ab initio data. Some techniques are based on the attachment of basis function to the atomic centers, e.g. Mulliken's population analysis. In the majority of these approaches, the molecular electron density is partitioned over atomic domains with either strict boundaries (as in Baders' Quantum Theory of Atoms In Molecules) or fuzzy boundaries (the Hirshfeld methods, the Iterated Stockholder Approach, Mayer's fuzzy atoms,...). Although the AIM is not an observable and cannot be defined in a unique way, these methods have been successfully applied to determine partial atomic charges, atomic valences, covalent bond orders and atom-condensed reactivity descriptors, and to investigate e.g. dispersion interactions and hydrogen bonding. Since not all molecular properties have an explicit expression in terms of the electron density, e.g. the kinetic energy, AIM values often rely on a partitioning of these properties themselves over the atomic domains. Especially for atoms with fuzzy boundaries, this approach is shown to be inherently ambiguous and problematic.

In this thesis, AIM techniques are developed that are based on the partitioning of the full one-electron density matrix (1DM) rather than its diagonal in $\boldsymbol{r}$-space (the electron density). All one-electron properties (e.g. the kinetic energy) and two-electron properties at the mean-field level (e.g. the exchange energy) have a simple expression in terms of the 1DM. A partitioning of the 1DM over the atoms therefore naturally determines well-defined AIM properties. In a first approach, we require consistency with established density-based techniques, as the successful application of most

AIM procedures relies on the local and positive character of the AIM densities. This nontrivial requirement is, however, not sufficient to determine the atom-condensed 1DMs. Several paths were identified that lead to single-atom 1DMs with a Hirshfeld-I density diagonal. These are obtained through a double-atom partitioning of the molecular 1DM over local atomic and bond contributions, and a subsequent distribution of the bond matrices over the atoms. From the choice of a distribution mechanism for the bond contributions (weighted-nonweighted) and from the constraint of consistency with the Hirshfeld-I densities, atomic weight functions are constructed for that double-atom partitioning scheme. For the "fuzzy" Hirshfeld-I AIM model, atomic and interaction energies as well as (chemical) bond energies were determined using the atom-condensed 1DMs. The advantage of this procedure, in comparison to the conventional methods, is that the energy terms are fully consistent with the underlying partitioning of the electronic structure. Without this consistency, the terms are ambiguous and are shown to be potentially misleading. In addition, this procedure can be implemented efficiently in one-electron Hilbert space to avoid the cumbersome numerical integrations that appear in the conventional methods.

We have observed that for the Hirshfeld-I AIM model, the partitioning of the 1DM inherits some shortcomings from the partitioning of the electron density: (i) The atomic weight functions are not selective for the valence orbitals involved in bonding, (ii) these weight functions don't have optimal convergence properties with respect to basis set size when they are expressed in one-electron Hilbert space, and (iii) the partitioning of space does not rely on projection onto orthogonal subspaces. The latter property is needed to interpret the double-atom fragments conveniently in terms of excited atoms and electronic deformations directed by the bonding process. To remedy these shortcomings, we developed a new AIM technique. This technique uses atomic weight matrices defined in one-electron Hilbert space, rather than weight functions in $\boldsymbol{r}$-space. The weight matrices are constructed by a stockholder approach from the 1DMs of isolated atoms, and a recursive scheme is used to transform them into orthogonal projection operators onto one-electron subspaces. These projectors are used to partition the molecular 1DM over atomic and bond matrices. From the populations of the atomic matrices, 1DMs are constructed for "excited" isolated atoms and these can be used to define new weight matrices. The process is repeated until self-consistency is reached for the atomic populations. The technique ensures that no significant contributions from core electrons and lone pairs appear in the bonds, and shows clear advantages for many properties associated with bonding, e.g. the covalent bond order indices.

AIM techniques based on the partitioning of the full one-electron density

matrix are therefore useful to improve the description of the atom in the molecule, to obtain more accurate AIM values for interesting properties and to avoid the cumbersome numerical integrations in alternative approaches. In addition, they allow the calculation of extra AIM quantities that are intimately related to the basic chemical concepts. Note that not all the properties of "molecular fragments" can be derived from the AIM densities or from a fragmentation of the molecular property. A simple example is the electronegativity of the Interacting Quantum Atom (IQA). The IQA is a combination of an AIM fragment of the electronic structure with a single nucleus. Practical calculation of the electronegativity of this "excited" and "deformed" atom requires orbital energies and thus the atom-condensed 1DM at the mean-field level. We show that the electronegativity of the IQA is related to the inductive effects of atoms and functional groups in molecules. This demonstrates that the 1DM-approach to the atom in the molecule provides important information about the bonding mechanisms in molecules that can not be obtained from one of the many previously investigated AIM methods.

Part I

The Atom-In-Molecule concept from a density-matrix perspective

1 Theoretical background: the atom in the molecule

1.1 The concept of atoms in molecules.

Chemistry is the science that studies the properties of matter and the changes it undergoes. Essential to the chemical approach is that it considers matter to be built up by "atoms" that are held together by "chemical bonds" and form "chemical compounds". The properties and behavior of a chemical are mainly determined by its functional groups: characteristic groups of bonded atoms within the compound bearing specific names, e.g. alkyl, alkenyl, alkynyl, benzyl, hydroxyl, aldehyde, carbonyl, carbonate, ester, ether, acyl halide, amine, imine, sulfhydryl, sulfide, etc ... The characteristics of functional groups are highly transferable in a sense that they are remarkably constant over different compounds [1, 2]. They are responsible for the "chemical reactivity", the behavior of the compound when it is exposed to other compounds. Chemical reactivity is conveniently explained in terms of electronegativity, nucleophilicity, acidity, inductive effects,...[3]. These typically chemical concepts have a long historic and phenomenological pedigree, often predating the 20'th century advent of quantum mechanics as the fundamental physical theory for atomic and molecular phenomena.

Quantum chemistry applies quantum theory to the explanation and prediction of chemical behavior. Essential to this more fundamental approach is that it considers matter as an interacting many-body system of nuclei and electrons [4]. Most quantum chemical studies assume that the nuclei are at rest (Born-Oppenheimer approximation). The physics of a quantum mechanical

system with N electrons is governed by the Schrödinger equation:

$$\hat{H}^{elec}\psi(\boldsymbol{x_1}, \boldsymbol{x_2}, ...\boldsymbol{x_N}) = E\psi(\boldsymbol{x_1}, \boldsymbol{x_2}, ...\boldsymbol{x_N}), \tag{1.1}$$

where the electronic "Hamiltonian" $\hat{H}^{elec}$ is an energy operator that acts on the "wavefunction" $\psi(\boldsymbol{x_1}, \boldsymbol{x_2}, ...\boldsymbol{x_N})$ of the electronic system and returns the expectation value E for the energy of the system. The electronic wavefunction is closely related to the probability distribution of the electrons and determines all chemical properties of a molecule. We use the notation $\boldsymbol{x} = \boldsymbol{r}\sigma$, where $\boldsymbol{r}$ specifies the coordinates of an electron in 3D space and σ represents the spin degree of freedom. The electronic Hamiltonian is defined (in atomic units) as

$$\hat{H}^{elec} = -\frac{1}{2}\sum_i \nabla_i^2 + \sum_{A,i} \frac{Z_A}{|\boldsymbol{r_i} - \boldsymbol{R_A}|} + \sum_{i<j} \frac{1}{|\boldsymbol{r_i} - \boldsymbol{r_j}|}, \tag{1.2}$$

where $-\frac{1}{2}\nabla_i^2$ is the kinetic energy operator for electron i, Z_A is the nuclear charge and $\boldsymbol{R_A}$ represents the coordinates of nucleus A. There is a discrete set of solutions to the Schrödinger equation, providing electronic energies and associated wavefunctions for the system. In principle, all "ground-state" properties of a chemical compound can be derived from the wavefunction associated with the lowest energy. Chemical reactions are simulated from modeling of the "transition-state" wavefunction [5]. A quantum chemical approach is also needed for an accurate simulation of the fingerprints of a compound in chemistry: the molecular infrared (IR) spectra [6], Raman spectra, ultraviolet-visible (UV-VIS) spectra, etc.

In contrast to traditional chemistry, the quantum mechanical approach does not rely on the concept of atoms and functional groups to determine the properties of matter and the changes it undergoes during chemical reactions. However, it is undoubtedly useful to define atoms and functional groups in the quantum mechanical framework. This provides the physical understanding of existing concepts of chemistry, such as the electronegativity of functional groups, inductive effects in a molecule, nucleophilic and acidic regions in molecules,.... It also allows to quantify these efficient chemical concepts, which is useful for the design and synthesis of new molecules and new materials with specific desirable properties. The quantum chemical atoms can be applied to improve force-field models [7] and make theoretical calculations faster and more efficient. The idea of treating atoms and functional groups as building blocks of molecules is called the AIM (atoms in molecules) concept [8].

1.2 Molecular properties

1.2.1 The expectation value of a molecular property

In quantum chemistry, a property of a molecule with N electrons in a state $\psi(\boldsymbol{x_1}, \boldsymbol{x_2}, ...\boldsymbol{x_N})$ is computed as the expectation value O of the corresponding operator $\hat{O}$ [9],

$$\begin{aligned} O &= \langle\psi|\hat{O}|\psi\rangle \\ &= \int d\boldsymbol{x_1} \int d\boldsymbol{x_2} \ldots \int d\boldsymbol{x_N} \psi^*(\boldsymbol{x_1}, \boldsymbol{x_2}, ...\boldsymbol{x_N}) \hat{O} \psi(\boldsymbol{x_1}, \boldsymbol{x_2}, ...\boldsymbol{x_N}). \end{aligned} \quad (1.3)$$

In Eq. (1.3), the subscripts label the individual electrons. For the most relevant properties, the operator depends on the coordinates and the spins of only few electrons. For a k-electron operator $\hat{O}^{(k)}$, the knowledge of a reduced density matrix of order k [10],

$$\begin{aligned} &\rho(\boldsymbol{x_1}, \boldsymbol{x_2}, ..., \boldsymbol{x_k}; \boldsymbol{x'_1}, \boldsymbol{x'_2}, ..., \boldsymbol{x'_k}) = \\ &\frac{N!}{k!(N-k)!} \int d\boldsymbol{x_{k+1}} \int d\boldsymbol{x_{k+2}} \ldots \int d\boldsymbol{x_N} \psi^*(\boldsymbol{x_1}, \boldsymbol{x_2}, ..., \boldsymbol{x_k}, \boldsymbol{x_{k+1}}, ..., \boldsymbol{x_N}) \\ &\qquad \times \psi(\boldsymbol{x'_1}, \boldsymbol{x'_2}, ..., \boldsymbol{x'_k}, \boldsymbol{x_{k+1}}, ..., \boldsymbol{x_N}) \ , \end{aligned} \quad (1.4)$$

is sufficient to determine the associated expectation value:

$$O^{(k)} = \int d\boldsymbol{x_1} \int d\boldsymbol{x_2} \ldots \int d\boldsymbol{x_k} \left(\hat{O}^{(k)} \rho(\boldsymbol{x_1}, \boldsymbol{x_2}, ..., \boldsymbol{x_k}; \boldsymbol{x'_1}, \boldsymbol{x'_2}, ..., \boldsymbol{x'_k}) \right) |_{\boldsymbol{x_1} = \boldsymbol{x'_1}, ...}. \quad (1.5)$$

The reduced density matrices of order 1 and 2 are henceforth called the one-electron density matrix (1DM) and the two-electron density matrix (2DM). The 2DM is sufficient to determine the ground state energy of a molecular electronic system since the associated operator, the electronic Hamiltonian in Eq. (1.2), is a two-electron operator.

1.2.2 Basic variables

Calculation of the wavefunction

The determination of the molecular wavefunction remains computationally challenging. It is convenient that the time-independent Schrödinger equation,

$$\hat{H}^{elec} \psi(\boldsymbol{x_1}, \boldsymbol{x_2}, ...\boldsymbol{x_N}) = E \psi(\boldsymbol{x_1}, \boldsymbol{x_2}, ...\boldsymbol{x_N}), \quad (1.6)$$

is solved in Hilbert space using a finite set of one-electron basis functions, rather than pointwise in $\boldsymbol{r}$-space. For a given basis set, the exact solution can

be expanded,

$$\psi(\boldsymbol{x_1}, \boldsymbol{x_2}, ...\boldsymbol{x_N}) = c_0\Phi_0(\boldsymbol{x_1}, \boldsymbol{x_2}, ...\boldsymbol{x_N}) + \sum_{ip} c_{ip}\Phi_i^p(\boldsymbol{x_1}, \boldsymbol{x_2}, ...\boldsymbol{x_N}) + \sum_{ijpq} c_{ijpq}\Phi_{ij}^{pq}(\boldsymbol{x_1}, \boldsymbol{x_2}, ...\boldsymbol{x_N}) + \ldots, \tag{1.7}$$

in terms of antisymmetrized Slater products $\Phi_0(\boldsymbol{x_1}, \boldsymbol{x_2}, ...\boldsymbol{x_N})$ of orthonormalized one-electron functions $\chi(\boldsymbol{x})$ [11–13]:

$$\Phi_0(\boldsymbol{x_1}, \boldsymbol{x_2}, ...\boldsymbol{x_N}) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(\mathbf{x}_1) & \chi_2(\mathbf{x}_1) & \cdots & \chi_N(\mathbf{x}_1) \\ \chi_1(\mathbf{x}_2) & \chi_2(\mathbf{x}_2) & \cdots & \chi_N(\mathbf{x}_2) \\ \vdots & \vdots & & \vdots \\ \chi_1(\mathbf{x}_N) & \chi_2(\mathbf{x}_N) & \cdots & \chi_N(\mathbf{x}_N) \end{vmatrix}. \tag{1.8}$$

The Φ_i^p are obtained from Φ_0 by replacing χ_i with χ_p, etc. The coefficients of the expansion are found by minimizing the expectation value of the energy under the constraint that the wavefunction is normalized.

$$\begin{aligned} L &= \langle\psi|\hat{H}|\psi\rangle - \lambda[\langle\psi|\psi\rangle - 1] \\ \frac{\delta L}{\delta c_0} &= \frac{\delta L}{\delta c_{ip}} = \frac{\delta L}{\delta c_{ijpq}} = \cdots = 0 \end{aligned} \tag{1.9}$$

Because the number of determinants required in the "full-CI" expansion grows factorially with the number of electrons and orbitals, the exact solution can only be obtained for very small molecules (with only a few electrons) in a small basis. Therefore, approximate wavefunctions are determined. Several methods can be used, each with its own scaling behavior in terms of the number of basis functions N_b [14]. In order of decreasing quality the most popular methods are Coupled-Cluster Theory [15–20] (CCSD, scales $\sim N_b^7$), Møller-Plesset Perturbation Theory [21–23] using a second order approximation (MP2, scales $\sim N_b^5$) or fourth order approximation (MP4, scales $\sim N_b^6$) and Hartree-Fock Theory [24–26] (HF, scales $\sim N_b^{3-4}$). In the bulk of the calculations in this thesis, HF theory is used. In Hartree-Fock theory, the wavefunction is expressed in terms of a single Slater determinant and a set of orthonormalized one-electron functions $\chi(\boldsymbol{x})$ is found that minimizes the expectation value of the energy.

Direct calculation of the 2DM

Using the strategy outlined above, the exact wavefunction is expanded (see Eq. (1.7)) in the huge N-electron Hilbert space. Approximate wavefunctions

are found by truncation of that Hilbert space, and the corresponding 1DM and 2DM are determined with Eq. (1.4). All one- and two- electron properties can be calculated using Eq. (1.5).

Instead of truncating the N-electron Hilbert space, an alternative strategy is to determine the 2DM (and 1DM) directly [27–29]. Then solutions must be found in the 2-electron Hilbert space, which scales only quadratically with system size. The 2DM that corresponds to the ground state wavefunction is the one that minimizes the expectation value of the electronic energy in the subspace of N-representable matrices.

$$E\left[\rho(\boldsymbol{x_1}, \boldsymbol{x_2}; \boldsymbol{x_1'}, \boldsymbol{x_2'})\right] = \int d\boldsymbol{x_1} \int d\boldsymbol{x_2}\ \hat{H}^{elec}\ \rho(\boldsymbol{x_1}, \boldsymbol{x_2}; \boldsymbol{x_1'}, \boldsymbol{x_2'})|_{\boldsymbol{x_1}=\boldsymbol{x_1'}, \boldsymbol{x_2}=\boldsymbol{x_2'}}. \quad (1.10)$$

A 2DM is N-representable if it can be derived from a physical wavefunction ensemble:

$$\rho(\boldsymbol{x_1}, \boldsymbol{x_2}; \boldsymbol{x_1'}, \boldsymbol{x_2'}) = \sum_i w_i \int d\boldsymbol{x_3} \ldots \int d\boldsymbol{x_N} \left(\psi_i^*(\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}, \ldots, \boldsymbol{x_N}) \times \psi_i(\boldsymbol{x_1'}, \boldsymbol{x_2'}, \boldsymbol{x_3}, \ldots, \boldsymbol{x_N})\right), \quad (1.11)$$

where $\sum_i w_i = 1$. The difficulty is to impose the sufficient N-representability conditions [30] and the method scales $\sim N_b^6$.

The electron density as basic variable

In Sec. 1.2.1 it was put forward to derive the properties of an electronic system from the wavefunction and the density matrices. An alternative strategy to determine these properties is provided by Density Functional Theory (DFT, [31]). In DFT, the electron density is the basic variable. According to the first Hohenberg-Kohn theorem [32], the ground-sate electron density,

$$\rho(\boldsymbol{r}) \quad = \quad \int d\sigma \int d\boldsymbol{x_2} \ldots \int d\boldsymbol{x_N} \psi^*(\boldsymbol{r}\sigma, \boldsymbol{x_2}, \ldots \boldsymbol{x_N}) \psi(\boldsymbol{r}\sigma, \boldsymbol{x_2}, \ldots \boldsymbol{x_N}), \quad (1.12)$$

which depends on only 3 spatial coordinates and is just the diagonal element of the spin-summed 1DM,

$$\rho(\boldsymbol{r}) = \int d\sigma \rho(\boldsymbol{x}, \boldsymbol{x}) = \rho(\boldsymbol{r}, \boldsymbol{r}), \quad (1.13)$$

uniquely determines the external potential and the electronic Hamiltonian of a many electron system. Therefore, it also determines the electronic states ψ and all properties of the system. Each property can be expressed as a universal

functional of the electron density. The second Hohenberg-Kohn theorem [32] defines an energy functional for the system,

$$E[\rho(\boldsymbol{r})], \tag{1.14}$$

and proves that the correct ground-state electron density minimizes this energy functional. The exact universal energy functional is unknown and the goal of Density Functional Theory is to design approximate functionals that connect the electron density with the energy. DFT offers the ability to reduce a many-electron problem depending on 4N coordinates, with a complexity that increases with the number of electrons, to only 3 coordinates. DFT methods scale similar to Hartree-Fock (N_b^{3-4}), but with a larger proportionality term.

1.3 Defining atomic regions

It is not trivial to introduce the concepts of atoms (and functional groups) into the quantum mechanical framework outlined in Sec. 1.2. The simplest strategy is based on the electron density. According to the Hohenberg-Kohn theorems (see Sec. 1.2.2), the ground-state electron density determines all properties of a molecule. A partitioning of the electron density over atomic "basins" with hard or fuzzy boundaries in principle defines the properties of the atoms in the molecule. This is by far the most popular AIM strategy. A more complicated approach is based on the density matrices. It was originally proposed by Li [33] in 1986. Although it clearly has its merits (as we will see in the following chapters), such approach remains largely unexplored. In this section we briefly summarize the most important techniques (for the current work) that have been developed to describe the elusive concept of atoms in the molecule.

1.3.1 Partitioning of space

Most techniques can be redefined in terms of a partitioning of space into atomic regions [34],

$$1 = \sum_A w_A(\boldsymbol{r}) \tag{1.15}$$

$$I = \sum_A (w_A)_{ij}, \tag{1.16}$$

where $w_A(\boldsymbol{r})$ is an "atomic weight function" in $\boldsymbol{r}$-space and $(w_A)_{ij}$ is a matrix element of the "atomic weight matrix" in one-electron Hilbert space. The indices refer to an a set of orthogonal basis functions $\phi_i(\boldsymbol{r})$. Such approach permits to treat the different decomposition schemes in a unified manner. It is possible to express the same atomic weight definition both in $\boldsymbol{r}$-space and one-electron Hilbert space,

$$w_A(\boldsymbol{r}) = \sum_{ij} (w_A)_{ij} \phi_i(\boldsymbol{r}) \phi_j(\boldsymbol{r}) \tag{1.17}$$

$$(w_A)_{ij} = \int d\boldsymbol{r} \phi_i(\boldsymbol{r}) w_A(\boldsymbol{r}) \phi_j(\boldsymbol{r}), \tag{1.18}$$

but a weight function defined in $\boldsymbol{r}$-space might be poorly approximated in one-electron Hilbert space when a small basis set is used (see Fig. 1.1). In the general case, the equivalence of Eqs. 1.17 and 1.18 is only guaranteed for a complete basis set.

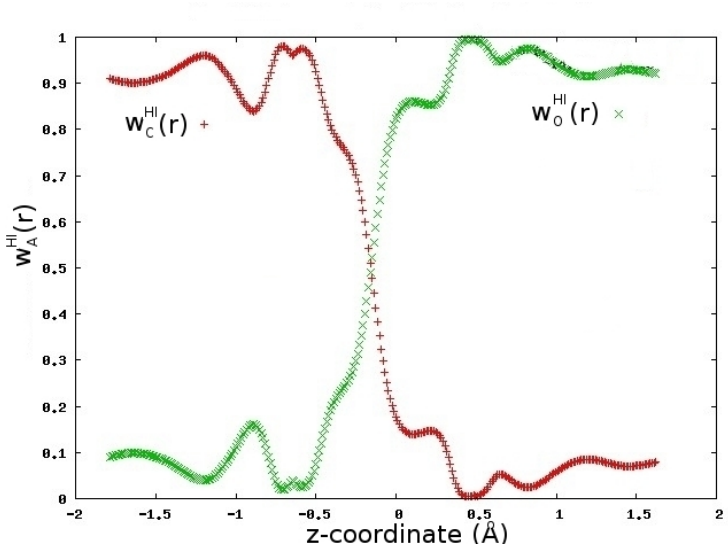


Figure 1.1: Values for the Hirshfeld-I atomic weight functions $w_A^{HI}(\boldsymbol{r})$ on the internuclear axis of a CO-molecule at the ROHF/Aug-cc-pVDZ level. The weight functions were originally defined in $\boldsymbol{r}$-space as smooth functions (resembling logistic functions), then expressed in the one-electron Hilbert space and finally transformed back to $\boldsymbol{r}$-space.

1.3.2 Overview of AIM techniques

Mulliken population analysis and the minimal deformation partitioning

In one group of techniques, one uses the attachment of non-orthogonal basis functions $\phi_i(\boldsymbol{r})$ to atomic centers A. The best known method of this type is the Mulliken population analysis [35]. Suppose that the electron density is given in terms of real basis functions $\phi_i(\boldsymbol{r})$ centered on the atoms:

$$\rho(\boldsymbol{r}) = \sum_{ij} \rho_{ij} \phi_i(\boldsymbol{r}) \phi_j(\boldsymbol{r}). \tag{1.19}$$

The two-center contributions are partitioned half-half between the centers

$$(\rho_A^{Mull})_{ij} = \frac{1}{2} \Big(\delta_{iA} + \delta_{jA} \Big) \rho_{ij}. \tag{1.20}$$

In Eq. (1.20) δ_{iA} is 1 or 0 if $\phi_i(\boldsymbol{r})$ is respectively centered or not centered on atom A. Mulliken weight functions can be constructed as follows:

$$w_A^{Mull}(\boldsymbol{r}) = \frac{\rho_A^{Mull}(\boldsymbol{r})}{\rho(\boldsymbol{r})} = \frac{\sum_{ij} (\rho_A^{Mull})_{ij} \phi_i(\boldsymbol{r}) \phi_j(\boldsymbol{r})}{\sum_{ij} (\rho)_{ij} \phi_i(\boldsymbol{r}) \phi_j(\boldsymbol{r})}. \tag{1.21}$$

In alternative approaches, e.g. the minimal deformation partitioning [36, 37], the two-center contributions are assigned to the center that corresponds with the least diffuse basis function. These atomic weight functions are obviously basis set dependent and do not converge in the basis set limit.

Bader's Quantum Theory of Atoms In Molecules (QTAIM)

In the second group of methods, one uses the molecular electron density and its properties as the guide for obtaining the atom in the molecule. Since the electron density depends on 3 spatial coordinates, it is a natural choice to use a three-dimensional partitioning of space with sharp or fuzzy boundaries between different atoms in the molecule. In Bader's Quantum Theory of Atoms In Molecules (QTAIM) [8, 38, 39], a "zero-flux" boundary condition for the electron density is satisfied at every point $\boldsymbol{r}_S$ of an interatomic surface $S(\boldsymbol{r}_S)$,

$$\nabla \rho(\boldsymbol{r}_S) \cdot n(\boldsymbol{r}_S) \quad = \quad 0, \tag{1.22}$$

where $n(\boldsymbol{r}_S)$ is the unit vector normal to the surface at $\boldsymbol{r}_S$. The interatomic surface delimits the atomic domains. This implies that $\boldsymbol{r}$-space is partitioned in a binary way, since each point in space corresponds completely to a single atomic domain Ω_A:

$$w_A^{QTAIM}(\boldsymbol{r}) \quad = \quad \delta(\boldsymbol{r} \in \Omega_A). \tag{1.23}$$

The standard Hirshfeld method

In the standard Hirshfeld method [40], the atomic regions are allowed to have fuzzy boundaries. A model electron density for the molecule $\rho^0(\boldsymbol{r})$ is constructed from the electron densities $\rho_A^0(\boldsymbol{r})$ of neutral isolated atoms placed at the positions of the nuclei in the molecule,

$$\rho^0(\boldsymbol{r}) = \sum_A \rho_A^0(\boldsymbol{r}), \tag{1.24}$$

and the overlapping atomic regions are defined by the contribution of each isolated atom to the model electron density

$$w_A^H(\boldsymbol{r}) = \frac{\rho_A^0(\boldsymbol{r})}{\rho^0(\boldsymbol{r})}. \tag{1.25}$$

The Hirshfeld atoms,

$$\rho_A(\boldsymbol{r}) = w_A^H(\boldsymbol{r})\rho(\boldsymbol{r}), \tag{1.26}$$

minimize the Kullback-Leibler functional [41–43], expressing the entropy deficiency in information theory,

$$I = \sum_A \int d\boldsymbol{r}\rho_A(\boldsymbol{r}) ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right) \tag{1.27}$$

and are thus the ones best preserving the information contained in the $\rho_A^0(\boldsymbol{r})$.

The iterative extension to the Hirshfeld method

For ionic species, it is counterintuitive to use neutral isolated atoms in the Hirshfeld model. Different results are obtained when the model is constructed from isolated ions. To eliminate this source of arbitrariness in an iterative extension of the Hirshfeld method [44–47], it is imposed that the densities of the isolated atoms $\rho_A^0(\boldsymbol{r})$ have the same (fractional) number of electrons N_A as the resulting AIM densities $\rho_A(\boldsymbol{r})$. The iterative process is started from a regular Hirshfeld calculation. At the end of iteration i, an AIM electron density is obtained and the electron number

$$N_A^{(i)} = \int d\boldsymbol{r} w_A^{(i)}(\boldsymbol{r})\rho(\boldsymbol{r}), \tag{1.28}$$

is used to update the densities of the isolated atoms $\rho_A^0(\boldsymbol{r}, N_A^{(i)})$. The densities of isolated atoms with a fractional number of electrons are, e.g., linear interpolations between the densities of isolated atoms with an integer number of

electrons [48, 49]:

$$N_A^{(i)} = I_A + q \tag{1.29}$$

$$\rho_A^0(\boldsymbol{r}, N_A^{(i)}) = (1-q)\ \rho_A^0(\boldsymbol{r}, I_A) + q\ \rho_A^0(\boldsymbol{r}, I_A+1). \tag{1.30}$$

In Eq. (1.30), I_A and q are the integer and fractional part of $N_A^{(i)}$. New atomic weight functions are derived,

$$w_A^{(i+1)}(\boldsymbol{r}) = \frac{\rho_A^0(\boldsymbol{r}, N_A^{(i)})}{\sum_B \rho_B^0(\boldsymbol{r}, N_B^{(i)})}, \tag{1.31}$$

and the final Hirshfeld-I weights are obtained when the cycle converges :

$$N_A^{(i)} - N_A^{(i-1)} \leq \varepsilon \tag{1.32}$$

$$w_A^{HI}(\boldsymbol{r}) \equiv w_A^{(i)}(\boldsymbol{r}). \tag{1.33}$$

The Hirshfeld-Iλ method

In the Hirshfeld-Iλ approach [50] the Kullback-Leibler functional (see Eq. (1.27)), expressing the entropy deficiency in the information theory, is minimized with respect to the constraints:

$$\rho(\boldsymbol{r}) = \sum_A \rho_A(\boldsymbol{r}), \tag{1.34}$$

$$N_A^0 = \int d\boldsymbol{r}\rho_A^0(\boldsymbol{r}) = \int d\boldsymbol{r}\rho_A(\boldsymbol{r}) = N_A. \tag{1.35}$$

Both constraints are placed in one minimization, whereas the second constraint is imposed by an outer iterative cycle in the Hirshfeld-I model. The two algorithms do not lead to the same end result since the weight function of the Hirshfeld-Iλ method has an extra factor of $e^{-\lambda_A}$ (both in the nominator and the denominator):

$$w_A^{HI\lambda}(\boldsymbol{r}) = \frac{\rho_A^0(\boldsymbol{r}, N_A^0)e^{-\lambda_A}}{\sum_B \rho_B^0(\boldsymbol{r}, N_B^0)e^{-\lambda_B}}, \tag{1.36}$$

where

$$\lambda_A = -\int d\boldsymbol{r}\rho_A(\boldsymbol{r})\frac{\delta(\rho_A^0(\boldsymbol{r}, N_A^0))}{\delta N_A}\frac{1}{\rho_A^0(\boldsymbol{r}, N_A^0)}. \tag{1.37}$$

Since $w_A^{HI\lambda}$ depends on λ_A, which in turn depends on $\rho_A(\boldsymbol{r}) = w_A^{HI\lambda}\rho(\boldsymbol{r})$, Eqs. (1.36) and (1.37) are interdependent and have to be solved iteratively. While the method is better founded in information theory than Hirshfeld-I (it is e.g. strictly variational), the results were rather disappointing.

The Iterated Stockholder Atoms (ISA) approach

The Iterated Stockholder Atoms (ISA) approach [47, 51] resembles the iterative-Hirshfeld method, but the model electron density is constructed differently. Suppose that we can obtain a provisional AIM electron density,

$$\rho_A^{(i)}(\boldsymbol{r}) = w_A^{(i)}(\boldsymbol{r})\rho(\boldsymbol{r}), \tag{1.38}$$

during iteration i. Every point in space $\boldsymbol{r}$ lies on the surface of a sphere of radius $|\boldsymbol{r} - \boldsymbol{R}_A|$ from a given nucleus A. The average density over the surface of that sphere is denoted $\langle\ \rho_A^{(i)}(\boldsymbol{r} - \boldsymbol{R}_A)\ \rangle$ and can be used to define new atomic weight functions:

$$w_A^{(i+1)}(\boldsymbol{r}) \quad = \quad \frac{\langle\ \rho_A^{(i)}(\boldsymbol{r} - \boldsymbol{R}_A)\ \rangle}{\sum_B \langle\ \rho_B^{(i)}(\boldsymbol{r} - \boldsymbol{R}_B)\ \rangle} \tag{1.39}$$

An updated AIM electron density can be constructed with Eq. (1.38) and the process can be repeated until convergence:

$$w_A^{ISA}(\boldsymbol{r}) \quad = \quad w_A^{(\infty)}(\boldsymbol{r}). \tag{1.40}$$

1.4 Atom-condensed properties

1.4.1 Fragments of molecular properties

In accordance with the first Hohenberg-Kohn theorem (see Sec. 1.2.2), the ground-state electron density in principle determines all properties. The second Hohenberg-Kohn theorem provides a strategy to calculate the ground-state density and its energy components. A partitioning of the molecular ground-state density over the atoms defines AIM densities. The properties of the atoms in the molecule are in general not calculated as functionals of AIM densities. Even though there are explicit formulas for some properties, these formulas are often computationally intractable. Moreover, the atoms in the molecule are not in the ground state with respect to e.g. the Hamiltonian of the isolated atoms. Therefore, the common strategy to determine the properties of the AIM electron densities $\rho_A(\boldsymbol{r})$, e.g. the kinetic energy, consists of fragmenting the molecular expectation value expressions using the same atomic weight functions $w_A(\mathbf{r})$ that are used to partition the molecular density [34].

One-electron operators

The expectation value of a one-electron operator $\hat{F}$ is condensed to the atoms using:

$$F_A = \int d\boldsymbol{x}\; w_A(\boldsymbol{r})\hat{F}\rho(\boldsymbol{x},\boldsymbol{x}')\big|_{\boldsymbol{x}=\boldsymbol{x}'}\,. \tag{1.41}$$

where the notation $|_{x=x'}$ indicates that $\boldsymbol{x'}$ is replaced by $\boldsymbol{x}$ after the action of $\hat{F}$ on $\rho(\boldsymbol{x},\boldsymbol{x}')$ but before the integration is carried out. Within the Quantum Theory of Atoms in Molecules (QTAIM) [8], e.g., atomic energies are obtained by partitioning the expectation value of the molecular kinetic energy over the atomic domains (as in Eq. (1.41)) and using the virial theorem [52] to construct the corresponding atomic potential energy.

It is important to note that Eq. (1.41) may be ambiguous if the operator $\hat{F}$ has multiple representations [53, 54]. The kinetic energy operator, e.g., has two distinct representations. Whereas the molecular kinetic energy t is independent of the representation of the operator,

$$t = -\frac{1}{2}\int d\boldsymbol{r}\; \nabla^2\rho(\boldsymbol{r},\boldsymbol{r'})\Big|_{\boldsymbol{r}=\boldsymbol{r'}} = \frac{1}{2}\int d\boldsymbol{r}\; \nabla\cdot\nabla'\rho(\boldsymbol{r},\boldsymbol{r'})\big|_{r=r'}\,, \tag{1.42}$$

the atom-condensed value

$$t_A^{(1)} = -\frac{1}{2}\int d\boldsymbol{r}\; w_A(\boldsymbol{r})\nabla^2\rho(\boldsymbol{r},\boldsymbol{r'})\Big|_{\boldsymbol{r}=\boldsymbol{r'}} \tag{1.43}$$

yields a result that differs in general from

$$t_A^{(2)} = \frac{1}{2} \int d\boldsymbol{r}\; w_A(\boldsymbol{r}) \nabla \cdot \nabla' \rho(\boldsymbol{r}, \boldsymbol{r'})\big|_{\boldsymbol{r}=\boldsymbol{r'}} . \tag{1.44}$$

Only in special cases, e.g. when QTAIM weight functions are used, do Eq. (1.43) and Eq. (1.44) coincide [55]. For the Hirshfeld-I atom-in-molecule model, which will be used extensively throughout this work, Eq. (1.41) is unavoidably ambiguous.

A second remark concerning nonlocal operators (such as the kinetic energy operator), is the fact that the exact place where the partitioning of unity ($1 = \sum_A w_A(\boldsymbol{r})$) is introduced in the expectation value expression can have an important influence on the resulting AIM condensed values [56]. For a nonlocal operator $\hat{F}$,

$$F_A^{(2)} = \int d\boldsymbol{x}\; \hat{F} \left(w_A(\boldsymbol{r}) \rho(\boldsymbol{x}, \boldsymbol{x'}) \right)\Big|_{\boldsymbol{x}=\boldsymbol{x'}} \tag{1.45}$$

yields a result that may differ from Eq. (1.41). However, in Eq. (1.41) the weight function restricts the integration effectively to a local area while for Eq. (1.45) this is clearly not the case. For this reason, several authorities in the field propose to introduce the unity and its partitioning into weight functions before any operator [34].

Two-electron operators

When an operator $\hat{V}$ strictly depends on the coordinates of two electrons, its expectation value is conveniently condensed to the atom pairs using [34]:

$$V_{AB} = \int\int d\boldsymbol{x}_1 d\boldsymbol{x}_2\; w_A(\boldsymbol{r}_1) w_B(\boldsymbol{r}_2) \hat{V} \rho(\boldsymbol{x}_1, \boldsymbol{x}_2; \boldsymbol{x}_1', \boldsymbol{x}_2')\big|_{\boldsymbol{x_1}=\boldsymbol{x}_1', \boldsymbol{x_2}=\boldsymbol{x}_2'} . \tag{1.46}$$

For the two-electron energy V_{ee} within the Hartree-Fock approximation, we obtain:

$$(V_{ee})_{AB}^{(HF)} = \frac{1}{2} \int\int d\boldsymbol{x}_1 d\boldsymbol{x}_2 w_A(\boldsymbol{r}_1) w_B(\boldsymbol{r}_2) \frac{\left[\rho(\boldsymbol{x}_1, \boldsymbol{x}_1)\rho(\boldsymbol{x}_2, \boldsymbol{x}_2) - (\rho(\boldsymbol{x}_1, \boldsymbol{x}_2))^2\right]}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|}. \tag{1.47}$$

Note that the fragmentation proposed in Eqs. (1.46) and (1.47) is not unique and depends on the exact partitioning of unity in two-atom terms,

$$\begin{aligned} 1 &= \sum_{AB} w_A(\boldsymbol{r}_1) w_B(\boldsymbol{r}_2) \\ &= \frac{1}{2} \sum_{AB} \Big(w_A(\boldsymbol{r}_1) w_B(\boldsymbol{r}_1) + w_A(\boldsymbol{r}_2) w_B(\boldsymbol{r}_2) \Big) \\ &= \frac{1}{4} \sum_{AB} \Big(w_A(\boldsymbol{r}_1) w_B(\boldsymbol{r}_1) + w_A(\boldsymbol{r}_1) w_B(\boldsymbol{r}_2) + w_A(\boldsymbol{r}_2) w_B(\boldsymbol{r}_1) + w_A(\boldsymbol{r}_2) w_B(\boldsymbol{r}_2) \Big) \\ &\vdots \quad . \end{aligned}$$

The formulation in Eqs. (1.46) and (1.47) has the advantage that the weight functions restrict the integration effectively to a local two-atom area. It has been used in several studies to partition the molecular two-electron energy into two-atom terms [57–60].

1.4.2 Properties of molecular one-electron fragments

Depending on the operator, the partitioning of the molecular expectation value expressions can be ambiguous (see Sec. 1.4.1). Based on the criterion that the integration should be restricted to a local area, the number of possible fragmentation schemes is narrowed down. However, there is no guarantee that the atom-condensed values $\int d\boldsymbol{x}\; w_A(\boldsymbol{r}) \hat{F} \rho(\boldsymbol{x}, \boldsymbol{x}') \big|_{\boldsymbol{x}=\boldsymbol{x}'}$ are in principle consistent with the evaluation of a universal density functional $F[\rho_A(\boldsymbol{r})]$ over the AIM domains.

The correspondence between the molecular fragment and its properties is less problematic when density-matrix fragments are considered [33] (instead of density fragments). The expectation values of a one- or two- electron operator are expressed directly in terms of respectively a 1DM and a 2DM (Eq. (1.5)). Therefore, a partitioning of the density matrices leads naturally to the partitioning of the expectation values:

$$F_A = \int d\boldsymbol{x}\; \hat{F} \rho_A(\boldsymbol{x}, \boldsymbol{x}') \big|_{\boldsymbol{x}=\boldsymbol{x}'} . \tag{1.48}$$

$$V_{AB} = \int\int d\boldsymbol{x}_1 d\boldsymbol{x}_2\; \hat{V} \rho_{AB}(\boldsymbol{x}_1, \boldsymbol{x}_2; \boldsymbol{x}_1', \boldsymbol{x}_2') \big|_{\boldsymbol{x}_1=\boldsymbol{x}_1', \boldsymbol{x}_2=\boldsymbol{x}_2'} . \tag{1.49}$$

Since at the Hartree-Fock level of theory the 2DM can be expressed in terms of the 1DM,

$$\rho(\boldsymbol{x}_1, \boldsymbol{x}_2; \boldsymbol{x}_1', \boldsymbol{x}_2') = \rho(\boldsymbol{x}_1, \boldsymbol{x}_1') \rho(\boldsymbol{x}_2, \boldsymbol{x}_2') - \rho(\boldsymbol{x}_1, \boldsymbol{x}_2') \rho(\boldsymbol{x}_2, \boldsymbol{x}_1') \tag{1.50}$$

a partitioning of the Hartree-Fock 1DM defines atom-condensed values for all one- and two- electron properties. In chapters 2 and 3, we present two novel and promising methods to partition the molecular 1DM over the atoms (and bonds) in the molecule. We first briefly review the most important applications of AIM techniques in the following section.

1.5 Applications of AIM techniques

Chemists consider molecules as atoms held together by chemical bonds. Physicists and quantum chemists treat them as a many-body system of electrons and nuclei. AIM techniques are used to translate the results of a quantum chemical calculation into typical chemical concepts.

1.5.1 Partial atomic charges

To have a crude indication of the reactive sites in a molecule, a chemist considers partial charges on the atoms. Partial charges reflect the observation that electrons are partially drawn away when an electrically neutral atom bonds chemically to another neutral atom that is more electronegative. This leaves the region about that atom's nucleus with a partial positive charge, and it creates a partial negative charge on the atom to which it is bonded. A quantitative treatment of these partial charges has been central to numerous scientific investigations in the last decades [61, 62]. Atomic charges have been proposed that fit experimental data (e.g. reaction rate constants, spectroscopic data, dielectric constants,...), that fit a theoretically determined property (e.g. the molecular electrostatic potential [63, 64]) or fulfil a theoretical principle (e.g. equalization of the electronegativity in the molecule [65–71]).

In a more rigorous approach, partial charges are obtained from partitioning the molecular wavefunction (or a quantity that is derived from it) using one of the atom-in-molecule models described in Sec. 1.3.2. Once the atomic density $\rho_A(\boldsymbol{r})$ is derived, the charge q_A of an atom with nuclear charge Z_A is determined using:

$$q_A = Z_A - \int d\boldsymbol{r}\rho_A(\boldsymbol{r}). \tag{1.51}$$

Since there is no unique way to delineate the atomic regions, partial charges from different schemes differ considerably. Bader charges are large in comparison to charges from most other schemes, while the standard Hirshfeld scheme provides charges that are to too small for ionic compounds [44].

Partial atomic charges are particularly useful in force-field models in molecular mechanics [72]. The energy term that describes the electrostatic interaction energy in atom pairs is computed using Coulomb's law,

$$U_{AB}^{Coul} = \frac{q_A q_B}{\boldsymbol{R}_{AB}} \tag{1.52}$$

The electrostatic potential $V(\boldsymbol{r})$ generated by the nuclear charges and the AIM

charge distribution $\rho_A(\boldsymbol{r})$,

$$V(\boldsymbol{r}) = \frac{Z_A}{|R_A - \boldsymbol{r}|} - \int d\boldsymbol{r'} \frac{\rho_A(\boldsymbol{r'})}{|\boldsymbol{r'} - \boldsymbol{r}|}, \tag{1.53}$$

can be conveniently expressed by a traditional multipole expansion [73],

$$V(\boldsymbol{r}) = \frac{q_A}{\boldsymbol{r}} + \frac{1}{\boldsymbol{r}^3} (\boldsymbol{\mu}_A \cdot \boldsymbol{r}) + \cdots, \tag{1.54}$$

in terms of the atomic charges q_A, the dipole moments $\boldsymbol{\mu}_A$, the quadrupole moments $\boldsymbol{Q}_A$ and higher multipole moments. For the application in Eq. (1.52), it is important that this multipole expansion is dominated by the atomic charges. However, it is observed that the dipoles on the atoms from an AIM model, e.g. the Hirshfeld-I model, are often non-negligible. This implies that the molecular electrostatic potential is not optimally fit by only the charges on the atoms. Nevertheless, Hirshfeld-I charges are a valuable alternative for the charges derived from the electrostatic potential (ESP) in force field models [74]. One could argue that charges fitted to reproduce the ESP easily outperform the Hirshfeld-I charges in this comparison. This is correct but ESP-fitted partial charges are statistically ill defined [75] and create a noise on the training data used for the parameterization of the force field.

1.5.2 Bond orders and atomic valences

In Lewis structures, chemical bonds are identified with electron pairs shared between bonded atoms. The bond order is the multiplicity of a chemical bond (single, double or triple). The valence of an atom is the maximal number of electrons that the atom may provide for bonding (the number of bonds plus the number of unpaired electrons). The free valence of an atom is the number of electrons that is available for additional bonding (the number of unpaired electrons). In classical chemistry, these quantities have integer values. AIM techniques are used to separate the bond order in covalent and ionic contributions and to construct quantum chemical indicators for covalent bonds orders, atomic valences and free valences. For homonuclear diatomic species at the Hartree-Fock level, values for the quantum chemical indicators are real numbers that are close the integer classical values. The main advantage of quantum chemical indicators for bonds orders is that these can be computed also between e.g. atoms connected by a hydrogen bridge, where the classical expression can no longer be used [76].

The shared electron distribution index (SEDI) [77–86] is a commonly used quantum chemical bond order index. SEDI are obtained from integration

of the exchange density at the Hartree-Fock level or exchange-correlation density at correlated levels of theory. The exchange-correlation density expresses the correlation between the positions of two electrons in the molecule. It reads

$$\rho^{xc}\left(\boldsymbol{x},\boldsymbol{x}'\right)=\rho\left(\boldsymbol{x}\right)\rho\left(\boldsymbol{x}'\right)-\rho\left(\boldsymbol{x},\boldsymbol{x}';\boldsymbol{x},\boldsymbol{x}'\right) \tag{1.55}$$

where the first term corresponds to the independent movement of electrons in the molecule and second term represents the fully correlated movement. The latter term is identified with the diagonal elements of the 2DM. At the single-determinant level, this simplifies to the Hartree-Fock exchange density:

$$\rho^{xc}\left(\boldsymbol{x},\boldsymbol{x}'\right)=\rho\left(\boldsymbol{x},\boldsymbol{x}'\right)\rho\left(\boldsymbol{x}',\boldsymbol{x}\right). \tag{1.56}$$

In terms of separate 1DMs for α and β spin,

$$\begin{aligned}\rho_{\alpha}\left(\boldsymbol{r},\boldsymbol{r}'\right) &= \rho\left(\boldsymbol{r}\alpha,\boldsymbol{r}'\alpha\right) &(1.57)\\ \rho_{\beta}\left(\boldsymbol{r},\boldsymbol{r}'\right) &= \rho\left(\boldsymbol{r}\beta,\boldsymbol{r}'\beta\right) &(1.58)\\ \rho_{\alpha-\beta}\left(\boldsymbol{r},\boldsymbol{r}'\right) &= \rho_{\alpha}\left(\boldsymbol{r},\boldsymbol{r}'\right)-\rho_{\beta}\left(\boldsymbol{r},\boldsymbol{r}'\right) &(1.59)\end{aligned}$$

corresponding expressions can be defined as:

$$\begin{aligned}\rho_{\alpha}^{xc}\left(\boldsymbol{r},\boldsymbol{r}'\right) &= \rho_{\alpha}\left(\boldsymbol{r},\boldsymbol{r}'\right)\rho_{\alpha}\left(\boldsymbol{r}',\boldsymbol{r}\right) &(1.60)\\ \rho_{\beta}^{xc}\left(\boldsymbol{r},\boldsymbol{r}'\right) &= \rho_{\beta}\left(\boldsymbol{r},\boldsymbol{r}'\right)\rho_{\beta}\left(\boldsymbol{r}',\boldsymbol{r}\right) &(1.61)\\ \rho_{\alpha-\beta}^{xc}\left(\boldsymbol{r},\boldsymbol{r}'\right) &= \rho_{\alpha-\beta}\left(\boldsymbol{r},\boldsymbol{r}'\right)\rho_{\alpha-\beta}\left(\boldsymbol{r}',\boldsymbol{r}\right). &(1.62)\end{aligned}$$

Integrating $\boldsymbol{r}$ and $\boldsymbol{r}'$ over the atomic domains of atoms A and B and multiplying by two to account for the symmetrical integration over B and A, eventually results in the bond order index B_{AB}:

$$B_{AB}=2\int d\boldsymbol{r}\int d\boldsymbol{r}'w_{A}(\boldsymbol{r})w_{B}(\boldsymbol{r}')\left(\rho_{\alpha}^{xc}\left(\boldsymbol{r},\boldsymbol{r}'\right)+\rho_{\beta}^{xc}\left(\boldsymbol{r},\boldsymbol{r}'\right)\right). \tag{1.63}$$

The atomic valence V_A is simply,

$$V_{A}=\sum_{A<B}B_{AB}+F_{A}, \tag{1.64}$$

where F_A is the free valence defined by:

$$F_{A}=\int d\boldsymbol{r}\int d\boldsymbol{r}'w_{A}(\boldsymbol{r})w_{A}(\boldsymbol{r}')\rho_{\alpha-\beta}^{xc}\left(\boldsymbol{r},\boldsymbol{r}'\right). \tag{1.65}$$

1.5.3 Atom-condensed reactivity descriptors

Chemical reactivity is associated with the presence of specific atoms and functional groups in the molecule and is efficiently interpreted in terms of the electrophilicity, nucleophilicity, polarizability, electronegativity, acidity, basicity, hardness, softness, etc. of the functional groups. In quantum chemistry, reactivity descriptors [3] have been developed to express the information contained in the wavefunction in terms of these typical chemical quantities. These descriptors represent the response of a system to a specific perturbation. Within the DFT framework, the descriptors are derived from a Taylor series expansion of the electronic energy $E[N, v(\boldsymbol{r})]$ in terms of the total number of electrons N and the external potential $v(\boldsymbol{r})$ [87]:

$$\begin{aligned} E[N, v(\boldsymbol{r})] &= E_0 + \Delta N \left(\frac{\partial E}{\partial N}\right)_{v(\boldsymbol{r})} + \int \left(\frac{\delta E}{\delta v(\boldsymbol{r})}\right)_N \Delta v(\boldsymbol{r}) d\boldsymbol{r} \\ &\quad + \frac{1}{2}(\Delta N)^2 \left(\frac{\partial^2 E}{\partial N^2}\right)_{v(\boldsymbol{r})} + \Delta N \int \left(\frac{\delta^2 E}{\delta v(\boldsymbol{r}) \delta N}\right) \Delta v(\boldsymbol{r}) d\boldsymbol{r} \\ &\quad + \frac{1}{2} \int \left(\frac{\delta^2 E}{\delta v(\boldsymbol{r}) \delta v(\boldsymbol{r'})}\right)_N \Delta v(\boldsymbol{r}) \Delta v(\boldsymbol{r'}) d\boldsymbol{r} d\boldsymbol{r'} + \cdots \end{aligned} \tag{1.66}$$

The reactivity indicators are the derivatives in Eq. (1.66). Global indicators provide a single value and describe the global reactivity of the molecule. The most important indicators are the chemical potential μ [88, 89] and the global hardness η [90],

$$\mu = \left(\frac{\partial E}{\partial N}\right)_{v(\boldsymbol{r})} \qquad \eta = \frac{1}{2}\left(\frac{\partial^2 E}{\partial N^2}\right)_{v(\boldsymbol{r})} \tag{1.67}$$

that express respectively the electronegativity of the molecule and the resistance towards charge transfer. From these quantities, other global descriptors can be derived: the global softness S [91], the electrophilicity ω [92–94] and the nucleophilicity ν [94, 95],

$$S = \frac{1}{2\eta} \qquad \omega = \frac{\mu^2}{2\eta} \qquad \nu = \frac{1}{\omega}. \tag{1.68}$$

The global softness is associated with polarizability on a molecular level. The electrophilicity is the change in energy that occurs when the molecule is brought in contact with a perfect nucleophile (a reservoir of electrons that are completely polarizable and have a zero chemical potential). In practice, the global indicators are not calculated with Eq. (1.67), but approximated using

finite differences as [96]:

$$\mu = \frac{-(IP + EA)}{2} \qquad \eta = \frac{(IP - EA)}{2}, \tag{1.69}$$

where IP and EA represent the ionization potential and the electron affinity of the molecule. The global reactivity indicators do not provide information about the reactive sites in the molecule.

Information about the reactive sites in the molecule is contained in the local indicators. Local indicators are functions of the position $\boldsymbol{r}$. The most important local indicators are the electron density $\rho(\boldsymbol{r})$ and the Fukui function $f(\boldsymbol{r})$ [97–99],

$$\rho(\boldsymbol{r}) = \left(\frac{\delta E}{\delta v(\boldsymbol{r})}\right)_N \qquad f(\boldsymbol{r}) = \left(\frac{\delta^2 E}{\delta v(\boldsymbol{r})\delta N}\right) = \left(\frac{\partial \rho(\boldsymbol{r})}{\partial N}\right)_{v(\boldsymbol{r})}. \tag{1.70}$$

The Fukui function is an intramolecular descriptor that is normalized to 1 and indicates the preferential site for electron addition or removal. Other local reactivity indicators are derived from it, e.g. the local softness $s(\boldsymbol{r})$ and the local hardness $\eta(\boldsymbol{r})$ [100],

$$s(\boldsymbol{r}) = Sf(\boldsymbol{r}) \qquad \eta(\boldsymbol{r}) = \frac{1}{2s(\boldsymbol{r})} \tag{1.71}$$

In practice, atom-condensed values are computed for these three-dimensional reactivity descriptors [101]:

$$f_A^+ = \int d\boldsymbol{r}\, w_A(\boldsymbol{r}) \left(\rho^{N+1}(\boldsymbol{r}) - \rho^N(\boldsymbol{r})\right) = N_A^{N+1} - N_A^N \tag{1.72}$$

$$f_A^- = \int d\boldsymbol{r}\, w_A(\boldsymbol{r}) \left(\rho^{N}(\boldsymbol{r}) - \rho^{N-1}(\boldsymbol{r})\right) = N_A^{N} - N_A^{N-1} \tag{1.73}$$

According to Ayers et al. [102] the condensed Fukui functions are even more instructive indicators of molecular site reactivity than the actual Fukui function. The electron populations on the atoms N_A are determined from either the atom-in-molecule approaches in Sec. 1.3.2 or from the other methods discussed in Sec. 1.5.1

2 Atom-condensed 1DMs with a local electron density

2.1 Requirements for an improved AIM approach

2.1.1 A fundamental approach based on density matrices

The large majority of AIM methods are based on the partitioning of the electron density. This is the most accessible approach to the atom in the molecule, and both Baders' QTAIM technique and the Hirshfeld scheme are cornerstones in the AIM field. Furthermore, the advent of DFT and the development of DFT reactivity descriptors boosted even more the popularity of the density-based approach (see Sec. 1.5.3). In this approach, the properties of the atomic densities $\rho_A(\boldsymbol{r})$, e.g. their energy components, are obtained from a fragmentation of the molecular expectation value expressions using the AIM weight functions. Especially for fuzzy-AIM schemes, such fragmentation can be ambiguous because (i) the weight functions and the operators do not necessarily commute or (ii) several combinations of the integration variables with the atomic indices can be identified (see Sec. 1.4.1).

The QTAIM scheme does not suffer from these problems. QTAIM weight functions commute with the kinetic energy operator and the atomic potential energy is derived using the virial theorem [8]. The virial theorem, which governs the relationship between the potential and kinetic energies of a molecule, has been generalized by Bader from its global statement (which applies to the molecule as a whole) to a local statement defined at every point in space [52]. More precisely, the local virial theorem relates the potential energy density and the kinetic energy density distributions locally to the Laplacian of the electron density. Bader also postulated [103] and showed [104–106] that the integrated form of this theorem translates into a virial theorem satisfied

by each atom within a molecule. However, the uniqueness of QTAIM was recently disputed by Cioslowski and Karwowski who showed that arbitrary choices in the Lagrangian density can have an important influence on the atom-condensed values [107].

A unique correspondence between the molecular fragments and their properties is obtained in a more fundamental approach to the AIM based on density matrices [33] (see Sec. 1.4.2). Note that each N-electron property can be directly expressed in terms of the (partitioned) N-electron density matrix (see Sec. 1.2.1). In this chapter, we restrict ourselves to the partitioning of the molecular 1DM over the atoms, to determine atom-condensed values for the one-electron properties at all levels of theory and for the two- electron properties at the Hartree-Fock level of theory. A partitioning of the 2DM is possible, but is not considered in this thesis. To determine the spin-dependent properties of the atoms, the spin up and down 1DMs,

$$\rho^{\sigma}(\boldsymbol{r},\boldsymbol{r}') = \rho(\boldsymbol{r}\sigma,\boldsymbol{r}\sigma), \tag{2.1}$$

should be partitioned separately over the atoms. For spin-independent properties, partitioning of the spin-summed 1DM,

$$\rho(\boldsymbol{r},\boldsymbol{r}') = \sum_{\sigma}\rho^{\sigma}(\boldsymbol{r},\boldsymbol{r}') \tag{2.2}$$

is sufficient. For spin-singlet molecules the electron spin can also be discarded without loss of information, since

$$\rho(\boldsymbol{r},\boldsymbol{r}') = 2\rho^{\sigma}(\boldsymbol{r},\boldsymbol{r}'). \tag{2.3}$$

2.1.2 Localization of the 1DM to atomic regions

Single-atom 1DMs

Single-atom 1DMs are obtained if a single partitioning of unity in atomic weight functions ($1 = \sum_A w_A(\boldsymbol{r})$) is introduced in the expression for the molecular 1DM. The partitioning of unity can be introduced at various places, e.g.

$$\rho(\boldsymbol{r},\boldsymbol{r}') = \sum_A \frac{1}{2}\Big(w_A(\boldsymbol{r}) + w_A(\boldsymbol{r}')\Big)\rho(\boldsymbol{r},\boldsymbol{r}') \tag{2.4}$$

$$\rho(\boldsymbol{r},\boldsymbol{r}') = \sum_A \int d\boldsymbol{r}''\rho^{\frac{1}{2}}(\boldsymbol{r},\boldsymbol{r}'')\Big(w_A(\boldsymbol{r}'')\Big)\rho^{\frac{1}{2}}(\boldsymbol{r}'',\boldsymbol{r}') \tag{2.5}$$

$$\vdots$$

The resulting atom-condensed 1DMs are inherently non-local. This can be inferred from the corresponding expressions in terms of the eigenvalue decomposition of the molecular 1DM,

$$\rho(\boldsymbol{r},\boldsymbol{r'}) = \sum_i n_i \varphi_i(\boldsymbol{r})\varphi_i(\boldsymbol{r'}), \tag{2.6}$$

where n_i represent the natural populations and $\varphi_i(\boldsymbol{r})$ the natural orbitals. For the definition in Eq. (2.4), the atom-condensed 1DM reads

$$\begin{aligned}\rho_A(\boldsymbol{r},\boldsymbol{r'}) &= \frac{1}{2}\Big(\, w_A(\boldsymbol{r}) + w_A(\boldsymbol{r'})\Big)\,\rho(\boldsymbol{r},\boldsymbol{r'}) \\ &= \frac{1}{2}\sum_i n_i \left[\Big(\, w_A(\boldsymbol{r})\varphi_i(\boldsymbol{r})\Big)\,\varphi_i(\boldsymbol{r'}) + \Big(\, w_A(\boldsymbol{r'})\varphi_i(\boldsymbol{r'})\Big)\,\varphi_i(\boldsymbol{r})\right]. \end{aligned}\tag{2.7}$$

For a delocalized molecular valence orbital φ_i, the decomposition in Eq. (2.7) contains both a function $w_A(\boldsymbol{r})\varphi_i(\boldsymbol{r})$ localized to the atomic region A and a delocalized molecular natural $\varphi_i(\boldsymbol{r'})$. The definition in Eq. (2.5) leads to,

$$\begin{aligned}\rho_A(\boldsymbol{r},\boldsymbol{r'}) &= \int d\boldsymbol{r''}\rho^{\frac{1}{2}}(\boldsymbol{r},\boldsymbol{r''})\,\Big(\, w_A(\boldsymbol{r''})\Big)\,\rho^{\frac{1}{2}}(\boldsymbol{r''},\boldsymbol{r'}) \\ &= \sum_{ij}\left(n_i^{\frac{1}{2}} n_j^{\frac{1}{2}} \int d\boldsymbol{r''}\;\varphi_i(\boldsymbol{r''})w_A(\boldsymbol{r''})\varphi_j(\boldsymbol{r''})\right)\varphi_i(\boldsymbol{r})\varphi_j(\boldsymbol{r'}) \\ &= \sum_{ij}\Big(n_i^{\frac{1}{2}} n_j^{\frac{1}{2}}\,(w_A)_{ij}\Big)\,\varphi_i(\boldsymbol{r})\varphi_j(\boldsymbol{r'}) \\ &= \sum_{ij} n_{A,ij}\;\;\varphi_i(\boldsymbol{r})\varphi_j(\boldsymbol{r'}), \end{aligned}\tag{2.8}$$

where the atomic 1DM is expressed in terms of the delocalized molecular naturals $\varphi_i(\boldsymbol{r})$, and $n_{A,ii}$ represents the occupancies of the molecular naturals in the atoms. Consider the rather common case in which the antibonding orbitals of the molecule have vanishing occupancies ($n_i \approx 0$), e.g. H_2 at the Hartree-Fock level. It is clear from Eq. (2.8) that molecular orbitals with vanishing occupancies are also barely occupied in the atoms ($\forall_j : \;\; n_{A,ij} \approx 0$). Hence, only the bonding naturals appear in the summation (2.8), and the eigenfunctions of $\rho_A(\boldsymbol{r},\boldsymbol{r'})$ cannot be balanced linear combinations of the delocalized bonding and antibonding molecular naturals. As a logical consequence, these eigenfunctions are not localized to the atomic regions. It is clear that no single-atom partitioning scheme of the 1DM can have good localization properties.

Double-atom 1DMs

Double-atom 1DMs are obtained if the partitioning of unity in atomic weight functions ($1 = \sum_A w_A(\boldsymbol{r})$) is introduced twice in the expression for the molecular 1DM:

$$\begin{aligned}
\rho(\boldsymbol{r},\boldsymbol{r'}) &= \frac{1}{2}\sum_{AB}\Big[\Big(w_A(\boldsymbol{r})\Big)\,\rho(\boldsymbol{r},\boldsymbol{r'})\,\Big(w_B(\boldsymbol{r'})\Big) \\
&\qquad + \Big(w_B(\boldsymbol{r})\Big)\,\rho(\boldsymbol{r},\boldsymbol{r'})\,\Big(w_A(\boldsymbol{r'})\Big)\Big] \qquad (2.9)\\
\rho(\boldsymbol{r},\boldsymbol{r'}) &= \sum_A \int d\boldsymbol{r''}\rho^{\frac{1}{2}}(\boldsymbol{r},\boldsymbol{r''})\,\Big(w_A(\boldsymbol{r''})w_B(\boldsymbol{r''})\Big)\,\rho^{\frac{1}{2}}(\boldsymbol{r''},\boldsymbol{r'}). \qquad (2.10)\\
&\vdots
\end{aligned}$$

We distinguish between atomic ($A = B$) and bond matrices ($A \neq B$), according to the atomic indices. The eigenvectors of the bond matrices provide an orbital perspective on the changes in the atoms when bonds are formed, and are reminiscent in philosophy to the so-called Natural Orbitals for Chemical Valence introduced by Nalewajski et al. [108] and used to describe chemical bonds by Ziegler and co-workers [109, 110].

In contrast to the single-atom partitioning of the molecular 1DM, the double-atom scheme in Eq. (2.9) defines contributions that are clearly localized to the atom pairs:

$$\begin{aligned}
\rho_{AB}(\boldsymbol{r},\boldsymbol{r'}) &= \frac{1}{2}\Big[\Big(w_A(\boldsymbol{r})\Big)\,\rho(\boldsymbol{r},\boldsymbol{r'})\,\Big(w_B(\boldsymbol{r'})\Big) \\
&\qquad + \Big(w_B(\boldsymbol{r})\Big)\,\rho(\boldsymbol{r},\boldsymbol{r'})\,\Big(w_A(\boldsymbol{r'})\Big)\Big] \\
&= \sum_i n_i\Big[\frac{1}{2}\Big(w_A(\boldsymbol{r})\varphi_i(\boldsymbol{r})\Big)\Big(w_B(\boldsymbol{r'})\varphi_i(\boldsymbol{r'})\Big) \\
&\qquad + \frac{1}{2}\Big(w_B(\boldsymbol{r})\varphi_i(\boldsymbol{r})\Big)\Big(w_A(\boldsymbol{r'})\varphi_i(\boldsymbol{r'})\Big)\Big]. \qquad (2.11)
\end{aligned}$$

Also Mayer and Salvador [111] recently argued that a two-index approach is necessary to split the molecular 1DM in localized contributions. However, not all two-index approaches define localized contributions. The atomic and bond matrices derived from the definition in Eq. (2.10), e.g., are not localized to the regions of the atom pairs, as can be shown by an analogous reasoning as in Eq. (2.8)

2.1.3 The localization requirement for the electron density

Recently, Alcoba and coworkers [112] introduced a single-atom 1DM partitioning based on the Hilbert space analogue of the expression in Eq. (2.5) and the atomic weight functions of QTAIM,

$$\rho_{ij} = \sum_{A}\sum_{kl} \rho_{ik}^{\frac{1}{2}} \left(\int d\boldsymbol{r} \varphi_k(\boldsymbol{r}) w_A^{QTAIM}(\boldsymbol{r}) \varphi_l(\boldsymbol{r}) \right) \rho_{lj}^{\frac{1}{2}}. \tag{2.12}$$

In Eq. (2.12), the indices *i,j,k,l* refer to the orthonormal basis spanned by the eigenfunctions $\varphi_i(\boldsymbol{r})$ of the molecular 1DM. Note that in the basis set limit the expressions in $\boldsymbol{r}$-space and Hilbert space are equivalent. Although QTAIM weight functions are derived from a topological partitioning of the electron density over atomic basins with strict boundaries, the atomic densities obtained by Alcoba's method,

$$\rho_A(\boldsymbol{r}) = \sum_{ij} \left(n_i^{\frac{1}{2}} n_j^{\frac{1}{2}} (w_A)_{ij}^{QTAIM} \right) \varphi_i(\boldsymbol{r}) \varphi_j(\boldsymbol{r}), \tag{2.13}$$

are clearly not localized to the atomic basins (see the discussion following Eq. (2.8)), even not in the basis set limit. Alcoba's atomic densities differ qualitatively from QTAIM atomic densities.

This inconsistency between the partitioning of the 1DM and the underlying AIM method is avoided when Eq. (2.4) is used to partition the 1DM over the atoms. The corresponding atomic density reads (see Eq. (2.7))

$$\begin{aligned} \rho_A(\boldsymbol{r}) &= w_A(\boldsymbol{r}) \rho(\boldsymbol{r},\boldsymbol{r}) \\ &= \sum_i n_i \, w_A(\boldsymbol{r}) \varphi_i(\boldsymbol{r}) \varphi_i(\boldsymbol{r}), \end{aligned} \tag{2.14}$$

and coincides with the electron density from the AIM techniques discussed in Sec. 1.3.2:

$$\rho_A(\boldsymbol{r}) = w_A(\boldsymbol{r}) \rho(\boldsymbol{r}). \tag{2.15}$$

The requirement of consistency between the 1DM partitioning and the underlying AIM method is very important as it guarantees that the diagonal of the single-atom 1DM, the atomic electron density, is localized to the atomic regions. This facilitates considerably the interpretation of the properties of the quantum atoms. If the electron density of the atoms is not sufficiently localized and spreads over the molecular space, then (i) the properties of those quantum atoms are less likely to be transferable from one molecule to than the properties of the atoms in traditional chemistry [113, 114], (ii) the contribution of the

dipoles in the multipole expansion of the electrostatic potential is significantly larger, which hampers the applicability of the atomic charges in forcefield methods (see section 1.5.1), (iii) the interaction energies between the molecular atoms in energy decomposition schemes are possibly not within the range of typical dissociation energies [115], (iv) the electrons of several atoms, e.g. the hydrogen atom, are possibly not bound with respect to the Hamiltonian of the isolated atoms, (v) ... In the following, we construct a double-atom partitioning scheme of the 1DM that is consistent with an underlying electron density partitioning and has correct localization properties.

2.2 Density matrix extension of the Hirshfeld-I atom-in-molecule model

2.2.1 Theoretical derivation

To determine the properties of fuzzy atoms unambiguously (see Sec. 2.1), a density matrix extension of the Hirshfeld-I atom-in-molecule model was developed. In this new approach, that was introduced in **Paper 1**, the local Hirshfeld-I atomic densities are retrieved from the diagonal of the atomic 1DMs. In Sec. 2.1.3 we noted that the single-atom expression for partitioning of the molecular 1DM in Eq. (2.7) trivially satisfies this consistency requirement. However there are several alternative expressions that are chemically meaningful and also satisfy this condition. These are easily derived from the double-atom partitioning scheme

$$\rho(\boldsymbol{r},\boldsymbol{r}') = \sum_{AB} \rho_{AB}(\boldsymbol{r},\boldsymbol{r}'), \tag{2.16}$$

that defines atomic ($A = B$) and bond ($A \neq B$) matrices that are localized to the regions of the atom pairs:

$$\rho_{AB}(\boldsymbol{r},\boldsymbol{r}') = \frac{1}{2}\, w_A(\boldsymbol{r})\rho(\boldsymbol{r},\boldsymbol{r}')w_B(\boldsymbol{r}') + \frac{1}{2}\, w_B(\boldsymbol{r})\rho(\boldsymbol{r},\boldsymbol{r}')w_A(\boldsymbol{r}'). \tag{2.17}$$

To obtain single-atom 1DMs, the bond matrices can be distributed over the atoms in different ways. Two of these can be considered to be extreme cases and are studied in detail.

In the nonweighted scheme (indicated with a superscript n), the atoms A and B receive an equal share of the bond matrices,

$$\begin{aligned} \rho^n_A(\boldsymbol{r},\boldsymbol{r}') &= \sum_B \rho_{AB}(\boldsymbol{r},\boldsymbol{r}') \\ &= \frac{1}{2}\Big(w_A(\boldsymbol{r}) + w_A(\boldsymbol{r}')\Big)\, \rho(\boldsymbol{r},\boldsymbol{r}'), \end{aligned} \tag{2.18}$$

and the expression in Eq. (2.7) appears. On the diagonal,

$$\rho_A^n(\boldsymbol{r}) \quad = \quad w_A(\boldsymbol{r})\rho(\boldsymbol{r}) \tag{2.19}$$

the Hirshfeld-I AIM density is simply retrieved when AIM weight functions $w_A^{HI}(\boldsymbol{r})$ are used in the 1DM partitioning scheme:

$$w_A(\boldsymbol{r}) = w_A^{HI}(\boldsymbol{r}). \tag{2.20}$$

In the weighted scheme, the atoms receive a fraction $x_{A_{(B)}}(\boldsymbol{r},\boldsymbol{r}')$ of the AB bond matrices that differs at each point in space

$$\rho_A^w(\boldsymbol{r},\boldsymbol{r}') \quad = \quad \sum_B x_{A_{(B)}}(\boldsymbol{r},\boldsymbol{r}')\rho_{AB}(\boldsymbol{r},\boldsymbol{r}'). \tag{2.21}$$

The fractions should obey

$$x_{A_{(B)}}(\boldsymbol{r},\boldsymbol{r}') + x_{B_{(A)}}(\boldsymbol{r},\boldsymbol{r}') = 1, \tag{2.22}$$

and can be defined from the atomic decomposition of space:

$$x_{A_{(B)}}(\boldsymbol{r},\boldsymbol{r}') = \frac{1}{2}\left(\frac{w_A(\boldsymbol{r})}{w_A(\boldsymbol{r}) + w_B(\boldsymbol{r})} + \frac{w_A(\boldsymbol{r}')}{w_A(\boldsymbol{r}') + w_B(\boldsymbol{r}')}\right). \tag{2.23}$$

On the diagonal of the single-atom 1DM,

$$\rho_A^w(\boldsymbol{r}) \quad = \quad \sum_B \left(\frac{2w_A^2(\boldsymbol{r})w_B(\boldsymbol{r})}{w_A(\boldsymbol{r}) + w_B(\boldsymbol{r})}\right)\rho(\boldsymbol{r}), \tag{2.24}$$

the AIM density is recovered when weights functions $w_A(\boldsymbol{r})$ are used for the partitioning of the molecular 1DM that differ from the AIM weight functions $w_A^{HI}(\boldsymbol{r})$ and obey the following set of nonlinear equations:

$$\forall A: \quad w_A(\boldsymbol{r})\left[\sum_B \left(\frac{2w_A(\boldsymbol{r})w_B(\boldsymbol{r})}{w_A(\boldsymbol{r}) + w_B(\boldsymbol{r})}\right)\right] = w_A^{HI}(\boldsymbol{r}). \tag{2.25}$$

The set of nonlinear equations is easily solved through an iterative sequence that converges rapidly to a stable solution of Eq. (2.25) [116].

2.2.2 Characteristics of the 1DM fragments

The double-atom fragments

Although the double-atom fragments of the molecular 1DM are introduced in Sec. 2.2.1 as auxiliary matrices from which proper single-atom 1DM can be derived, they provide the most detailed insight into the characteristics of the partitioning scheme. The diagonal terms ρ_{AA} are positive semidefinite hermitian matrices, and their diagonalization leads to the natural orbitals $\varphi_{AA,i}$ and occupancies $n_{AA,i}$ for the atoms:

$$\rho_{AA}(\boldsymbol{r},\boldsymbol{r}') = \sum_i n_{AA,i}\varphi_{AA,i}(\boldsymbol{r})\varphi_{AA,i}(\boldsymbol{r}'). \tag{2.26}$$

Fig. 2.1 shows, for an exemplary calculation on CO, the characteristic features of these natural orbitals and occupancies. An important observation is

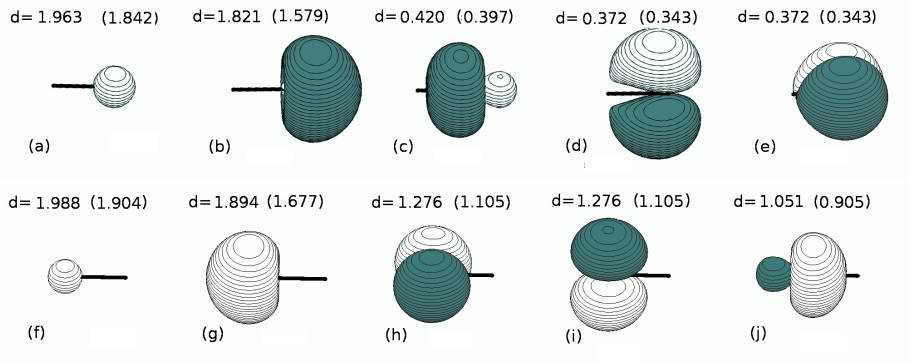


Figure 2.1: The dominant natural orbitals, and the corresponding occupations, of the atomic density matrix $\rho_{AA}(\boldsymbol{r},\boldsymbol{r}')$ of carbon (a-e) and oxygen (f-j) in a CO molecule calculated in the nonweighted scheme, at the HF/Aug-cc-pVDZ level. The occupations in the weighted scheme are also indicated, between brackets.

that these terms indeed qualify as one-electron density matrices, having occupancies $n_{AA,i}$ between 0 and 2. Furthermore, the atomic naturals are very well localized, in accordance with Eq. (2.11). Their shape resembles the naturals of isolated atoms, being only slightly deformed. When the weighted and nonweighted variants of the partitioning scheme are compared, the dominant natural orbitals of both approaches are visually indistinguishable, but their occupancies differ. In general, the occupancies of $\rho_{AA}(\boldsymbol{r},\boldsymbol{r}')$ are smaller for the weighted scheme. The core 1s atomic orbitals, represented in Fig. 2.1 by orbitals (a) and (f), have a population well below 2.000

(the expected value of an orbital not involved in bonding) in the weighted scheme (1.842 and 1.904) and more acceptable values (1.963 and 1.988) in the nonweighted scheme. Also for orbitals (b) and (g), the population in the weighted scheme (1.579 and 1.677) drops well below the more intuitive values in the nonweighted scheme (1.821 and 1.894). The shape and occupancies of the latter orbitals suggests that these orbitals are occupied by the free electron pair and therefore should have an occupancy of 2.000. The difference between the occupancies in both schemes is typical, and due to the more diffuse character of the atomic weight function in the weighted scheme [116]. The weighted scheme assigns a smaller fraction of electrons to the atoms and a larger fraction, including non-negligible contributions from core electrons and free electron pairs, to the bonds.

The eigenvalue decomposition of the nondiagonal terms ρ_{AB} and ρ_{BA} provides an orbital perspective on the changes that occur when bonds are formed:

$$\rho_{AB}(\boldsymbol{r},\boldsymbol{r}') + \rho_{BA}(\boldsymbol{r},\boldsymbol{r}') = \sum_i n_{AB,i}\varphi_{AB,i}(\boldsymbol{r})\varphi_{AB,i}(\boldsymbol{r}'). \tag{2.27}$$

The bond orbitals $\varphi_{AB,i}(\boldsymbol{r})$ come in bonding-antibonding pairs, as shown in Fig. 2.2 for the exemplary calculation on CO. The bonding orbitals have positive occupancies between 0 and 1, while the antibonding orbitals have negative occupancies between 0 and -1. This can be readily understood by simply

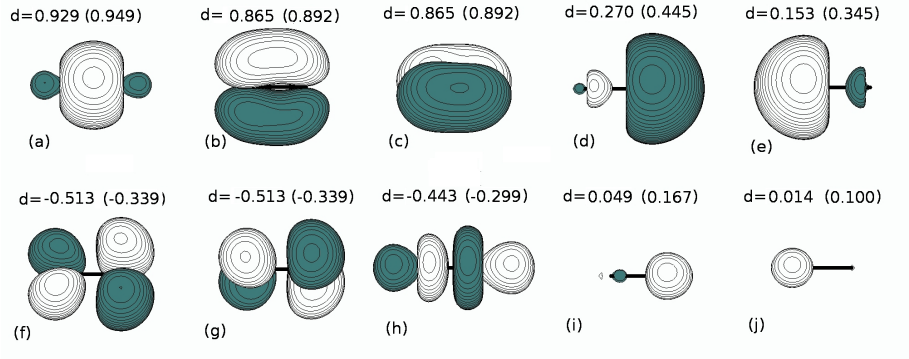


Figure 2.2: The dominant eigenvectors and occupations of $\rho_{AB}(\boldsymbol{r},\boldsymbol{r}') + \rho_{BA}(\boldsymbol{r},\boldsymbol{r}')$ in a CO molecule calculated in the nonweighted scheme, at the HF/Aug-cc-pVDZ level. The occupations in the weighted scheme are also indicated, between brackets.

rewriting the hermitian matrix as [116],

$$\begin{aligned}\rho_{AB}(\boldsymbol{r},\boldsymbol{r}') + \rho_{BA}(\boldsymbol{r},\boldsymbol{r}') &= \frac{1}{2}\sum_i n_i \varphi_{A+B,i}(\boldsymbol{r})\varphi_{A+B,i}(\boldsymbol{r}') \\ &\quad -\frac{1}{2}\sum_i n_i \varphi_{A-B,i}(\boldsymbol{r})\varphi_{A-B,i}(\boldsymbol{r}') \end{aligned} \tag{2.28}$$

where n_i represents the populations of the molecular natural orbitals $\varphi_i(\boldsymbol{r})$ and

$$\varphi_{A\pm B,i}(\boldsymbol{r}) = \Big(w_A(\boldsymbol{r}) \pm w_B(\boldsymbol{r}) \Big)\, \varphi_i(\boldsymbol{r}). \tag{2.29}$$

When a homonuclear diatomic molecule is considered and binary (QTAIM) atomic weight functions are assumed, it is clear from Eq. (2.29) that the functions $\varphi_{A\pm B,i}(\boldsymbol{r})$ are orthonormal. In that case, the occupancies $n_{AB,i}$ of the bonding/antibonding pairs of orbitals $\varphi_{AB,i}(\boldsymbol{r})$ are exactly opposite and the bond matrix represents a shift of electrons from antibonding orbitals (occupied in $\sum_A \rho_{AA}(\boldsymbol{r},\boldsymbol{r}')$) to the corresponding bonding orbitals. This picture appeals to the chemical notion of bonding. The other extreme case is represented by atomic weight functions that are maximally delocalized, e.g. $w_A(\boldsymbol{r}) = w_B(\boldsymbol{r}) = ... = w_N(\boldsymbol{r})$. In that case, the second term in Eq. (2.28) vanishes and no negative occupancies appear. The present fuzzy-atom model is situated in between both extreme cases, and the negative occupancies are smaller (in absolute value) than the positive occupancies. Note that the negative occupancies are also less outspoken for the weighted scheme, due to the more diffuse character of the atomic weight functions in the weighted scheme [116]. Also note that some smaller contributions with positive occupancies appear in the bond matrix, that are not related to the molecular σ, π_1 and π_2 bonds. These are the contributions from the core electrons ((i) and (j)) and the free electron pairs ((d) and (e)) that are missing in the atomic matrices. They are invariably larger for the weighted scheme.

For AIM models based on fuzzy atom definitions, the difference in occupancies of the corresponding bonding and antibonding orbitals translates into a positive net population for the bond matrix. This also implies that the summed atomic matrices do not carry the total number of electrons,

$$\int d\boldsymbol{r} \Big(\rho_{AB}(\boldsymbol{r},\boldsymbol{r}) + \rho_{BA}(\boldsymbol{r},\boldsymbol{r}) \Big) > 0 \quad ; \quad \sum_A \int d\boldsymbol{r} \rho_{AA}(\boldsymbol{r},\boldsymbol{r}) < N. \tag{2.30}$$

The charge transfer from the atoms to the bonds is proportional to the bond order. For a test set of ca. 50 molecules calculated at the HF/Aug-cc-pVDZ level and using the Hirshfeld-I AIM model, Fig. 2.3 shows the correlation between the population of the bond matrix (as a matrix trace) and the shared

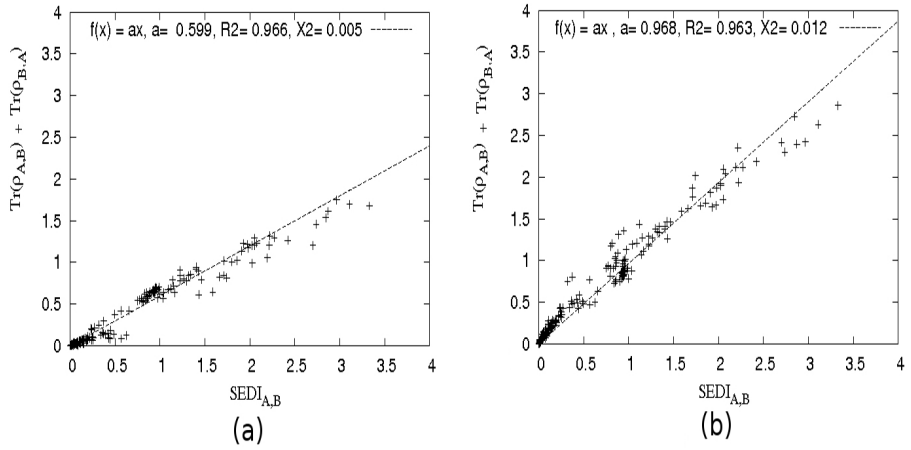


Figure 2.3: Correlation between the SEDI index and the traces of the bond matrices for the nonweighted scheme (a) and the weighted scheme (b) at the HF/Aug-cc-pVDZ level.

electron distribution index (SEDI, see Eq. (1.63)). A conceptually analogous correlation was first given by I. Mayer for the Mulliken case [117]. For both the weighted and the nonweighted schemes there is a strong linear correlation (R^2 > 0.96). The net population of the bond matrix is also quite stable with respect to basis set size, as can be inferred from Table 2.1, with differences less than 0.01 going from DZ to TZ and less than 0.004 going from TZ to QZ.

CO	Aug-cc-pVDZ	Aug-cc-pVTZ	Aug-cc-pVQZ
A , B	$\mathrm{Tr}(\rho_{A,B}+\rho_{B,A})$	$\mathrm{Tr}(\rho_{A,B}+\rho_{B,A})$	$\mathrm{Tr}(\rho_{A,B}+\rho_{B,A})$
C , C	4.955 (4.528)	4.956 (4.529)	4.958 (4.531)
O , O	7.508 (6.743)	7.516 (6.751)	7.518 (6.752)
C , O	1.536 (2.728)	1.528 (2.720)	1.524 (2.716)

Table 2.1: The number of electrons present in the atomic density and bond matrices for CO, in the nonweighted and weighted scheme. Bracketed values correspond to the weighted scheme. The values are converged with respect to grid size.

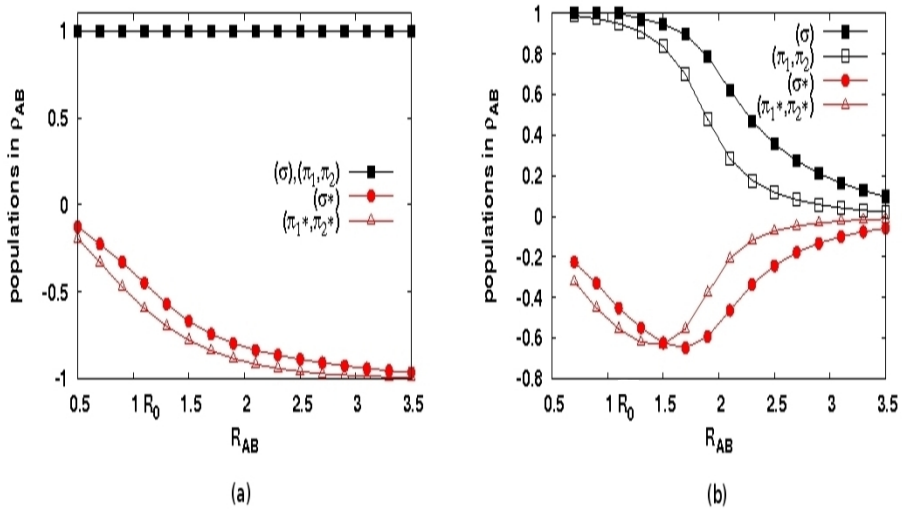


Figure 2.4: The populations of the dominant eigenvectors (σ, π_1,π_2, σ^*,π_1^* and π_2^*) in the bond matrix ($\rho_{AB} + \rho_{BA}$) of N_2, for bond lengths R_{AB}(Å). The molecular 1DMs were calculated at (a) the ROHF level and (b) the full-valence CASSCF level in the Aug-cc-pVTZ basis set using the nonweighted scheme. R_0 is the equilibrium bond length.

The double-atom 1DMs show a satisfactory behavior in dissociating molecules. Fig. 2.4 displays the populations of the dominant eigenvectors (σ, π_1,π_2, σ^*,π_1^* and π_2^*) in the bond matrix ($\rho_{AB} + \rho_{BA}$) of N_2. In accordance with the wrong dissociation behavior of molecules at the Hartree-Fock level, there still is a transfer of electrons from antibonding to bonding orbitals in the bond matrix at infinite atomic separation. A more realistic picture is obtained by the full-valence CASSCF (active space spanned by the 2s and 2p orbitals) molecular calculation, where a smooth vanishing of the bond matrix is observed for increasing bond lengths. In accordance with chemical expectations [118], it is observed that the π-bonds 'snap' at an earlier distance than the σ-bond.

The single-atom fragments

The characteristics of the single-atom 1DM follows from (i) their derivation from double-atom 1DMs and (ii) the constraint that local and positive semidefinite atom densities are retrieved on the diagonal. Diagonalization of these hermitian matrices,

$$\rho_A(\boldsymbol{r},\boldsymbol{r}') = \sum_i n_{A,i}\varphi_{A,i}(\boldsymbol{r})\varphi_{A,i}(\boldsymbol{r}'), \tag{2.31}$$

generates atomic orbitals $\varphi_{A,i}(\boldsymbol{r})$ with mostly positive occupancies $n_{A,i}$ and tails that stretch into the bonding regions. Negative eigenvalues, corresponding with antibonding orbitals from the bond matrices, do appear but are suppressed by the constraint that the diagonal is positive definite in $\boldsymbol{r}$-space. The single-atom 1DMs in the weighted and nonweighted scheme are very similar, since for both schemes the diagonal in $\boldsymbol{r}$-space coincides with the Hirshfeld-I electron density. An exemplary result for the CO molecule is shown in Fig. 2.5.

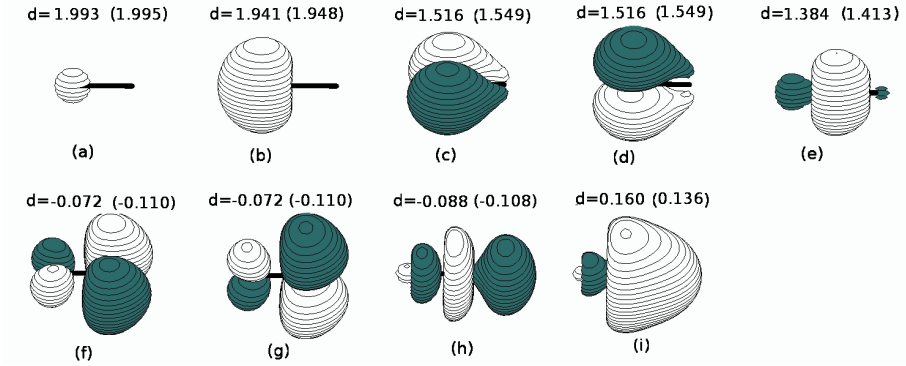


Figure 2.5: The dominant natural orbitals and populations of the single-atom 1DM of oxygen in a CO molecule, calculated in the nonweighted scheme at the HF/Aug-cc-pVDZ level. The populations in the weighted scheme are also indicated, between brackets.

2.3 Applications

We evaluated the energy contributions associated with the 1DM fragments at the Hartree-Fock level, as a pilot test to determine whether the new 1DM-based AIM-approach leads to chemically acceptable results. A detailed discussion can be found in **Paper 3**.

2.3.1 Energy decomposition based on single-atom 1DMs

Chemistry deals with energy differences between reagents and products that are several orders of magnitude smaller than typical electronic energies of molecules. This implies that the electronic arrangements of atoms in a molecular environment are only slightly distorted compared to the isolated atoms. In the theory of Interacting Quantum Atoms (IQA) [119–121], the molecular Hamiltonian $\hat{H}$ of Eq. (1.2) is decomposed into contributions $\hat{H}_A{}^{self}$ that closely resemble those of the isolated atoms and contributions $\hat{H}_{AB}{}^{int}$ that can be interpreted as interactions between the atoms:

$$\hat{H} = \sum_A \hat{H}_A^{self} + \sum_{A<B} \hat{H}_{AB}^{int} \tag{2.32}$$

$$\hat{H}_A^{self} = \sum_i \delta(\boldsymbol{r}_i \in \Omega_A) \left(\nabla_i^2 + \frac{Z_A}{|\boldsymbol{r}_i - \boldsymbol{R}_A|} \right) + \sum_{i<j} \frac{\delta(\boldsymbol{r}_i \in \Omega_A)\delta(\boldsymbol{r}_j \in \Omega_A)}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} \tag{2.33}$$

$$\hat{H}_{AB}^{int} = \frac{Z_A Z_B}{|\boldsymbol{R}_A - \boldsymbol{R}_B|} + \sum_i \left(\frac{Z_A \delta(\boldsymbol{r}_i \in \Omega_B)}{|\boldsymbol{r}_i - \boldsymbol{R}_A|} + \frac{Z_B \delta(\boldsymbol{r}_i \in \Omega_A)}{|\boldsymbol{r}_i - \boldsymbol{R}_B|} \right) + \sum_{i<j} \left(\frac{\delta(\boldsymbol{r}_i \in \Omega_A)\delta(\boldsymbol{r}_j \in \Omega_B)}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} + \frac{\delta(\boldsymbol{r}_i \in \Omega_B)\delta(\boldsymbol{r}_j \in \Omega_A)}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} \right) \tag{2.34}$$

Evaluation of the expectation value of these operators, with respect to the ground state of the molecule, leads to a corresponding decomposition of the molecular energy E:

$$E = \sum_A E_A^{self} + \sum_{AB} E_{AB}^{int}. \tag{2.35}$$

If $\delta(\boldsymbol{r}_i \in \Omega_A)$ in Eq. (2.33) is represented by a Hirshfeld-I weight function ambiguities arise with respect to e.g. the exact place of $w_A(\boldsymbol{r}_i)^{HI}$ in the expression for the expectation value and the representation of the kinetic energy operator (see Sec. 1.4.1). In an alternative 1DM-based definition for the

(closed-shell) Hartree-Fock case, the δ-functions in the operator are replaced by a partitioning of the 1DM [122]:

$$\begin{aligned} E_A^{self} &= \sum_{ij}(t+V_A^{ext})_{ij}(\rho_A)_{ij} \\ &\quad +\frac{1}{2}\sum_{ijkl}V_{ijkl}\left((\rho_A)_{ik}(\rho_A)_{jl}-\frac{1}{2}(\rho_A)_{il}(\rho_A)_{jk}\right) \end{aligned} \tag{2.36}$$

$$\begin{aligned} E_{AB}^{int} &= \frac{Z_A Z_B}{r_{AB}}+\sum_{ij}(V_A^{ext})_{ij}(\rho_B)_{ij}+\sum_{ij}(V_B^{ext})_{ij}(\rho_A)_{ij} \\ &\quad +\sum_{ijkl}V_{ijkl}\left((\rho_A)_{ik}(\rho_B)_{jl}-\frac{1}{2}(\rho_A)_{il}(\rho_B)_{jk}\right) \end{aligned} \tag{2.37}$$

Eqs. (2.36) and (2.37) are conveniently expressed in a one-electron Hilbert space and require only the Hamiltonian matrix elements (one- and two-electron integrals), that are routinely computed in ab initio programs,

$$\begin{aligned} t_{ij} &= -\frac{1}{2}\int \phi_i(\boldsymbol{r})\nabla^2\phi_j(\boldsymbol{r})d\boldsymbol{r} \\ (V_A^{ext})_{ij} &= \int \phi_i(\boldsymbol{r})\frac{Z_A}{|\boldsymbol{r}-\boldsymbol{R}_A|}\phi_j(\boldsymbol{r})d\boldsymbol{r} \\ V_{ijkl} &= \int \phi_i(\boldsymbol{r}_1)\phi_j(\boldsymbol{r}_2)\frac{1}{|\boldsymbol{r}_1-\boldsymbol{r}_2|}\phi_k(\boldsymbol{r}_1)\phi_l(\boldsymbol{r}_2)d\boldsymbol{r_1}d\boldsymbol{r_2}, \end{aligned} \tag{2.38}$$

and the single-atom 1DMs expressed in the orthonormal basis of real functions $\phi_i(\boldsymbol{r})$ (e.g. the Hartree-Fock orbitals) :

$$(\rho_A)_{ij} = \int d\boldsymbol{r}d\boldsymbol{r'}\phi_i(\boldsymbol{r})\rho_A(\boldsymbol{r},\boldsymbol{r'})\phi_j(\boldsymbol{r'}). \tag{2.39}$$

The molecular energy E is not an experimentally determined quantity in chemistry, in contrast to e.g. the atomization energy. To determine whether the new 1DM-based AIM-approach leads to chemically acceptable results, it is useful to consider the atomization energy ΔE^{at} instead, that can be obtained from Eq. (2.35) if the energies of the isolated atoms are subtracted:

$$\begin{aligned} E-\sum_A E_A^0 &= \sum_A\left(E_A^{self}-E_A^0\right)+\sum_{A<B}E_{AB}^{int}. \\ \Delta E^{at} &= \sum_A E_A^{prom}+\sum_{A<B}E_{AB}^{int}. \end{aligned} \tag{2.40}$$

The atomic promotion energy E_A^{prom} represents the energy change in an isolated atom that occurs when its electron density is deformed into the AIM density. E_{AB}^{int} is the interaction energy between these deformed atoms in the molecule. Both quantities are not observable, but it is clear that these cannot have chemical relevance if (i) their absolute values exceed the chemical energy range (0.05 - 0.5 Hartree) by several orders of magnitude, (ii) if the interactions energies between chemically non-bonded atoms are not singled out with a small absolute value, (iii) if the interaction energies are not similar for chemically similar bonds and (iv) if there is no linear correlation between the interaction energies and the bond order for a chosen pair of elements [115]. Since delocalization of the electron density leads to larger absolute values for the promotion and interaction energies, it is believed that the requirement of a local electron density for the atoms is a necessary condition to obtain a chemically relevant energy decomposition scheme [115].

In chemistry, it is more convenient to interpret the atomization energy solely in terms of contributions from the chemical bonds in the molecule.

$$\Delta E^{at} = \sum_{A<B} E_{AB}^{bond} \tag{2.41}$$

To rewrite Eq. (2.40) in terms of average bond energies E_{AB}^{bond}, we propose to distribute the atomic promotion energies over the atom pairs ($A < B$),

$$E_{AB}^{prom} = E_A^{prom} \left(\frac{E_{AB}^{int}}{\sum_C E_{AC}^{int}} \right) + E_B^{prom} \left(\frac{E_{AB}^{int}}{\sum_C E_{BC}^{int}} \right), \tag{2.42}$$

and average bond energies E_{AB}^{bond} are defined from pairwise energy contributions [122]:

$$E_{AB}^{bond} = E_{AB}^{prom} + E_{AB}^{int}. \tag{2.43}$$

The average bond energies must not be confused with bond dissociation energies (that also sum to the atomization energy of the molecule). While the bond dissociation energy differs for two equivalent bonds that are consecutively broken, the average bond energy is equal for two equivalent bonds in a molecule. It is important to note that the bond energies are obtained through the very delicate balance of promotion and interaction energies in Eq. (2.42), and therefore we only expect to get transferable values for the bond energies if the promotion and interaction energies are chemically relevant.

	bond	$SEDI_{AB}$	f_{AB}^{int}	E_{AB}^{int}	E_{AB}^{prom}	E_{AB}^{bond}
CH_4	C-H	0.99	-0.30	-0.45	0.33	-0.12
	H$\cdots$H	0.07	-0.01	-0.02	0.01	-0.01
CH_3CH_3	C-C	1.23	-0.39	-0.58	0.48	-0.10
	C-H	0.95	-0.30	-0.43	0.32	-0.12
	C$\cdots$H	0.10	-0.02	-0.03	0.02	-0.01
	H$\cdots$H(*)	0.07	-0.01	-0.02	0.01	-0.01
CH_2CH_2	C-C	2.06	-0.63	-0.98	0.81	-0.17
	C-H	0.96	-0.30	-0.44	0.32	-0.12
	C$\cdots$H	0.15	-0.03	-0.04	0.03	-0.01
	H$\cdots$H(*)	0.06	-0.01	-0.01	0.01	-0.01
CH-CH	C-C	2.98	-0.88	-1.38	1.13	-0.25
	C-H	0.96	-0.30	-0.43	0.33	-0.10
	C$\cdots$H	0.15	-0.02	-0.04	0.03	-0.01
	H$\cdots$H	0.01	0.00	0.01	0.00	0.00
CH_3OH	C-O	1.44	-0.48	-0.79	0.67	-0.12
	C-H	0.94	-0.29	-0.42	0.32	-0.10
	C-H	0.93	-0.29	-0.41	0.32	-0.10
	O-H	0.85	-0.29	-0.62	0.51	-0.11
	O$\cdots$H	0.14	-0.03	-0.04	0.03	-0.01
	O$\cdots$H	0.13	-0.03	-0.05	0.03	-0.01
H_2CO	C-O	2.29	-0.77	-1.34	1.15	-0.20
	C-H	0.94	-0.29	-0.41	0.32	-0.09
	O$\cdots$H	0.20	-0.03	-0.04	0.03	-0.02
CHOOH	C-O	1.35	-0.47	-1.03	0.92	-0.11
	C-O	1.98	-0.70	-1.40	1.21	-0.19
	C-H	0.85	-0.27	-0.40	0.32	-0.08
	O-H	0.75	-0.26	-0.62	0.51	-0.12
	O$\cdots$HC	0.17	-0.04	-0.04	0.03	-0.02
	O$\cdots$HC	0.20	-0.04	-0.04	0.02	-0.02
	O$\cdots$HO	0.04	-0.01	-0.09	0.07	-0.02
CO_2	C-O	2.08	-0.74	-1.52	1.31	-0.21
	O$\cdots$O	0.46	-0.07	0.02	-0.01	0.01
CO	C-O	2.88	-0.95	-1.58	1.29	-0.29

Table 2.2: Interaction energies (E_{AB}^{int}, in a.u.) for the molecular Hirshfeld-I fragments within the nonweighted scheme of some small molecules calculated at the ROHF/Aug-cc-pVTZ level of theory. Atomic promotion energies are attributed to the bonds (E_{AB}^{prom}) to derive Hartree-Fock bond energies (E_{AB}^{bond}). For comparison, the Hirshfeld-I SEDI-index is included as a bond order indicator. The Fock part of the interaction energy (f_{AB}^{int}) seems to be a robust indicator of covalent interaction energy [122]. The notation $[-]$ indicates chemical bonding, $[\cdots]$ refers to interaction between chemically non-bonded atoms and (*) indicates that geminal atoms are considered.

A selection from the results obtained in **Paper 3** is presented in Table 2.2, where one can see that the new 1DM-based method to decompose the energy leads to chemically acceptable results. The bond energies are clearly transferable, sufficiently small for non-bonded atom pairs, and correlate linearly with the bond order index (for a chosen pair of elements). This is all the more surprising since (i) the interaction energies between the 1DM fragments and the atomic promotion energies are about 3-7 times larger than the bond energies and (ii) the linear correlation between the interaction energies and the bond energies is not strong (R^2=0.6) [122]. This is a strong indication that the interaction energies and the atomic promotion energies are very well balanced and that the interaction energies fulfil the requirements of chemical relevance: (i) their magnitude is quite reasonable, (ii) their values single out chemically not bonded atoms with a small value, (iii) they have similar values for chemically similar bonds and (iv) their values correlate linearly with the bond order for a chosen pair of the elements. The interaction energies are all the more interesting since they do not depend on the choice of reference atoms E_A^0 (see Eq. 2.40).

The advantages of the 1DM-based approach

Two important advantages of the current approach are the reduction of ambiguities and the straightforward implementation in Hilbert space. These are studied intensively in **Paper 3**, while only the main results are presented here. With respect to the first issue, the kinetic energy of the molecular (1DM) fragments,

$$t_A[\rho_A] = -\frac{1}{2}\int d\boldsymbol{r}\ \nabla^2\rho_A(\boldsymbol{r},\boldsymbol{r'})\Big|_{\boldsymbol{r}=\boldsymbol{r'}} = \frac{1}{2}\int d\boldsymbol{r}\ \ \nabla\cdot\nabla'\rho_A(\boldsymbol{r},\boldsymbol{r'})\Big|_{\boldsymbol{r}=\boldsymbol{r'}} \qquad (2.44)$$

can be compared numerically to different fragments of the molecular kinetic energy,

$$t_A[w_A] = -\frac{1}{2}\int d\boldsymbol{r}\ w_A(\boldsymbol{r})\nabla^2\rho(\boldsymbol{r},\boldsymbol{r'})\Big|_{\boldsymbol{r}=\boldsymbol{r'}} \qquad (2.45)$$

$$t'_A[w_A] = \frac{1}{2}\int d\boldsymbol{r}\ w_A(\boldsymbol{r})\nabla\cdot\nabla'\rho(\boldsymbol{r},\boldsymbol{r'})\big|_{\boldsymbol{r}=\boldsymbol{r'}}\,, \qquad (2.46)$$

A typical result of such comparison is shown in Table 2.3. It can be analytically proven using integration by parts that values for $t_A[\rho_A]$ and $t_A[w_A]$ coincide [122], while values for $t'_A[w_A]$ differ in the order of tenths of a Hartree. For CH_3OH, the latter values show e.g. a counterintuitive increase of kinetic energy (+0.11 a.u.) in the highly electron deficient hydrogen atom (Q_A=+0.44) of the hydroxyl group. The kinetic energy of an electron deficient hydrogen

	A	t_A^0	$t_A[\rho_A]$ - t_A^0	$t_A[w_A]$ - t_A^0	$t'_A[w_A]$ - t_A^0
CH_3OH	C	37.68	0.00	0.00	-0.13
	H	0.50	0.10	0.10	0.21
	H	0.50	0.11	0.11	0.22
	O	74.80	0.37	0.37	-0.04
	H	0.50	-0.10	-0.10	0.11

Table 2.3: Kinetic energy ($t_A[\rho_A]$) of the single-atom 1DMs with a Hirshfeld-I diagonal in $\boldsymbol{r}$-space compared to the fragmented molecular kinetic energies ($t_A[w_A]$) and ($t'_A[w_A]$) obtained using the Hirshfeld-I atomic weight functions. Values (in a.u.) are presented relative to the kinetic energy (t_A^0) of the neutral isolated atom. Computations were performed at the ROHF/Aug-cc-pVTZ level of theory using the nonweighted partitioning scheme.

atom in its ground state is expected to vary linearly between 0.5 a.u. (neutral atom) and 0 a.u. (proton) [48, 49, 123]. With an atomic charge of +0.44, relatively strong deformations are needed to get a value for the kinetic energy larger than 0.6 a.u.. Note that in general the virial theorem is satisfied for the isolated ground-state atoms at the Hartree-Fock level [124], implying that the absolute values of the kinetic energy and the total electronic energy are equal.

	A	t_A^0	$\tilde{t_A}[(\rho_A)]$ - t_A^0	$t_A[\rho_A]$ - t_A^0
CH_3OH	C	37.68	0.05	0.00
	H	0.50	0.10	0.10
	H	0.50	0.11	0.11
	O	74.80	0.31	0.37
	H	0.50	-0.09	-0.10

Table 2.4: Kinetic energy of the single-atom 1DMs with a Hirshfeld-I diagonal in $\boldsymbol{r}$-space, calculated with the energy expressions in $\boldsymbol{r}$-space ($t_A[\rho_A]$) and Hilbert space ($\tilde{t_A}[\rho_A]$). Values (in a.u.) are presented relative to the kinetic energy (t_A^0) of the neutral isolated atom. The molecular 1DM was calculated at the ROHF/Aug-cc-pVTZ level of theory.

With respect to the straightforward implementation of the current approach in one-electron Hilbert space, Table 2.4 displays typical values for the atom-condensed kinetic energy obtained with both the $\boldsymbol{r}$-space expression in Eq. (2.44) and the corresponding expression in one-electron Hilbert space using a finite basis set:

$$\tilde{t_A}[(\rho_A)_{ij}] \quad = \quad \sum_{ij}(t)_{ij}(\rho_A)_{ij}. \tag{2.47}$$

The kinetic energies obtained with the *r*-space expression are converged with respect to grid size and basis set size [122]. Since for infinite basis sets both expressions should coincide, it is clear that even for the large Aug-cc-pVTZ basis set used in Table 2.4 the values obtained with the energy expression in one-electron Hilbert space are still not sufficiently converged (except for the hydrogen atoms).

On the other hand, the implementation of the full energy decomposition in one-electron Hilbert space is computationally much more efficient than the implementation in *r*-space, as long as the size of the molecule or the molecular basis set does not grow too large. This is evident from Fig. 2.6, which presents data for the molecular species investigated in [122]. Analysis of the

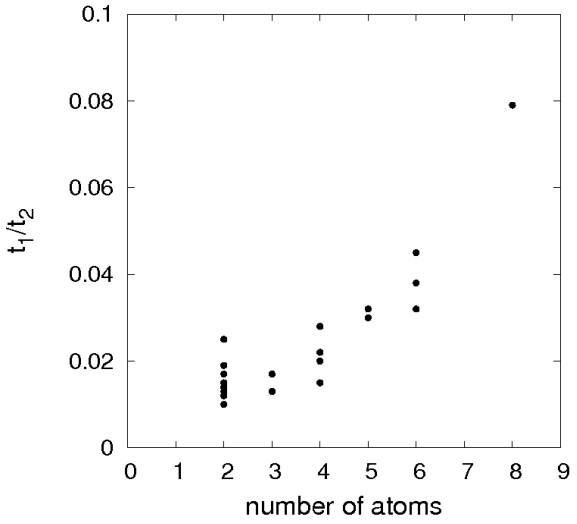


Figure 2.6: Calculation times for the energy decomposition (at the ROHF/Aug-cc-pVTZ level of theory) using the Hilbert-space approach (t_1) relative to the computational cost of using an *r*-space approach (t_2) as a function of the number of atoms in the molecule.

computational efficiency shows that there is a trade-off between the number of integrals that needs to be computed and the size of the grids required to reach an acceptable level of accuracy [122]. When using numerical integration techniques in *r*-space relatively few integrals need to computed, determined by the number of occupied naturals in the molecule. This advantage of the *r*-space approach would become less significant for post-HF levels of theory. On the other hand, large integration grids are required. For the energy

expressions in one-electron Hilbert space (see Eq. 2.36-2.37), the one- and two-electron integrals are computed fast by ab initio programs. The atomic overlap integrals,

$$S_{ij}^{A} = \int d\boldsymbol{r}\phi_i(\boldsymbol{r})w_A(\boldsymbol{r})\phi_j(\boldsymbol{r}), \tag{2.48}$$

require only modest grids and are sufficient to determine the single-atom 1DMs in the basis of the molecular naturals $\phi_i(\boldsymbol{r})$. On the other hand, the number of energy contributions that has to be calculated is large since the summations in Eq. (2.36) and Eq. (2.37) run unavoidably over the full dimension of the one-electron Hilbert space.

2.3.2 The energy of the double-atom bond matrix

The properties of the single-atom 1DMs are quite similar for the nonweighted and the weighted scheme. At the level of the double-atom 1DMs, the characteristics of both schemes diverge. We already mentioned that the bond matrices have a larger net population in the weighted scheme, including undesirable contributions from the core electrons and free electron pairs (see Sec. 2.2.2). To discriminate between both schemes, it is interesting to compare the Hartree-Fock energy terms associated with the double-atom contributions.

We identify the total energy associated with the bond contributions as the energy difference between the molecular 1DM with and without its bond contributions,

$$\sum_{A<B} E_{AB} = E_{mol} - E\left[\sum_A \rho_{AA}\right]. \tag{2.49}$$

For the closed-shell Hartree-Fock case, the energy of the molecular 1DM without its bond contributions is readily expressed in one-electron Hilbert space as:

$$\begin{aligned} E\left[\sum_A \rho_{AA}\right] &= \sum_{ij}\left(t_{ij} + (\sum_C V_C^{ext})_{ij}\right)\left(\sum_A \rho_{AA}\right)_{ij} \\ &+ \frac{1}{2}\sum_{ijkl} V_{ijkl}\left(\left(\sum_A \rho_{AA}\right)_{ik}\left(\sum_A \rho_{AA}\right)_{jl} - \frac{1}{2}\left(\sum_A \rho_{AA}\right)_{il}\left(\sum_A \rho_{AA}\right)_{jk}\right), \end{aligned} \tag{2.50}$$

where t_{ij}, $(V_C^{ext})_{ij}$ and V_{ijkl} are defined in Eq. (2.38). We introduce $\rho'_{AB} = \rho_{AB} + \rho_{BA}$ as a shorthand notation for the bond contributions of the molecular

1DM, to define the individual bond matrix energies E_{AB} as [122]:

$$\begin{aligned} E_{AB} &= \sum_{ij}\left(t_{ij}+(\sum_{C}V_{C}^{ext})_{ij}\right)(\rho'_{AB})_{ij} \\ &+ \sum_{ijkl}V_{ijkl}\sum_{C}\left((\rho'_{AB})_{ik}(\rho_{CC})_{jl}-\frac{1}{2}(\rho'_{AB})_{il}(\rho_{CC})_{jk}\right) \\ &+ \frac{1}{2}\sum_{ijkl}V_{ijkl}\sum_{C<D}\left((\rho'_{AB})_{ik}(\rho'_{CD})_{jl}-\frac{1}{2}(\rho'_{AB})_{il}(\rho'_{CD})_{jk}\right). \quad (2.51) \end{aligned}$$

The bond matrix energy E_{AB} includes all interactions of ρ'_{AB} with the atomic contributions and half of the interactions with other bond contributions.

Table 2.5 displays the bond matrix energies $E_{AB}^{w,n}$ (in a.u.) for the small set of mainly diatomic molecules with B3LYP/cc-pVDZ geometries and a ROHF/Aug-cc-pVTZ 1DM. We also include the HOMO-LUMO gaps in these molecules and experimental atomization energies. The order of magnitude of these quantities can be compared to the bond energy. The Hirshfeld-I SEDI index is the bond order indicator. [77–86].

mol	$SEDI_{AB}$	ΔE_{at}^{exp}	E_{HOMO} -E_{LUMO}	E_{AB}^{w}	E_{AB}^{n}	f_{AB}^{w}	f_{AB}^{n}	$f_{A,B}^{w}$	$f_{A,B}^{n}$
H_2	1.00	-0.17	-0.64	-0.77	-0.80	-0.15	-0.16	-0.24	-0.25
F_2	1.58	-0.06	-0.75	-4.73	-1.89	-0.31	-0.26	-0.35	-0.41
Cl_2	1.75	-0.09	-0.47	-6.38	-1.59	-0.25	-0.21	-0.32	-0.36
Li_2	1.03	-0.04	-0.18	-0.72	-0.30	-0.06	-0.06	-0.09	-0.10
O_2	2.69	-0.19	-0.70	-7.91	-3.52	-0.55	-0.49	-0.71	-0.78
N_2	3.33	-0.36	-0.71	-9.66	-4.70	-0.67	-0.61	-0.97	-1.03
LiH	0.26	-0.09	-0.29	-0.80	-0.41	-0.04	-0.02	-0.04	-0.05
CO	2.84	-0.41	-0.62	-8.88	-4.09	-0.57	-0.49	-0.84	-0.90
LiF	0.26	-0.22	-0.47	-1.08	-0.48	-0.04	-0.03	-0.05	-0.07
NaCl	0.32	-0.15	-0.33	-1.85	-0.37	-0.04	-0.02	-0.05	-0.08
HO-H	0.88	-0.17	-0.54	-2.54	-1.09	-0.14	-0.12	-0.26	-0.29
HF	0.80	-0.22	-0.68	-2.72	-1.00	-0.16	-0.11	-0.23	-0.27
HCl	1.12	-0.16	-0.51	-4.04	-1.15	-0.18	-0.15	-0.27	-0.29

Table 2.5: Energy contributions of the bond matrix (in a.u.) for some diatomic molecules, with B3LYP/cc-pVDZ geometries and a ROHF/Aug-cc-pVTZ 1DM. The bond matrix was obtained using the weighted (w) and the nonweighted (n) variants of the Hirshfeld-I consistent double-index partitioning scheme. For comparison, HOMO-LUMO band gaps and experimental atomization energies E_{at}^{exp} from the NIST data base [125] are also indicated for these molecules.

The bond matrix energies are clearly too large to be applicable in chemistry, but they contain valuable information on the partitioning schemes. First, the bond matrix energies $E_{AB}^{n,w}$ could be thought to be related to the atomization energies. This is, however, not the case, as clearly comes out of the statistical analysis with a linear regression of only R^2 =0.52 for the nonweighted scheme, and R^2 =0.37 for the weighted scheme. Since the bond matrix implies a transfer of electrons from antibonding to bonding orbitals, its energy $E_{AB}^{w,n}$ is expected to be related to the energy difference between these orbitals, which in a crude estimate equals the HOMO-LUMO gap (roughly 0.5 a.u. for a single electron). However, since there is a net number of electrons in the bond matrices, $E_{AB}^{w,n}$ is also expected to scale with the net number of electrons in the bond matrix, and therefore with the SEDI bond order index. We do find indeed that the energy of the bond matrix is correlated with the product of the bond order index and the HOMO-LUMO gap. This is especially the case for the nonweighted scheme (with a linear regression of R^2 =0.96), but also for the weighted scheme (with a regression of R^2 =0.87). The E_{AB} therefore have a rough interpretation as the energy associated with promoting electrons from the bonding to the antibonding orbitals when a bond is broken (while keeping the molecular geometry). The bond energies from the weighted scheme are about twice as large as those obtained using the nonweighted scheme. This is not surprising, since for the former scheme, the (net) number of electrons in the bond matrix is also significantly larger.

In order to distinguish between the weighted and the nonweighted schemes, values for similar molecules are compared. Inspection of Table 2.5 reveals that bond matrix energies for Cl_2 and F_2 are large in comparison to the bond matrix energy of H_2, especially for the weighted scheme. The weighted scheme also produces quite high bond matrix energies for ionic species like NaCl and LiF (in the case of NaCl even surpassing the bond matrix energy of H_2). The appearance of core electron contributions in the bond matrix is the likely cause. For similar compounds, values obtained using the nonweighted scheme are also often (relatively) closer than values obtained using the weighted scheme. E.g. the relative difference between the values for LiH and LiF obtained using the weighted scheme $(|E_{LiH}^{w}\text{-}E_{LiF}^{w}|)/E_{LiF}^{w}$ =0.26, almost doubles that obtained using the nonweighted scheme $(|E_{LiH}^{n}\text{-}E_{LiF}^{n}|)/E_{LiF}^{n}$ =0.15. From these data, a slight preference for the nonweighted scheme emerges. But there still is room for improvement, since even for the nonweighted scheme the bond energies of the ionic species are too large in comparison to the value for H_2

Although the full bond energies are not directly useful in chemistry, it is possible to identify a contribution which seems to be a robust indicator of the covalent bond strength [122]. It is (minus) the Fock-exchange energy of the ρ_{AB}

bond matrix,

$$f_{AB}^{n,w} = \sum_{ijkl} V_{ijkl} \left(\frac{1}{2} (\rho_{AB}^{n,w})_{il} (\rho_{AB}^{n,w})_{jk} \right). \tag{2.52}$$

The values in both the weighted and nonweighted scheme are quite close and clearly single out the ionic species with a small value. In contrast to the SEDI bond order index, the value is also suppressed for the covalent but weakly bound Li_2. The values in the weighted and nonweighted scheme are closely correlated (R^2=0.994). A measure for the exchange energy between atoms A and B is also available using the single-atom density matrices ρ_A and ρ_B,

$$f_{A,B}^{n,w} = \sum_{ijkl} V_{ijkl} \left(\frac{1}{2} (\rho_{A}^{n,w})_{il} (\rho_{B}^{n,w})_{jk} \right). \tag{2.53}$$

It is remarkable that the obtained energy is again scheme independent (R^2=0.998 for the linear regression between weighted and nonweighted scheme) and even though the ingredients entering f_{AB} and $f_{A,B}$ (see Eqs. 2.52-2.53) are rather different, they basically contain the same information, e.g. the linear regression between f_{AB} and $f_{A,B}$ has an $R^2 = 0.983$ in both the nonweighted and weighted scheme.

3 Hilbert-Space partitioning of the molecular 1DM with orthogonal projectors

3.1 Requirements for an improved double-atom partitioning

3.1.1 Problems in the Hirshfeld-I consistent scheme

In the previous chapter, we introduced a density-matrix extension of the Hirshfeld-I atom-in-molecule model (see Sec. 2.2.1). We identified several paths that lead to similar single-atom 1DMs with a Hirshfeld-I density diagonal. These are obtained through a double-atom partitioning of the molecular 1DM over local atom and bond contributions, and a subsequent distribution of the bond matrices over the atoms. From the choice of a distribution mechanism for the bond contributions (weighted-nonweighted) and from the constraint of consistency with the Hirshfeld-I densities, atomic weight functions are constructed for that double-atom partitioning scheme.

We observed that the single-atom 1DMs are useful in practical applications, while the double-atom fragments are efficient instruments for the theoretical evaluation of the model. The double-atom fragments may clarify the characteristics of the single-atom 1DM fragments. E.g., the presence of small negative occupancies in the single-atom 1DMs, associated with antibonding orbitals, is readily understood from the analysis of the double-atom fragments (see Sec. 2.2.2). The double-atom fragments also help to establish the chemical (ir)relevance of a particular path for obtaining the single-atom 1DMs, e.g. through the analysis of the transferability of properties

associated with the bond matrices (see Sec. 2.3.2).

Practical use of these double-atom 1DMs (in applications) remains quite prohibitive. The energy components associated with the bond matrices are too large to be chemically intuitive. The atomic matrices are all electron deficient, making it difficult to interpret the state of the atoms in the molecule. Other problems are the misallocation of contributions from core electrons and free electron pairs (especially in the weighted scheme), too large bond matrix energies for the ionic compounds and the slow basis set convergence of the energy components in one-electron Hilbert space.

3.1.2 Requirements for an improved double-atom partitioning

In this section we attempt to improve the description of atoms and bonds in a molecule, so that the double-atom partitioning is applicable to chemical problems. We introduce three requirements:

1. **Orbital-based atomic weights instead of density-based atomic weights**

   Density-based atomic weight functions for fuzzy atoms, e.g. the Hirshfeld atomic weight functions (introduced in Sec. 1.3.2)),

   $$w_A^H(\boldsymbol{r}) \quad = \quad \frac{\rho_A^0(\boldsymbol{r})}{\sum_B \rho_B^0(\boldsymbol{r})} = \frac{\rho_A^0(\boldsymbol{r})}{\rho^0(\boldsymbol{r})} \tag{3.1}$$

   pick up contributions from core electrons and free electron pairs on the other atoms with their tails. In general, density-based approaches are not selective with respect to the orbitals involved in chemical binding. Bultinck and Cooper recently proposed an orbital specific extension [126] of the ISA approach (see Sec. 1.3.2). In this approach, the average density of the individual molecular orbitals $\varphi_k(\boldsymbol{r})$ with occupancies n_k over a sphere of radius $|\boldsymbol{r} - \boldsymbol{R}_A|$ is denoted $\langle \rho_{A,k}(\boldsymbol{r} - \boldsymbol{r_A}) \rangle$. This density is used to construct orbital-specific weight matrices $w_{A,i}(\boldsymbol{r})$,

   $$w_{A,k} = \frac{\langle \rho_{A,k}(\boldsymbol{r} - \boldsymbol{r_A}) \rangle}{\sum_B \langle \rho_{B,k}(\boldsymbol{r} - \boldsymbol{r_B}) \rangle} \tag{3.2}$$

   for the atomic decomposition of the electron density

   $$\rho = \sum_{A,k} w_{A,k} n_k \varphi_k^2(\boldsymbol{r}). \tag{3.3}$$

In a more general approach, selectivity for the individual one-electron functions can be obtained by introducing atomic weight kernels $w_A(\boldsymbol{r},\boldsymbol{r'})$:

$$w_A(\boldsymbol{r},\boldsymbol{r'}) = \sum_k \omega_{A,k}\phi_{A,k}(\boldsymbol{r})\phi_{A,k}(\boldsymbol{r'}). \tag{3.4}$$

The selectivity for the eigenvectors $\phi_{A,k}$ is expressed by the corresponding eigenvalues $\omega_{A,k}$. It is now possible to introduce orbital-selective atomic weight matrices reminiscent of the Hirshfeld philosophy, since the eigenvectors and eigenvalues of w_A can be defined from the populations and natural orbitals (i.e. the density matrices) of the isolated atoms. In practice, a 1DM-based alternative for Eq. (3.1) must fulfil the following constraint:

$$0 \leq \frac{w_A(\boldsymbol{r},\boldsymbol{r'})}{\sum_B w_B(\boldsymbol{r},\boldsymbol{r'})} \leq 1, \tag{3.5}$$

i.e. the eigenvalues of the w_A weight kernel must be between zero and one. This requirement restricts the possible expressions to only a few, including:

$$w_A(\boldsymbol{r},\boldsymbol{r'}) = \int d\boldsymbol{r''} \int d\boldsymbol{r'''} \left[\left(\rho^0\right)^{-\frac{1}{2}}(\boldsymbol{r},\boldsymbol{r''})\right] \rho^0_A(\boldsymbol{r''},\boldsymbol{r'''}) \left[\left(\rho^0\right)^{-\frac{1}{2}}(\boldsymbol{r'''},\boldsymbol{r'})\right], \tag{3.6}$$

where $\rho^0_A(\boldsymbol{r},\boldsymbol{r'})$ is the 1DM of the isolated atom A, and $\rho^0(\boldsymbol{r},\boldsymbol{r'})$ equals $\sum_B \rho^0_B(\boldsymbol{r},\boldsymbol{r'})$.

2. **Smooth convergence of the expressions in one-electron Hilbert space with respect to basis set size**

When the Hirshfeld atomic weights are expressed in a finite basis set (e.g. spanned by the natural orbitals $\varphi^0_i(\boldsymbol{r})$ of the promolecule $\rho^0(\boldsymbol{r},\boldsymbol{r'})$),

$$(w_A)_{ij} \quad = \quad \int d\boldsymbol{r}\varphi^0_i(\boldsymbol{r}) \left(\frac{\rho^0_A(\boldsymbol{r})}{\rho^0(\boldsymbol{r})}\right) \varphi^0_j(\boldsymbol{r}), \tag{3.7}$$

the complicated integrand is not optimally reproduced by the usual atomic basis sets. This affects the convergence properties of the Hirshfeld method in one-electron Hilbert space and hampers the application of the finite basis set expressions in e.g. energy decomposition schemes. For the kernel defined in Eq. (3.6), the corresponding finite basis set expression is:

$$(w_A)_{ij} \quad = \quad (n^0_i n^0_j)^{-\frac{1}{2}} \int d\boldsymbol{r} \int d\boldsymbol{r'} \varphi^0_i(\boldsymbol{r}) \left(\rho^0_A(\boldsymbol{r},\boldsymbol{r'})\right) \varphi^0_j(\boldsymbol{r'}), \tag{3.8}$$

where the n_i are populations of the natural orbitals $\varphi_i^0(\boldsymbol{r})$ of the pro-molecule. Standard basis sets can reproduce integrands such as that in Eq. (3.8) quite well [127–129], and therefore we expect a smooth convergence for these 1DM-based atomic weights in one-electron Hilbert space. Henceforth, for convenience in notation, the indices i and j (indicating that the quantities are expressed in the matrix representation) will be omitted.

3. **Interpretative simplicity for the double-atom scheme**

To enhance the interpretation of the double-atom fragments,

$$\rho = \sum_{AB} \rho_{AB} = \sum_{AB} w_A \rho w_B, \tag{3.9}$$

we recall that the bonding and antibonding orbitals in the bond matrix tend to have opposite populations when Baders binary and mutually exclusive QTAIM weight functions are used (see Sec. 2.2.2). This result is readily generalized to the matrix space: if the atomic weight matrices are chosen as projectors onto orthogonal subspaces, which implies that they are both idempotent and orthogonal,

$$w_A w_B = w_A \delta_{AB}, \tag{3.10}$$

then the bond matrices (ρ_{AB}) with $A \neq B$ have a zero trace and only the atomic matrices ρ_{AA} will contribute to the number of electrons N in the molecule

$$N = Tr(\rho) = \sum_{A} Tr(\rho_{AA}). \tag{3.11}$$

Eq. (3.11) enhances the interpretation and applicability of the "state" that corresponds to ρ_{AA}. It eliminates the practical obstacle of having to deal with ρ_{AA}matrices that are overall "electron deficient" (in a mathematical sense). E.g., if Eq. (3.11) is valid, then it is possible the use the populations of the $\rho_{AA}$to define the state of the isolated atoms in Eq. (3.6).

Note that in general the orthogonality of the weight matrices does not necessarily imply a zero overlap between atomic electron densities $\rho_{AA}(\boldsymbol{r},\boldsymbol{r})$ of different atoms in coordinate space, as occurs in e.g. QTAIM. Consider, e.g., an orthonormal basis spanned by two delocalized functions $\varphi_A(\boldsymbol{r})$ and $\varphi_B(\boldsymbol{r})$ as depicted in Fig. 3.1. The orthogonality constraint is trivially satisfied for the following atomic weight matrices, expressed in this basis:

$$w_A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \qquad w_B = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}. \tag{3.12}$$

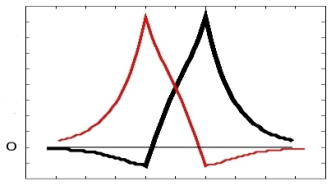


Figure 3.1: Basis of two orthonormal functions: $\varphi_A(\boldsymbol{r})$ (red) and $\varphi_B(\boldsymbol{r})$ (black)

Suppose the following "molecular" density matrix in this basis:

$$\rho = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{3.13}$$

It is clear that the atomic densities, e.g. defined as $\rho_A = \frac{1}{2}(w_A\rho + \rho w_A)$,

$$\rho_A(\boldsymbol{r}) = |\varphi_A(\boldsymbol{r})|^2 \qquad \rho_B(\boldsymbol{r}) = |\varphi_B(\boldsymbol{r})|^2 \tag{3.14}$$

have a significant overlap. Hence, the requirement that atomic weight matrices are chosen as projectors onto orthogonal subspaces does not exclude overlapping atomic regions and fuzzy atom definitions.

3.2 Double-atom partitioning of the molecular 1DM with orthogonal projectors in one-electron Hilbert space.

3.2.1 Introduction of the scheme

We found that, starting from the 1DM's $\rho_{AA}^{(0)}$ for isolated atoms, the following recursive scheme ($i = 0, 1, ...$)

$$\begin{aligned} \rho^{(i)} &= \sum_A \rho_{AA}^{(i)}\,; \quad w_A^{(i)} = (\rho^{(i)})^{-\frac{1}{2}} \rho_{AA}^{(i)} (\rho^{(i)})^{-\frac{1}{2}} \\ \rho_{AA}^{(i+1)} &= w_A^{(i)} \rho w_A^{(i)} \end{aligned} \tag{3.15}$$

converges and generates weight matrices $w_A^{(\infty)}$ that are by construction (i) chemically relevant, (ii) orbital-based, (iii) orthogonal projectors with (iv) good convergence properties with respect to basis set size (see **Papers 2 and 4**) [130],[131]. We elaborate on this in the Secs. 3.2.2 and 3.2.3. It is clear from Eq. (3.15) that the local double-atom contributions $\rho_{AA}^{(\infty)}$ themselves serve as a model for the partitioning of space. In contrast to what the iterative character of the procedure might suggest, the chosen starting point is crucial for several reasons:

1. In [131] it is shown that all sets of idempotent and orthogonal weight matrices may be the result of Eq. (3.15), including sets of orthogonal projectors that are delocalized and produce chemical nonsense. In practice, the iterative procedure only transforms the initial set of weight matrices into a similar set of orthogonal projectors.

2. It was also pointed out analytically that some non-idempotent sets of weight matrices may be the result of Eq. (3.15), but for the starting points chosen in [130] and [131] this was never observed.

3. When the 1DM of the isolated atom $\rho_{AA}^{(0)}$ contains only basis functions centered on atom A, the subspaces spanned by the eigenvectors of $\rho_{AA}^{(0)}$ are linearly independent, and as a result the weight matrices w_A^0 of the 0'th iteration already obey Eq. (3.10).

An efficient and chemically relevant set of starting points is provided by the 1DMs of the isolated atoms. As in ordinary Hirshfeld (see Sec. 1.3.2), different results are obtained depending on the charge and electronic state of the isolated atom. Following the ideas behind the iterative Hirshfeld procedure [44], the result can be made independent of the starting point by building in

self-consistency through an outer iterative scheme (where Eq. (3.15) would represent the inner iterative scheme).

In our simplest implementation (called **'charge equalization'**) we start from rotationally averaged 1DM's of the neutral isolated atoms. The recursive scheme in Eq. (3.15) generates effective electron numbers $N_A = Tr(\rho^{\infty}_{AA})$ for the atoms. These can be used to create a rotationally averaged 1DM of the charged isolated atom according to the linear interpolation between integer electron numbers $(k \leq N_A < k+1)$ [48, 49],

$$\begin{aligned} \rho^{(0)}_{AA}[N_A] &= (k+1-N_A)\,\rho^{(0)}_{AA}[k] \\ &+ (N_A-k)\,\rho^{(0)}_{AA}[k+1]. \end{aligned} \tag{3.16}$$

The charged $\rho^{(0)}_{AA}[N_A]$ can be used as the next starting point in Eq. (3.15) and the whole process is repeated until convergence for the effective electron numbers.

We noticed that, in contrast to the electron density, the 1DM is much more sensitive to a mismatch in the orientation of one-electron orbitals in the molecule and in the isolated atom used to set up the AIM. In some cases, this resulted in rather large AIM charges. In order to solve this problem we also implemented a more sophisticated scheme (called **'population equalization'**). In this scheme, the implementation of the previous paragraph is followed up to the point where $\rho_{AA}{}^{(0)}[N_A]$ is obtained. We then use the eigenvalue decompositions

$$\rho^{(\infty)}_{AA} = \sum_k n^{(\infty)}_{AA,k} \varphi^{(\infty)}_{AA,k} (\varphi^{(\infty)}_{AA,k})^T \tag{3.17}$$

$$\rho^{(0)}_{AA}[N_A] = \sum_l n^{(0)}_{AA,l} \varphi^{(0)}_{AA,l} (\varphi^{(0)}_{AA,l})^T \tag{3.18}$$

to generate the new starting point to Eq. (3.15). This is given by Eq. (3.18), but with the rotationally averaged occupation numbers $n^0_{AA,l}$ replaced with the occupation numbers $n^{(\infty)}_{AA,k}$ of the partitioned atom. The correspondence is made on the basis of maximal orbital overlap $|\langle \varphi_{AA,k}{}^{(\infty)} | \varphi_{AA,l}{}^{(0)} \rangle|$. By this procedure, the rotationally averaged 1DM's are replaced by 1DM's containing information on the molecular geometry. The whole process is again repeated until convergence.

3.2.2 Characteristics of the double-atom 1DM fragments

Chemically sound occupancies for AIM orbitals

The scheme introduced in Sec. 3.2.1 clearly benefits from the orbital-based character of the weight matrices. This is illustrated in Table 3.1 for an exemplary calculation on N_2. The eigenvalues of the atomic and bond matrices in the current approach are compared to those generated using Hirshfeld-I weight functions. Inspection of the eigenvalues of the atomic matrices $\rho_{N_1N_1}$ and $\rho_{N_2N_2}$ reveals that the current scheme generates occupancies of the orbitals not involved in bonding (core orbitals or orbitals containing free electron pairs) that are very close to their expected integer value 2. The corresponding occupancies are significantly lower when Hirshfeld-I weight functions are used for the double-atom partitioning (see Sec. 2.2).

	ρ_{N_1,N_1}	$\rho_{N_1,N_2}+\rho_{N_2,N_1}$	ρ_{N_1,N_1}	ρ_{N_1,N_2}	$+\rho_{N_2,N_1}$
(a)	2.000 (1s)	$\pm$ 1.000 (σ,σ^*)	1.977 (1s)	1.000 (σ)	-0.453 (σ^*)
(b)	1.995 (2s)	$\pm$ 1.000 (π_1,π_1^*)	1.866 (2s)	1.000 (π_1)	-0.600 (π_1^*)
(c)	1.000 (px)	$\pm$ 1.000 (π_2,π_2^*)	0.800 (px)	1.000 (π_2)	-0.600 (π_2^*)
(d)	1.000 (py)	$\pm$ 0.101	0.800 (py)	0.117	-0.119
(e)	1.000 (pz)	$\pm$ 0.101	0.706 (pz)	0.085	-0.010
(f)	0.010	$\pm$ 0.008	0.013	0.015	-0.002
(g)	0.000	$\pm$ 0.008	0.000	0.013	-0.001
(...)	$< 10^{-13}$	$-10^{-13} >< 10^{-13}$	$< 10^{-13}$	$< 10^{-13}$	$-10^{-13} >$
Sum	7.000	$\pm$ 3.218	6.162	3.460	-1.783

Table 3.1: All natural populations in the atomic density matrices ($\rho_{N_1N_1} = \rho_{N_2N_2}$) and eigenvalues of the bond matrix ($\rho_{N_1N_2} + \rho_{N_2N_1}$) for molecular N_2. The geometry was optimized at the B3LYP/cc-pVDZ level and the 1DM was calculated at the ROHF/Aug-cc-pVDZ level. The left side of the table shows the results for the approach in the current chapter (using the population equalization scheme), the right side of the table shows the results for the double-atom partitioning in 3D space based on the use of Hirshfeld-I weights.

The new scheme also benefits from the orthogonal projector property for the weight matrices. The double occupied molecular σ and π orbitals simply divide their occupancy over the atomic p_z, p_x and p_y orbitals and no charges appear on the atoms. In the bond matrix $\rho_{N_1N_2} + \rho_{N_2N_1}$, the bonding and antibonding pairs of orbitals have exactly opposite eigenvalues, representing an electronic deformation (see discussion on appearance of negative eigenvalues in Sec. 2.2.2). In contrast, when Hirshfeld-I weight functions are used for the

double-atom partitioning, large positive charges are generated on the nitrogen atoms ($Q_A = +0.84$) and a net bond density blurs the interpretation of the bond matrix.

Clear trends and similarities for the energy components of the bond matrix

As discussed in Sec. 2.3.2, the energy E_{AB} (see Eq. (2.51)) associated with the bond matrix ($\rho_{AB} + \rho_{BA}$) is indicative for the quality of the 1DM fragments. Table 3.2 list the bond energies obtained with the present scheme for some diatomic molecules with a B3LYP/cc-pVDZ geometry and a 1DM calculated at the ROHF/Aug-cc-pVTZ level of theory. Values clearly benefit from the orbital-based character of the weight matrices. The bond matrix energies of the ionic species LiH, LiF and NaCl (-0.05, -0.07 and -0.03 a.u.) are almost an order of magnitude smaller in absolute value than those obtained using the Hirshfeld-I weights (-0.41, -0.48, and -0.37 a.u). For the weakly bonded Li_2, the bond energy remains significantly smaller (-0.18 a.u.) than for the other compounds with a single covalent bond. The range for the E_{AB} values of the other compounds with a single covalent bond is more reasonable for the new scheme ([-0.78,-0.93] a.u.) than for the approach based on Hirshfeld-I weight functions ([-0.80,-1.89] a.u.), all the more since in the second approach the rather wide range is demarcated by the chemically similar compounds H_2 and F_2.

	E_{AB}	f_{AB}		E_{AB}	f_{AB}
H_2	-0.88 (-0.80)	-0.24 (-0.16)	LiH	-0.05 (-0.41)	-0.01 (-0.02)
F_2	-0.88 (-1.89)	-0.25 (-0.26)	LiF	-0.07 (-0.48)	-0.01 (-0.03)
Cl_2	-0.74 (-1.59)	-0.20 (-0.21)	NaCl	-0.03 (-0.37)	0.01 (-0.02)
Li_2	-0.18 (-0.30)	-0.09 (-0.06)	HF	-0.78 (-1.00)	-0.16 (-0.11)
O_2	-2.28 (-3.52)	-0.56 (-0.49)	CO	-2.99 (-4.09)	-0.62 (-0.49)
N_2	-3.73 (-4.70)	-0.85 (-0.61)	HCl	-0.92 (-1.15)	-0.22 (-0.15)
S_2	-1.45 (-2.52)	-0.39 (-0.34)	H-OH	-0.93 (-1.09)	-0.18 (-0.12)
P_2	-1.86 (-2.79)	-0.52 (-0.39)			

Table 3.2: Bond matrix energy E_{AB} and the energetic covalency measures f_{AB} (in Hartree) for some diatomic molecules with a B3LYP/cc-pVDZ geometry and a 1DM calculated at the ROHF/Aug-cc-pVTZ level of theory. Non bracketed values correspond to the present scheme based on orthogonal projectors (using the population equalization scheme), bracketed values correspond to the nonweighted variant of the Hirshfeld-I consistent scheme.

For the homonuclear diatomics S_2 and P_2, the E_{AB} values are suppressed in comparison to their second row counterparts O_2 and N_2. This is in line

with the less efficient overlap of the atomic valence orbitals of the third row elements. Note that Table 3.2 also lists plausible values for the energetic covalency indicator f_{AB} defined in Eq. (2.52).

Reasonable charges on the atoms

The atomic charges generated by the present approach show a strong linear correlation with charges from the Hirshfeld-I partitioning of the electron density, which is a reliable AIM approach. This is illustrated in Fig. 3.2 for the diverse set of 67 molecules used in [130]. Note that two different implementations of the present approach were proposed in Sec. 3.2.1: (i) an implementation satisfying a self-consistency requirement for the atomic charges ("charge equalization") and (ii) an implementation satisfying a self-consistency requirement for the individual populations of the atomic orbitals ("population equalization"). The linear correlation is more satisfying for the

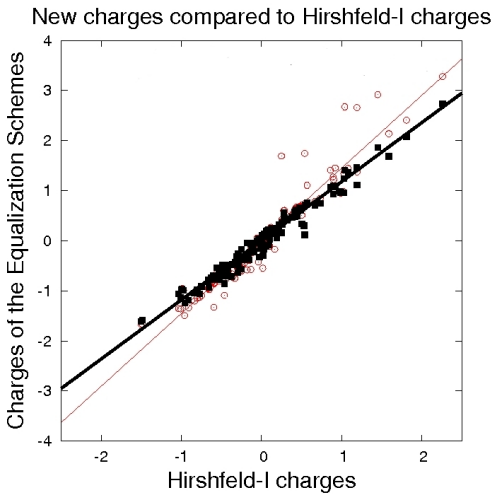


Figure 3.2: Comparison of the Hirshfeld-I charges (at the ROHF/Aug-cc-pVDZ level) with those generated by present approach using the charge equalization scheme (red circles) and the population equalization scheme (black squares). For details, see text.

"population equalization" scheme (R^2=0.96, slope=1.18) than for the "charge equalization" scheme (R^2=0.90, slope = 1.45). This is remarkable since the Hirshfeld-I scheme is itself based on a self-consistency requirement for the atomic charges. In some cases, the "charge equalization" scheme leads to

unexpectedly high charges. This is related to the inadequacy of the rotationally averaged 1DM's used as starting point in the charge equalization scheme. In CF_4, e.g., the procedure of rotational averaging transfers electrons from the (fully occupied) free electron pairs in F to the p_z orbital that is used for bonding. Since the atomic weight matrices are based on a "stockholder approach" (Eq. 3.15), this results in an artificially increased share of F for the bonding orbitals and thus in an exaggerated negative charge on the F-atoms. This problem seems to be solved using the "population equalization" procedure, which is more adapted to the molecular geometry. In the latter scheme, the atomic orbitals that are empty in the molecule (e.g. the p_z orbital on B in BH3) and orbitals that are occupied by lone pairs (e.g. the p_x and p_y orbitals on F in CF4) keep their specific populations in the starting points of the next iteration. This type of consistency is a clear advantage for fuzzy atoms defined by a stockholder approach. It clearly requires the use of orbital-based weight matrices in the 1DM partitioning scheme.

3.2.3 Orthogonal projectors

We already noted that the iterative procedure in Sec. 3.2.1 generates a set of orthogonal projectors $w_A^{(\infty)}$. This important result is not trivial and is proven rigorously in **Paper 4**. The analytical proof requires that the recursive scheme of Eq. (3.15) is generalized and rewritten more compactly,

$$\begin{aligned} w_A^{(i)} &= (\rho^{(i)})^{-\frac{1}{2}} w_A^{(i-1)} M w_A^{(i-1)} (\rho^{(i)})^{-\frac{1}{2}} \\ \rho^{(i)} &= \sum_A w_A^{(i-1)} M w_A^{(i-1)} \end{aligned} \tag{3.19}$$

where M is in general any positive-definite matrix. To obtain a chemically meaningful decomposition, we found that M can be chosen as either the full molecular 1DM itself ($M = \rho$ as in Eq.(3.15)), or the recursive estimate for $\rho^{(\infty)}$ (with $M = \rho^{(i-1)}$ in Eq.(3.19)). It is assumed that the matrices in Eq. (3.19) do not have a nullspace. In practice, zero eigenvalues can be substituted by a small value (e.g. 10^{-05}) without affecting results. For the generalized recursive scheme (Eq. (3.19)), we proved three lemmas [131]:

Lemma 3.2.1 *The converged weight matrices w_A^{∞} are idempotent when $M = \rho^{(i-1)}$ in Eq. (3.19).*

Lemma 3.2.2 *If the converged weight matrices w_A^{∞} are idempotent, then they are orthogonal.*

Lemma 3.2.3 *When idempotent weight matrices enter the recursion formula in Eq. (3.19), immediate convergence is obtained.*

When $M = \rho^{(i-1)}$ in Eq. (3.19), it is clear from lemma 1 and 2 that every set of (converged) weight matrices obtained with the recursive scheme is an exact set of orthogonal projectors. Lemma 3 implies that the chemical information should be contained in the starting points of the iterative scheme, since the iterative procedure would not alter a set of orthogonal projectors that is chemically irrelevant.

When $M = \rho$ in Eq. (3.19), the converged weight matrices are not necessarily idempotent. A trivial solution of the recursive scheme is provided by identical weight matrices for all N_A atoms in the molecule: $w_A^{(\infty)} = (1/N_A)I$. When bad starting points are chosen for the iterative scheme, we observed that the scheme has bad convergence properties and that eigenvalues 0 and 1/k appear in the weight matrices (where k is equal to or lower than N_A). However, when the starting points are close to idempotency, lemma 3 implies that a set of orthogonal projectors is a likely outcome of the recursive scheme. For the starting points based on the 1DMs of the isolated atoms, we observed that the recursion of Eq. (3.19) always transforms the initial weight matrices into perfect orthogonal projectors. This is illustrated in Fig. 3.3 for a set of 67 simple molecules with a singlet ground state [131]. The histogram displays the degree to which the weight matrices from the recursive scheme satisfy the relationship associated with orthogonal projectors (Eq. (3.10)). For each molecule in the entire test set, all atom pairs AB contribute to a matrix C_{AB} that reduces to the zero matrix for perfect orthogonal projectors:

$$C_{AB}^{(1)} = w_A^{(\infty)} w_B^{(\infty)} - \delta_{AB} w_A^{(\infty)}. \tag{3.20}$$

The matrix element with a maximal deviation from zero in C_{AB}, for all atom pairs, is chosen to represent the accuracy attained for that molecule. A histogram is created from the integer logarithms of these accuracy measures.

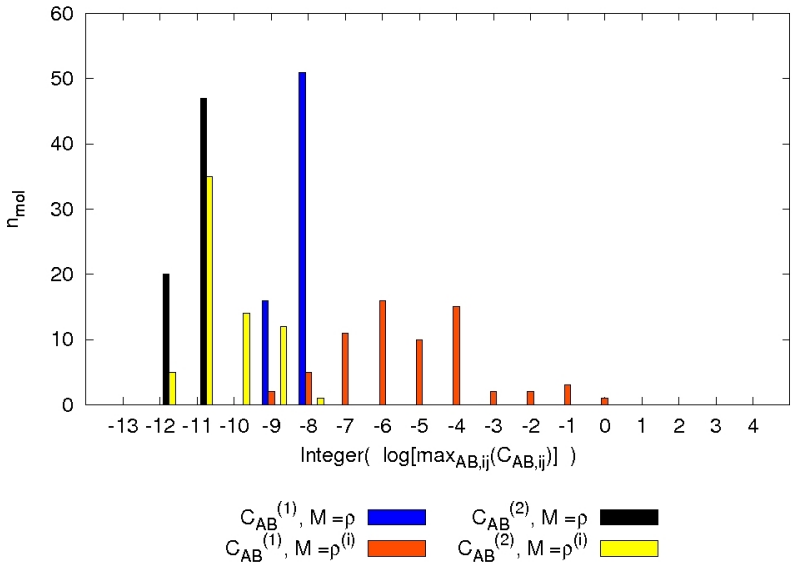


Figure 3.3: Attained numerical accuracy in obtaining weight matrices that are orthogonal projectors, through the use of the recursive scheme.

When $M = \rho$ in Eq. (3.19), we indeed find that the relationship associated with orthogonal projectors is extremely well fulfilled by the weight matrices from the recursive scheme (the largest nonzero element of the C_{AB} matrices being smaller than 10^{-8}). For most molecules, a less accuracy is attained when $M = \rho^{(i)}$, which is seemingly in conflict with the exact analytical proof that was provided for that scheme. The analytical proof assumes, however, the absence of nullspace issues. To obtain the results in Fig. 3.3, we only eliminated the nullspace of the molecular 1DM (by replacing zero eigenvalues with a small value, e.g. 10^{-5}) , while a repeated elimination of the nullspace of $\rho^{(i)}$ is required to make the numerical data coincide with the analytically derived predictions. We note that this nullspace issue hardly affects the atomic 1DM fragments. Multiplying Eq. (3.20) on both sides with $(\rho^{(\infty)})^{\frac{1}{2}}$, one obtains:

$$C^{(2)}_{AB} = \rho^{(\infty)}_{AA}(\rho^{(\infty)})^{-\frac{1}{2}}\rho^{(\infty)}_{BB} - \delta_{AB}\rho^{(\infty)}_{AA}, \tag{3.21}$$

where $\rho^{(\infty)}_{AA}$ is a shorthand notation for $w^{(\infty)}_A M w^{(\infty)}_A$. Expression 3.21 is an equivalent orthogonal-projector relationship, but is restricted to the relevant or "occupied" part of space. It is clear from Fig. 3.3 that the weight matrices

from both recursive schemes satisfy these relationships almost perfectly (the largest nonzero element of the C_{AB} matrices being consistently smaller than 10^{-8}).

From the analytical and numerical data presented in this section, we conclude that the iterative procedure introduced in Sec. 3.2.1 generates a set of chemically relevant orthogonal projectors $w_A^{(\infty)}$ when the 1DMs of the isolated atoms are chosen as a starting point for the scheme. It requires, however, that the 1DMs are positive definite. This implies that in MP2 and CCSD approaches, the small (unphysical!) negative eigenvalues that might appear in the 1DM (e.g. for larger bond lengths) must be neglected.

3.2.4 Types of atoms based on the use of orthogonal projectors

"Non-bonded" atoms in a molecular environment

The weight matrices $w_A^{(\infty)}$ from the iterative procedure in Sec. 3.2.1 are orthogonal projectors onto one-electron subspaces (see Sec. 3.2.3). This property guarantees that the atomic 1DMs from the double-atom fragmentation of the molecular 1DM,

$$\rho_{AA}^{(\infty)} = w_A^{(\infty)} \rho w_A^{(\infty)}, \tag{3.22}$$

generate the full number of electrons N in the molecule (see Eq. 3.11), and that no electrons are shared between different atoms:

$$Tr(\rho_{AA}^{(\infty)} \rho_{BB}^{(\infty)}) = Tr\left(\left(\rho_{AA}^{(\infty)}\right)^2\right) \delta_{AB}. \tag{3.23}$$

Since the sharing of electrons between the atoms in a molecule is identified with the process of chemical bonding, Eq. (3.23) is the basis for the interpretation of $\rho_{AA}^{(\infty)}$ as the 1DM of a non-bonded atom in a molecular environment. Note that this interpretation holds here for the normal molecular geometry, and has nothing to do with dissociation processes. Fig. 3.4 shows that the electron

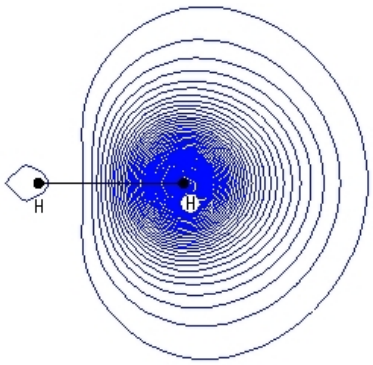


Figure 3.4: Contour plot for the electron density $\rho_{HH}^{(\infty)}(\boldsymbol{r})$ of a "non-bonded" hydrogen atom in a H_2 molecule with a B3LYP/cc-pVDZ geometry and a 1DM calculated at the ROHF/Aug-cc-pVTZ level. The outer contour has a value 0.01 (in atomic units) and the difference between two neighboring contours is 0.01.

density of a "non-bonded" atom is well localized but deformed by the presence of the other atoms.

The corresponding "bonded" atoms

The bond matrices do not hold a net electron density when the atomic weight matrices are orthogonal projectors (see Sec. 3.1.2). These matrices are associated with the deformation of the electron densities of the "non-bonded" atoms towards the bonding region, in order to allow a sharing of electrons. From this perspective, the 1DMs $\rho_{AA}^{(\infty)}$ of the "non-bonded" atoms are naturally associated with a particular set of single-atom 1DMs $\rho_A^{(\infty)}$:

$$\rho_A^{(\infty)} = \rho^{\frac{1}{2}} w_A^{(\infty)} \rho^{\frac{1}{2}}, \tag{3.24}$$

because the traces and the individual populations of both sets of atomic 1DMs coincide. The equality of the traces follows from the cyclic property of the trace and the idempotency of the weight matrices. The analytical proof for the equality of the individual populations is less trivial (see Ref. [131]). Both sets of atomic 1DMs only differ by the shape of the corresponding orbitals. While the orbitals of the "non-bonded atoms" seem to avoid the bonding region, the naturals of $\rho_A^{(\infty)}$are deformed towards the bonding region and resemble molecular orbitals. Fig. 3.5 shows that the electron density of an atom defined by Eq. (3.24) is heavily delocalized. Note that the expression in Eq. (3.24)

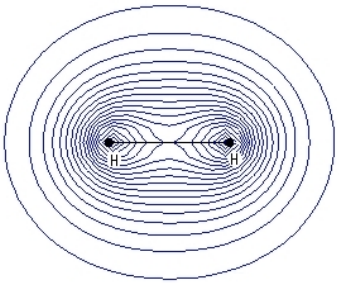


Figure 3.5: Contour plot for the electron density $\rho_H^{(\infty)}(\boldsymbol{r})$ of a "bonded" hydrogen atom in a H_2 molecule with a B3LYP/cc-pVDZ geometry and a 1DM calculated at the ROHF/Aug-cc-pVTZ level. The outer contour has a value 0.01 (in atomic units) and the difference between two neighboring contours is 0.01.

reduces to the expression introduced by Alcoba et al. [112] (see Eq. 2.12) when QTAIM weight functions are used. In Sec. 2.1.3 we already discussed the localization properties of the 1DM and the electron density based on this scheme. The same reasoning applies to Eq. (3.24) (apart from the fact that the

weight matrices are now nondiagonal in $\boldsymbol{r}$-space), and unsurprisingly $\rho_A^{(\infty)}$in Eq. (3.24) has the same "bad" localization properties. Henceforth, we refer to $\rho_A^{(\infty)}$as the 1DM of the bonded atoms, since these delocalized single-atom fragments of the molecular 1DM share electrons to an extent determined by the overlap of the bond matrices:

$$Tr(\rho_A^{(\infty)}\rho_B^{(\infty)}) = \frac{1}{2}Tr\left(\rho_{AB}^{(\infty)}\right)^2. \tag{3.25}$$

Combining Eqs. (3.15) and (3.24), one observes that the idempotent weight matrices $w_A^{(\infty)}$can be expressed in terms of both the localized "non-bonded" atom $\rho_{AA}^{(\infty)}$and the nonlocalized "bonded" atoms $\rho_A^{(\infty)}$:

$$w_A^{(\infty)} = (\rho^{(\infty)})^{-\frac{1}{2}}\rho_{AA}^{(\infty)}(\rho^{(\infty)})^{-\frac{1}{2}} = \rho^{-\frac{1}{2}}\rho_A^{(\infty)}\rho^{-\frac{1}{2}}. \tag{3.26}$$

This indicates that it is a nontrivial issue whether the binary Hilbert partitioning of Hilbert space should be associated with "hard" or "fuzzy" boundaries in coordinate space. Examining the eigenfunctions spanning the projector $w_A^{(\infty)}$, we notice that they have small contributions in the vicinity of the other nuclei. Fig. 3.6 shows an example of the density of (coordinate) space $w_A^{(\infty)}(\boldsymbol{r},\boldsymbol{r})$

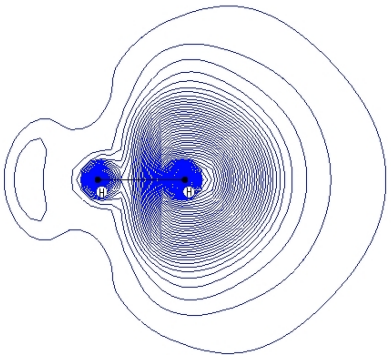


Figure 3.6: Contour plot for the density of the atomic weight matrix $w_A^{(\infty)}(\boldsymbol{r},\boldsymbol{r}')$ of hydrogen in a H_2 molecule with a B3LYP/cc-pVDZ geometry and a 1DM calculated at the ROHF/Aug-cc-pVTZ level. The outer contour has a value 0.05 (in atomic units) and the difference between two neighboring contours is 0.05.

assigned to an atom. It reaches a maximum at the positions of its own nucleus and a smaller one at the position of the other nucleus ($w_A^{(\infty)}(\boldsymbol{r},\boldsymbol{r})$=4.2. and 1.5 respectively). [1]

[1]Note that in a finite basis set, the unity matrix $I(\boldsymbol{r},\boldsymbol{r}) = \sum_A w_A(\boldsymbol{r},\boldsymbol{r})$ does not have value of 1 for every point in r-space. The value of $I(\boldsymbol{r},\boldsymbol{r})$ increases with the number of basis functions.

Alternative "bonded" atoms with a local density

The "bonded" single-atom 1DMs defined in Eq. (3.24) are positive semidefinite by construction, but have a delocalized electron density (see Sec. (2.1.3)) that makes them useless in the energy decomposition scheme introduced in Sec. 2.3.1. For the latter purpose, one might consider an alternative expression for the single-index atoms that is similar to the expression used for the nonweighted variant of the Hirshfeld-I consistent scheme in Sec. 2.2.1:

$$\rho_A^{\prime(\infty)} = \frac{1}{2}\left(w_A^{(\infty)}\rho + \rho w_A^{(\infty)}\right). \tag{3.27}$$

In Eq. (3.27), the single-atom 1DMs are constructed from the double-atom 1DMs by partitioning the bond matrices half-half over the atoms. However, the weight matrices are not diagonal in $\boldsymbol{r}$-space and the electron density,

$$\rho_A^{\prime(\infty)}(\boldsymbol{r}) = \sum_i n_i \varphi_i(\boldsymbol{r}) \int d\boldsymbol{r}'' \varphi_i(\boldsymbol{r}'') \left(w_A(\boldsymbol{r}, \boldsymbol{r}'')\right), \tag{3.28}$$

expressed in terms of the populations n_i and natural orbitals $\varphi_i(\boldsymbol{r})$ of the molecule, is not necessarily positive semidefinite. This is also clear from Fig.

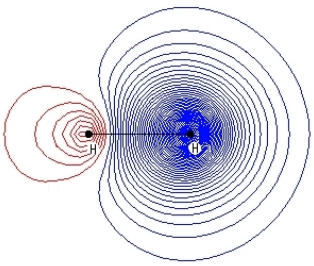


Figure 3.7: Contour plot for the electron density $\rho_H^{\prime(\infty)}(\boldsymbol{r})$ of an alternative "bonded" hydrogen atom in a H_2 molecule with a B3LYP/cc-pVDZ geometry and a 1DM calculated at the ROHF/Aug-cc-pVTZ level. Blue and red contours represent respectively positive and negative values. The outer contour has a value of $\pm$ 0.01 (in atomic units) and the difference between two neighboring contours is 0.01.

3.7. For an exemplary calculation on N_2 at the ROHF/Aug-cc-pVTZ level of theory, we observe that the interaction energy (as defined in Eq. (2.37)) between the atomic fragments in Eq. (3.27) is no longer attractive. As we failed to obtain any reasonable result with the fragments in Eq. (3.27), we did not investigate this expression for the single-index atoms any further.

It appears to be crucial to require a positive semidefinite character for the single-atom electron densities. Note that there are two easy ways to fulfil this requirement: (i) Either the single-atom 1DMs are positive semidefinite, as in Eq (3.24). However, we fail to combine that property with the locality requirement for the atomic densities. (ii) Another possibility is to use weight matrices that are diagonal in $\boldsymbol{r}$-space in Eq. (3.27). In this case, the constraint that the weight matrices must be orthogonal projectors (see Eq. (3.10)) reduces to the restriction that they are binary in $\boldsymbol{r}$-space. The situation is then very much like the QTAIM framework, and the atomic densities in Eq. (3.27) cannot be reconciled with a fuzzy atom picture (like Hirshfeld-I).

3.3 Applications

3.3.1 Bond orders

A quantum mechanical indicator for the covalent bond order is provided by the number of electron pairs shared between atoms. In the present framework, this is expressed as:

$$B'_{AB} = \frac{1}{2} Tr(\rho_A^{(\infty)} \rho_B^{(\infty)}) = \frac{1}{4} Tr\left(\rho_{AB}^{(\infty)}\right)^2. \tag{3.29}$$

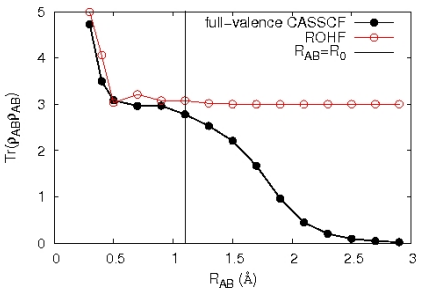
At the Hartree-Fock level, B'_{AB} can be considered as an extension of the Hartree-Fock SEDI-index B_{AB} (see Eq. (1.63)), where now weight kernels $w_A(\boldsymbol{r}, \boldsymbol{r}')$ are used instead of weight functions $w_A(\boldsymbol{r})$. It is clear from Table 3.3 that the indices for nonpolar bonds are much closer to the classically expected bond orders if they are constructed from the (orbital-based) orthogonal projectors ${w_A}^{(\infty)}$ than if they are constructed from the (density-based) Hirshfeld-I weight functions ${w_A}^{HI}$. The indices based on ${w_A}^{(\infty)}$ are also much more suppressed for ionic compounds, and thus more in line with the chemical notion of ionicity.

	ROHF	ROHF	CASSCF		ROHF	ROHF	CASSCF
	$B_{AB}[w_A^{HI}]$	$B'_{AB}[w_A^{(\infty)}]$	$B'_{AB}[w_A^{(\infty)}]$		$B_{AB}[w_A^{HI}]$	$B'_{AB}[w_A^{(\infty)}]$	$B'_{AB}[w_A^{(\infty)}]$
H_2	1.00	1.00	0.95	LiH	0.26	0.08	0.13
F_2	1.58	1.01	0.77	LiF	0.26	0.05	0.08
Cl_2	1.75	1.04	0.93	NaCl	0.32	0.03	0.05
Li_2	1.03	1.00	0.79	HF	0.80	0.48	0.54
O_2	2.69	2.02	1.35	CO	2.84	2.15	2.15
N_2	3.33	3.02	2.78	HCl	1.12	0.91	0.87
S_2	2.77	2.04	1.44	H_2O	0.88	0.65	
P_2	3.29	3.01	2.64				

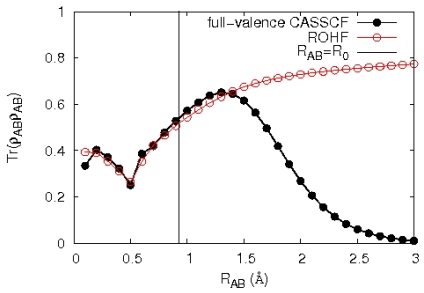
Table 3.3: Comparison of different indicators for the covalent bond order, for some small molecules with a singlet state and a B3LYP/cc-pVDZ geometry. The first column contains the well established iterative-Hirshfeld SEDI-indices at the ROHF/Aug-cc-pVDZ level. The second and third columns contain values for B'_{AB} at the ROHF/Aug-cc-pVDZ level and the CASSCF/Aug-cc-pVDZ level respectively.

At the post-Hartree-Fock level, the number of electron pairs shared between the 1DMs ${\rho_A}^{(\infty)}$of the bonded atoms is -in general- lower than in Hartree-Fock. This can be roughly understood by considering that previously unoccupied orbitals in these atoms now have a small occupation, but contribute less efficiently to the bond order index than the previously fully

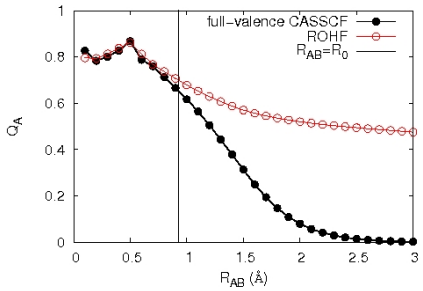
occupied orbitals (with e.g. σ and π shapes) whose occupation is now somewhat reduced. For the polar bonds, it is observed that some bond order indices are larger with respect to the Hartree-Fock case. This is related to the less polar character of the bonds in these compounds at the correlated level (smaller atomic charges). Calculations at the post-Hartree-Fock level are required to obtain reasonable indices for the covalent bond orders of molecules undergoing bond breaking and formation. As illustrated in Fig. 3.8 for N_2 and HF, Eq. (3.29) has the correct dissociation behavior at the full-valence CASSCF level, whereas the bond order index incorrectly evolves towards a constant nonzero value in the Hartree-Fock dissociation limit. As can be deduced from the right-hand side of Eq. (3.29), this behavior is linked to the dissociation behavior of the bond matrix (see the discussion following Fig. 2.4).



(a) Covalent bond order in N_2



(b) Covalent bond order in HF



(c) Charge on the H-atom in HF

Figure 3.8: Covalent bond order indices B'_{AB} in the N_2 and HF molecules for bond lengths R_{AB} varying from the equilibrium (B3LYP/Aug-cc-pVDZ) bond length R_0. The covalent bond order index was determined at the ROHF/Aug-cc-pVDZ and full-valence CASSCF/Aug-cc-pVDZ levels of theory.

The deformation matrices $\Delta\rho_A^{(\infty)}$that transform the "non-bonded" atoms $\rho_{AA}^{(\infty)}$ into the corresponding "bonded" atoms $\rho_A^{(\infty)}$,

$$\Delta\rho_A^{(\infty)} \quad = \quad \rho_A^{(\infty)} - \rho_{AA}^{(\infty)}, \tag{3.30}$$

are traceless and in addition may be considered to be atom-condensed bond matrices:

$$\sum_A \Delta\rho_A^{(\infty)} \quad = \quad \sum_{A,B\neq A} \rho_{AB}^{(\infty)}. \tag{3.31}$$

Using these atom-condensed bond matrices, it is possible to define the contribution of atom A to the covalent AB bond order as

$$B'_{A_{(B)}} = \frac{1}{2} Tr\left(\Delta\rho_A^{(\infty)}\rho_{AB}^{(\infty)}\right). \tag{3.32}$$

In case of a diatomic, it is easy to see (using the orthogonal projector relationship) that

$$B'_{AB} = \frac{1}{4} Tr\left(\rho_{AB}^{(\infty)}\rho_{AB}^{(\infty)}\right) = \frac{1}{2} Tr\left((\Delta\rho_A^{(\infty)} + \Delta\rho_B^{(\infty)})\rho_{AB}^{(\infty)}\right) = B'_{A_{(B)}} + B'_{B_{(A)}}. \tag{3.33}$$

Numerically, we found that Eq. (3.33) is also satisfied in the general case. From Eq. (3.32), it is possible to assess the share of each atom in the creation of the bond.

	Q_A	Q_B	B'_{AB}	$\frac{B'_{A_{(B)}}}{B'_{AB}}$	$\frac{\widetilde{B'_{A_{(B)}}}}{B'_{AB}}$	$\frac{B'_{B_{(A)}}}{B'_{AB}}$	$\frac{\widetilde{B'_{B_{(A)}}}}{B'_{AB}}$
LiH	0.91	-0.91	0.13	0.04	0.04	0.96	0.96
LiF	0.95	-0.95	0.08	0.04	0.02	0.96	0.98
NaCl	0.96	-0.96	0.05	0.02	0.02	0.98	0.98
HF	0.65	-0.65	0.54	0.21	0.18	0.79	0.83
CO	0.40	-0.40	2.15	0.32	0.27	0.68	0.73
HCl	0.31	-0.31	0.87	0.36	0.35	0.64	0.66

Table 3.4: Atomic charges (Q_A), indicators for the covalent bond order (B'_{AB}), fraction of the bond order condensed to atoms ($B'_{A_{(B)}}/B'_{AB}$)and a simple approximation to these fractions ($\widetilde{B'_{A_{(B)}}}/B'_{AB}$). Calculations were performed at the Full-valence CASSCF/Aug-cc-pVDZ level of theory for some polar diatomic molecules with a B3LYP/cc-pVDZ geometry.

Classical chemistry distinguishes between normal covalent bonding and dative covalent bonding when a bond is formed between neutral atoms. In a

normal covalent bond, each atom has an equal share to the bond order. In a dative covalent bond, only one atom provides the electrons for bonding. In the AIM framework, a bond is formed between partially charged atoms and covalent bonds display a continuous spectrum between "normal" and "dative" bonding. From the splitting of the bond order in Eq. (3.33), it is then possible to assess the "normal" and "dative" character of the covalent bond. Clearly the numbers in Table 3.4 indicate that the 1DMs of the "bonded" atoms ($\rho_A^{(\infty)}$) do not contribute equally to the covalent bond order index in Eq. (3.29). The contribution is mainly determined by the atomic charge and can be approximated (in the diatomic case) with the simple formula:

$$\widetilde{B'_{A_{(B)}}} = \frac{1}{2}\left((b.o.)_{AB} - Q_A + FC_A \right) B'_{AB}. \tag{3.34}$$

where $(b.o.)_{AB}$ is the classical number of bonds between the atoms (covalent or ionic), Q_A is the atomic charge determined from the AIM and FC_A is the formal charge.

In Table 3.5, the different indicators for the covalent bond order are compared for some polyatomic molecules at the Hartee-Fock level. It appears that the covalent bond order indicator that is based on orthogonal weight matrices ($B'_{AB}[w_A^{(\infty)}]$) has significantly lower values for the non-bonded interactions than the Hirschfeld-I SEDI index ($B_{AB}[w_A^{HI}]$). For atoms that are attached to the same center, values for the SEDI index are rather large (typically 0.05-0.20), while values for B'_{AB} are much smaller (0.01-0.06). However, for the $O\cdots O$ interaction in CO_2 and HCOOH, both indicators generate a large value. For chemically bonded atom pairs, the values generated by both indicators are clearly transferable between different molecules. Note that the number of electron pairs shared in a single bond is often significantly larger than 1.00 when the prediction is obtained from the SEDI index (e.g. 1.23 for C-C in CH_3CH_3 and 1.42 for C-O in CH_3OH). This is not the case for the new bond order indicator B'_{AB} (the corresponding values are 1.03 and 0.90 respectively).

		B_{AB}	B'_{AB}			B_{AB}	B'_{AB}
CH_4	C-H	0.99	0.95	HCN	C-N	3.11	2.96
	H$\cdots$H(*)	0.07	0.01		C-H	0.94	0.93
CH_3CH_3	C-C	1.23	1.03		N$\cdots$H	0.15	0.02
	C-H	0.95	0.95	CH_3OH	C-O	1.42	0.90
	C$\cdots$H	0.10	0.01		C-H	0.94	0.95
	H$\cdots$H(*)	0.07	0.02		C-H	0.93	0.94
	H$\cdots$H	0.01	0.01		O-H	0.86	0.64
CH_2CH_2	C-C	2.05	2.02		O$\cdots$H	0.13	0.04
	C-H	0.96	0.94		O$\cdots$H	0.13	0.02
	C$\cdots$H	0.15	0.01		C$\cdots$H	0.07	0.01
	H$\cdots$H(*)	0.06	0.01		H$\cdots$H	0.01	0.01
	H$\cdots$H	0.01	0.02	H_2O	O-H	0.88	0.64
	H$\cdots$H	0.01	0.00		H$\cdots$H(*)	0.04	0.01
CH-CH	C-C	2.97	2.95	HCOOH	C-O	1.98	1.65
	C-H	0.95	0.87		C-O	1.34	0.95
	C$\cdots$H	0.14	0.03		C-H	0.86	0.93
	H$\cdots$H	0.01	0.01		O-H	0.77	0.61
CH_3NH_2	C-N	1.40	1.04		C$\cdots$H	0.05	0.02
	C-H	0.94	0.95		O$\cdots$O	0.39	0.10
	C-H	0.92	0.95		CH$\cdots$O	0.19	0.07
	N-H	0.93	0.78		CH$\cdots$O	0.17	0.02
	N$\cdots$H	0.13	0.06		OH$\cdots$O	0.04	0.01
	N$\cdots$H	0.13	0.02		H$\cdots$H	0.01	0.02
	C$\cdots$H	0.08	0.02	H_2CO	C-O	2.27	1.83
	CH$\cdots$HN	0.01	0.02		C-H	0.94	0.94
	CH$\cdots$HN	0.01	0.00		O$\cdots$H	0.19	0.06
	CH$\cdots$HC(*)	0.08	0.00		H$\cdots$H	0.09	0.01
	NH$\cdots$HN(*)	0.05	0.01	CO_2	C-O	2.08	1.76
NH_3	N-H	0.96	0.81		O$\cdots$O	0.43	0.21
	H$\cdots$H(*)	0.05	0.01				

Table 3.5: Comparison of different indicators for the covalent bond order, for some polyatomic molecules with a singlet state and a B3LYP/cc-pVDZ geometry. The first column specifies the molecules under consideration. In the second column, the notation $[-]$ indicates chemical bonding, $[\cdots]$ refers to interaction between chemically non-bonded atoms and (*) indicates that geminal atoms are considered. The third column contains the well established iterative-Hirshfeld SEDI-indices $B_{AB}[w_A^{HI}]$ at the ROHF/Aug-cc- pVDZ level. The fourth column contains values for $B'_{AB}[w_A^{(\infty)}]$ at the ROHF/Aug- cc-pVDZ level.

4 Electronegativities of the Interacting Quantum Atoms and substituent effects

4.1 Introduction

The atom-in-molecule approach provides values, calculated from quantum mechanics, for some basic chemical concepts that are related to the molecular electronic structure (e.g. atomic charges, atomic valences, covalent bond order indices, bond energies) or its inherent reactivity (e.g. nucleophilicity, electrophilicity, local hardness and softness,...). We showed in the previous chapters that these values clearly benefit from a 1DM-based approach, compared to the more common density-based schemes.

A 1DM-based approach to the AIM also allows to obtain ab-initio values for essential chemical concepts that are difficult to treat within the density framework. An example is provided by the substituent effects in molecules. In organic chemistry, a substituent is an atom or group of atoms substituted in place of a hydrogen atom on the parent chain of a hydrocarbon. In general, it can be identified with any functional group of atoms in the molecule. Substituent effects are experimentally observable effects of the transmission of charge through the chain of atoms in a molecule [132]. These are related to the electron attracting/donating power of functional groups in the molecule and influence nearby reaction sites. The introduction of substituents in molecules provides the chemist with the means to model the reactivity of a compound and influence the equilibrium of a chemical reaction, reaction rates, reaction selectivity, etc.

In the most simple case, where the functional group is a single atom, the

relative importance of the substituent effect is estimated from the electronegativities of the elements. For larger functional groups, it is not trivial to estimate the relative strength of the substituent effect. Functional groups can withdraw electrons from σ bonds (the inductive effect, short range) and through a conjugated π system (the resonance effect, long range). Furthermore, there are several empirical lists of the relative strength of the electron withdrawing and electron donating effect. Although these lists differ based the experiment performed, there is a consensus in chemistry that $-CF_3$, $-NO_2$, $-[NR_3]^+$ have a strong electron withdrawing effect, while the effect is moderate for $-$COOH, $-$CHO and the halogens. Among the electron donors are $-NR_2$, $-$OR, $-[O]^-$ and $-CR_3$ [133]. The strong electron withdrawing effect of $-CF_3$ cannot be deduced from the electronegativity of the C element, nor from methods to determine an averaged electronegativity for the entire functional group (e.g. Sandersons method, see Eq. (4.12)). Indeed, the latter technique does not predict a significantly higher electronegativity for $-CF_3$ than for $-$F. The strong electron withdrawing effect of $-CF_3$ can neither be deduced directly from the charges on the atoms in the molecule, e.g. at the ROHF/6-31G* level of theory the ESP and QTAIM charges on H in HCF_3 are only 0.05 and 0.11 respectively, while the ESP and QTAIM charges on H in HF are 0.46 and 0.72. Huheey already concluded in the early 1970 that it is necessary to treat both the "inherent electronegativity" and the "charge capacity" when the inductive effect of a group is discussed [134]. Komorowski came to a similar conclusion in 1993, based upon his concept of group electronegativity [135]. He stated that the obtained group electronegativities cannot be correlated directly with Hammett substituent constants without taking the hardness into account.

It should be possible to obtain -at least qualitatively- an estimate for the electron withdrawing effect of functional groups. In this chapter, we propose that an estimate can be obtained from the electronegativity of the Interacting Quantum Atoms (IQA, see Sec. 2.3.1). The Interacting Quantum Atoms [119–121] can be obtained using an AIM scheme. The calculation of their electronegativity requires orbital energies and therefore a 1DM-based approach to the atom in the molecule (see Sec. 4.2). To avoid confusion, we already note here that the electronegativity of the IQA is not the same as the electronegativity of the AIM defined in the DFT framework (see Eq. (4.16) and Eqs. (4.19-4.20)). We also stress that the IQA electronegativity values are not only applicable to assess the inductive effects in molecules, but that they are also correlated with other very basic concepts in chemical reactivity, e.g. basicity and nucleophilicity. Electronegative IQAs tend to be electron withdrawing, acid, and bad nucleophiles. Note that in organic chemistry acidity and nucleophilicity are usually estimated from a common scale (the "organic" pKa scale of the conjugate acid associated with the functional

group [136]), although both concepts are clearly very different since acidity is determined from thermodynamic data while nucleophilicity stems from reaction kinetics. Finally, we emphasize that we do not want to present the IQA electronegativities as autonomous reactivity descriptors. Rather, they provide an explanation for inductive effects, i.e. an increase or reduction in reactivity.

In the next sections, we investigate which electronegativity definition is best suited to tackle the problem of deriving IQA electronegativity values. We then use these values to assess the relative strength of the electron withdrawing and electron donating effect resorted by different functional groups as qualitative lists.

4.2 Electronegativity definitions

Electronegativity is one of the key concepts in chemistry. Whenever two atoms form a bond, the nucleus of each atom attracts the other's electrons. Electronegativity is a measure of the strength of this attraction. Pauling first proposed the concept of electronegativity in 1932 [137] when he tried to explain why the dissociation energy $E_{\text{diss}}(\text{AB})$ of a heteronuclear diatomic is larger then the average of the dissociation energies $[E_{\text{diss}}(\text{AA}) + E_{\text{diss}}(\text{BB})]/2$ of the homonuclear diatomics. The extra stabilization energy was found to be related to "atomic electronegativity differences" as:

$$\chi_{\text{A}}^{Paul} - \chi_{\text{B}}^{Paul} \approx \sqrt{E_{\text{diss}}(\text{AB}) - [E_{\text{diss}}(\text{AA}) + E_{\text{diss}}(\text{BB})]/2}\ (\text{eV})^{-1/2}. \quad (4.1)$$

In Eq. (4.1), the quantities are expressed in eV and the factor $(\text{eV})^{-1/2}$ is included to define χ_{A}^{Paul} as a dimensionless property of the elements. Mulliken proposed in 1934 that the arithmetic mean of the first ionization energy $I = E(N-1) - E(N)$ and the electron affinity $A = E(N) - E(N+1)$ should be a measure of the tendency of an atom to attract electrons of other atoms [96],

$$\chi_A^{Mull(1)} = \frac{1}{2}(I + A). \quad (4.2)$$

The Mulliken technique was refined in the early 1960s by Iczkowski and Margrave [88], who showed, on the basis of experimental atomic ionization energies and electron affinities, that the energy of an atom could be expressed as a polynomial in the atomic charge $Q = N - Z$,

$$E_A(Q) - E_A(0) = aQ^4 + bQ^3 + cQ^2 + dQ. \quad (4.3)$$

Iczkowski and Margrave proposed to define the electronegativity as the slope

of this function:

$$\chi_A^{Mull(2)} = -\left(\frac{\partial E_A}{\partial Q}\right), \tag{4.4}$$

They noted that the first two terms in Eq. (4.3) are small and obtained a reasonable similarity to the Mulliken scale.

Until the development of DFT, electronegativity was mainly treated as a property of the atoms in molecules [138]. Studies on molecular electronegativities were conducted mainly in the context of Sanderson's electronegativity equalization method [139–144], where the emphasis was to find a set of atomic charges. In 1978, Parr et al. laid the cornerstone of conceptual DFT [145]. As the second Hohenberg-Kohn theorem states that the ground state density of a system minimizes the energy functional of a system, they tried to interpret the Lagrangian multiplier μ appearing in the variational ansatz,

$$\frac{\delta\left(E[\rho] - \mu \int d\boldsymbol{r}\rho(\boldsymbol{r})\right)}{\delta\rho(\boldsymbol{r})} = 0 \tag{4.5}$$

and the Euler-Lagrange equation derived from it:

$$\mu = v(\boldsymbol{r}) + \frac{\delta\left(T[\rho] + V_{ee}[\rho]\right)}{\delta\rho(\boldsymbol{r})}. \tag{4.6}$$

In Eq. (4.6), $v(\boldsymbol{r})$ is the external potential, $T[\rho]$ the electronic kinetic energy functional and $V_{ee}[\rho]$ the electron-electron energy functional. In analogy with the thermodynamic chemical potential μ_i of a component i in a macroscopic system at temperature T and pressure P,

$$\mu_i = \left(\frac{\partial G}{\partial n_i}\right)_{P,T,n_{j\neq i}}, \tag{4.7}$$

they proposed that μ is the electronic chemical potential,

$$\mu = \left(\frac{\delta E[\rho]}{\delta\rho(\boldsymbol{r})}\right)_v = \left(\frac{\partial E}{\partial N}\right)_v, \tag{4.8}$$

that can be identified with the electronegativity definition of Iczkowski and Margrave (Eq. (4.4)) if the implicit constraint of a fixed nuclear charge is generalized to the constraint of a fixed external potential:

$$\chi^{Mull(3)} = -\left(\frac{\partial E}{\partial N}\right)_v = -\mu. \tag{4.9}$$

Kohn, Parr, and Becke [146] stated that (i) μ characterizes the tendency of electrons to escape from the equilibrium system (atoms or molecules), (ii) systems coming together must attain a common chemical potential at equilibrium and (iii) this chemical potential is none other than the negative of the electronegativity concept of classical structural chemistry. Note that Gyftopoulos and Hatsopoulos already suggested in 1965 [89] that electronegativity is equivalent to a chemical potential. In practice, values for Eq. (4.9) can be obtained using the finite difference approach of Mulliken (see Eq. (4.2)). At the Hartree-Fock level, the ionization energy and electron affinity are represented by (minus) the energies of the frontier orbitals ϵ_{homo} and ϵ_{lumo} (Koopmans' theorem [147]) and Eq. (4.2) reduces to:

$$\chi^{Mull(1)} = -\frac{1}{2}(\epsilon_{homo} + \epsilon_{lumo}). \tag{4.10}$$

The definition in Eq. (4.9), and the assumption of N as a continuously varying property in a closed quantum mechanical system, was criticized by various people, e.g. by Bader et al. [148]. Also Allen was opposed to the definition in Eq. (4.9) as there are some fundamental differences between this electronegativity scale and the Pauling scale [149–153]. He proposed the average valence electron (~ionization) energy as an electronegativity measure. E.g., at the Hartree-Fock level one defines the "spectroscopic" electronegativity:

$$\chi^{Spec} = -\frac{\sum_i n_{\mathrm{i}} \epsilon_{\mathrm{i}}}{\sum_j n_{\mathrm{j}}}. \tag{4.11}$$

where n_{i} represents the occupancy of a valence orbital with energy ϵ_{i}. For free atoms, this measure shows a remarkably good linear correlation with the Pauling scale. For atoms in the molecule, which are generally not in their ground state, it is suggested here that a measure for the attraction of the electrons of other atoms is provided by the "averaged" ionization energy rather than only the first ionization energy.

4.3 Electronegativity equalization

The principle of electronegativity equalization was originally proposed by Sanderson [139–142] in the early 1950s. Sanderson postulated that when a molecule is formed from the atoms, the electronegativities of the atoms are altered to yield a collective molecular value equal to the geometric mean of the original electronegativities of the isolated atoms χ_A^0:

$$\chi_M = (\prod_A \chi_A^0)^{\frac{1}{m}} \tag{4.12}$$

where m is the number of atoms in the molecule. Sanderson studied the electronegativity change in the atom to find the corresponding charges on the atoms. Unfortunately, when this approach is used, all atoms of the same element adopt the same atomic charge within a molecule. The method received little interest in literature until solid proof was provided for the electronegativity equalization principle. First, Donnelly and Parr [154] showed that for a molecule in the ground state, the chemical potential in Eq. (4.8) is constant over the entire system as a consequence of the variational principle for the energy. Later, Politzer and Weinstein [155] reached a similar conclusion. Parr and Bartolotti [156] provided theoretical and numerical support for the geometric mean postulate. In 1985, Mortier et al. used the ideas of Sanderson to establish the well-known electronegativity equalization method (EEM) for obtaining AIM charges [65–71]. The method is based on the equality requirement for the electronegativities of the atoms in the molecule,

$$\chi_A = \chi_B = \cdots = \chi_Z = \chi_{mol}, \tag{4.13}$$

and a second order expansion of the molecular energy in terms of the partial atomic charges. It leads to the following expression for the χ_A:

$$\chi_A = (\chi_A^0 + \Delta\chi_A) + (\eta_A^0 + \Delta\eta_A)q_A + \sum_{B \neq A} \frac{q_B}{R_{AB}} \tag{4.14}$$

where χ_A^0 and η_A^0 are the electronegativity and hardness (Eq. (1.67)) of the isolated atoms, $\Delta\chi_A$ and $\Delta\eta_A$ are correction terms that describe the molecular environment and the final term accounts for the external potential. Since the electronegativity values of Eq. (4.14) are equal for all atoms in the molecule, they do not reflect the electron-attracting/-donating power of the substituents in the molecule. It is clarified in the next section that a rough indication of this power would be provided by the first two terms in Eq. (4.14), which represent the electronegativity of isolated atoms with deformations and charges determined by the molecule.

4.4 Substituent effects in molecules from the electronegativity of the IQA

4.4.1 Theoretical derivation

The electron attracting/donating power of substituents in the molecule is not a trivial concept, as we already mentioned in the introduction (see Sec. 4.1). It can clearly not be identified with the electronegativity of the AIM as defined in the DFT framework $(\partial E/\partial N_A)$, which is constant throughout the molecule. Rather, it expresses the extra (de)stabilization of electrons when molecules with different substituents are compared. We propose that the electronegativities of the charged Interacting Quantum Atoms (that are composed of an atom-condensed 1DM ρ_A and nucleus A, see Sec. 2.3.1), i.e. $(\partial E^{self}/\ \partial N_A)$, provide a measure for the electron-attracting/-donating power of the substituents in the molecule.

In the following paragraph we supply a set of arguments to justify our proposal. Note that it is not a rigorous derivation. The 1DM provides the populations and orbital energies that are necessary to calculate electronegativity values with Eqs. (4.10) and (4.11). A partitioning of the molecular 1DM over the atoms provides Interacting Quantum Atoms. In Sec. 2.3.1 we derived the energy components associated with these molecular fragments:

$$E = \sum_A E_A^{Self} + \sum_{A \neq B} E_{AB}^{Int} \tag{4.15}$$

where E_A^{Self} are the energies of the noninteracting IQA and E_{AB}^{Int} are the interaction energies between the IQA. Eq. (4.15) can be used to decompose the AIM electronegativity in the DFT electronegativity scale (Eq. (4.9)),

$$\chi_A = \left(-\frac{\partial E}{\partial N_A}\right) = \left(-\frac{\partial E_A^{Self}}{\partial N_A}\right) + \sum_{B \neq A} \left(-\frac{\partial E_{AB}^{Int}}{\partial N_A}\right). \tag{4.16}$$

A constant external potential is assumed in Eq. (4.16). We notice that the terms appearing in this decomposition show large conceptual resemblance to the terms appearing in the EEM scheme (Eq.(4.14)). The E_A^{Self} are based on an atomic Hamiltonian and represent the energies of charged and deformed (noninteracting) atoms. Therefore, the first term in Eq. (4.16) corresponds to the first two terms in Eq. (4.14):

$$\left(-\frac{\partial E_A^{Self}}{\partial N_A}\right) \sim (\chi_A^0 + \Delta\chi_A) + (\eta_A^0 + \Delta\eta_A) q_A \tag{4.17}$$

Since the E_{AB}^{Int} are interaction energies, the second term in Eq. (4.16) might be interpreted as the interaction potential created by the environment of IQA A. Therefore, it corresponds to the third term in Eq. (4.16):

$$\sum_{B\neq A}\left(-\frac{\partial E_{AB}^{Int}}{\partial N_A}\right) \sim \sum_{B\neq A}\frac{q_B}{R_{AB}} \tag{4.18}$$

Note that we assume that the equivalence of the concepts of electronegativity and chemical potential (see E. (4.8)) also applies to individual terms of the decomposition. Also note that E_{AB}^{Int}has both electrostatic and Fock components. We introduce a shorthand notation for the terms in Eq.(4.16),

$$\chi_A = \left(-\frac{\partial E}{\partial N_A}\right) \tag{4.19}$$

$$\chi_A^{Self} = \left(-\frac{\partial E_A^{Self}}{\partial N_A}\right) \tag{4.20}$$

$$U_A^{Int} = \sum_{B\neq A}\left(-\frac{\partial E_{AB}^{Int}}{\partial N_A}\right) \tag{4.21}$$

where χ_A is the AIM electronegativity that equals the molecular electronegativity and $\chi_A^{Self}$is the IQA electronegativity. $U_A^{Int}$is the potential generated by the environment of the IQA that stabilizes (destabilizes) its orbital energies ϵ_i. Note that "electron withdrawing" functional groups tend to stabilize the electrons of the IQA to which it is attached (in a sense that the ϵ_i become more bound), while "electron donating" groups destabilize these electrons. Comparing the (de)stabilization of the electrons of an IQA between molecules with different substituents is difficult since in general the "reference" level $\chi_A = \chi_{mol}$ differs:

$$-U_A^{Int} = \chi_A^{Self} - \chi_A. \tag{4.22}$$

The difference in (de)stabilization between the atoms A and B that attach the functional group (X-B) to the substrate (R-A) in the molecule (R-A-B-X) does not depend on a reference level,

$$-(U_A^{Int} - U_B^{Int}) = \chi_A^{Self} - \chi_B^{Self}, \tag{4.23}$$

and is a measure for the (inductive) substituent effect in the molecule. It can be calculated as a difference between IQA electronegativities. At the self-consistent field level, the electronegativities are commonly evaluated in terms of the energies of the frontier orbitals:

$$\chi_A^{Self} \approx -\frac{1}{2}(\epsilon_{A,homo} + \epsilon_{A,lumo}), \tag{4.24}$$

Unfortunately, such an expression cannot be used here since the IQA are not in the ground state. The only electronegativity definition that is computationally practical for IQA is the spectroscopic electronegativity expression of Allen:

$$\chi_A^{Self} \approx -\frac{\sum_i n_{A,i}\epsilon_{A,i}}{\sum_j n_{A,j}}. \tag{4.25}$$

We only expect to get sensible results from Eq. (4.25) if the electron densities of the IQA are local and positive semi-definite (see Sec. 2.3.1 and Sec. 3.2.4). This requirement is needed to guarantee that the electrons are bound with respect to the Hamiltonian of the isolated atom. Good candidates for the IQA atoms are obtained from an AIM partitioning of the 1DM in the Hirshfeld-I consistent scheme (see Sec. 2.2):

$$\rho_A \longleftarrow \rho_A^{HI} = \frac{1}{2}\left(w_A^{HI}\rho + \rho w_A^{HI}\right). \tag{4.26}$$

Some preliminary calculations were performed using Eq. 4.26 and promising results were obtained, but we do not yet have sufficient results to include them in this thesis. The partitioning scheme with orthogonal projectors in one-electron Hilbert space (see Sec. 3.2.1) does not provide single-atom 1DMs that are both local and positive semi-definite (see Sec. 3.2.4). To obtain at least a crude measure for the substituent effects, we rely on the model 1DM used in that scheme ($\rho^{(0)}[N_A, N_B, ...] = \sum_A \rho_{AA}^{(0)}[N_A]$, see Eq. (3.18)),

$$\rho_A \longleftarrow \rho_{AA}^{(0)}[N_A] \tag{4.27}$$

where $\rho_{AA}^{(0)}[N_A]$ is the 1DM of the isolated atom with the same occupancies as the 1DM $\rho_{AA}^{(\infty)}$ of the "non-bonded" atoms and the 1DM $\rho_A^{(\infty)}$ of the "bonded" atoms. A set of orbital energies for ρ_A is obtained with a mean-field expression. A Fock matrix $(F_A)_{ij}^{\alpha,\beta}$is constructed from the one- $(h_A)_{ij}$and two electron integrals $(V_{ijkl})$and the (possibly correlated) atom-condensed 1DM's $(\rho_A)_{ij}^{\alpha}$and $(\rho_A)_{ij}^{\beta}$:

$$(F_A)_{ij}^{\alpha,\beta} = (h_A)_{ij} + \sum_{kl}\left(V_{ikjl}\sum_{\sigma}(\rho_A)_{kl}^{\sigma} - V_{iklj}(\rho_A)_{kl}^{\alpha,\beta}\right). \tag{4.28}$$

Diagonalization of the (generalized) Fock matrix $(F_A)_{ij}^{\alpha,\beta}$generates a set of orthonormal orbitals $\phi_{A,i}$and the corresponding orbital energies $\epsilon_{A,i}$:

$$F_A^{\alpha,\beta}\phi_{A,i}^{\alpha,\beta} = \epsilon_{A,i}^{\alpha,\beta}\phi_{A,i}^{\alpha,\beta}. \tag{4.29}$$

The occupancies $n_{A,i}^{\alpha,\beta}$ of the orbitals $\phi_{A,i}^{\alpha,\beta}$ are determined as:

$$n_{A,i}^{\alpha,\beta} = \langle\phi_{A,i}^{\alpha,\beta}|\rho_A^{\alpha,\beta}|\phi_{A,i}^{\alpha,\beta}\rangle. \tag{4.30}$$

Note that the $\phi_{A,i}^{\alpha,\beta}$ are not eigenfunctions of $\rho_A^{\alpha,\beta}$, so

$$\rho_A^{\alpha,\beta} \neq \sum_i n_{A,i}^{\alpha,\beta} \phi_{A,i}^{\alpha,\beta} \phi_{A,i}^{\alpha,\beta}. \tag{4.31}$$

Finally, Allen electronegativities χ_A for the IQA are calculated from the valence orbital energies and occupancies,

$$\chi_A^{Self} = -\sum_{i,\sigma} \frac{n_{A,i}^{\sigma} \epsilon_{A,i}^{\sigma}}{\sum_{j,\sigma} n_{A,j}^{\sigma}}, \tag{4.32}$$

where i runs over all valence orbitals.

4.4.2 Results

All data in this section are based on the IQA obtained with the partitioning scheme that relies on orthogonal projectors (see Eq. (4.27). We checked that comparable results can be obtained for IQA based on the Hirshfeld-I AIM model, although it requires a larger basis set. In Fig. 4.1, IQA

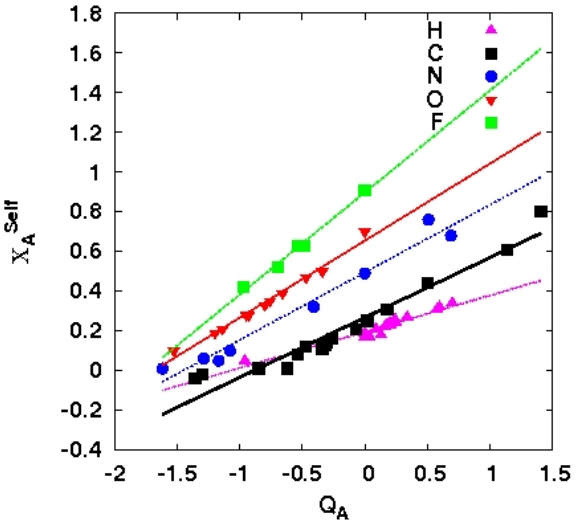


Figure 4.1: IQA electronegativities (χ_A^{Self}), for model atoms obtained with the orthogonal projectors scheme, as a function of the atomic charges (Q_A). Calculations were performed at the ROHF/Aug-cc-pVDZ level of theory.

electronegativities (χ_A^{Self}) are plotted as a function of the atomic charges (Q_A) for a test set of 35 small molecules (all of which are presented in Table 4.1 and/or Table 4.2) with geometries obtained by B3LYP-cc-pVDZ optimizations. Calculations were performed at the ROHF/Aug-cc-pVDZ level. There are three interesting observations. (i) For each element, there is a very strong linear correlation between the IQA electronegativities and Q_A (R^2 =0.92, 0.94, 0.97, 0.98, 0.99 for respectively H, C, N, O, and F). It appears that the IQA electronegativities are mainly determined by the atomic charge rather than the actual atomic state. For H and C, the correlation is slightly lower than for the more electronegative elements since the average energy of the valence electrons in the IQA does not tend to become positive in the region of large negative charges. (ii) The fit value at $(Q_A = 0)$ (0.20, 0.27,

0.49, 0.65 and 0.90 for respectively H, C, N, O, and F) correlates linearly with the element electronegativities of Pauling (R^2=0.99). (iii) The slope predicts an increasing atomic hardness for elements that have larger element electronegativity (slope=0.18, 0.30 ,0.34 ,0.39 ,0.51 for respectively H, C, N, O, and F). The hardness for H is rather low, but note that the IQA electronegativity values deviate from the linear interpolation when $\chi_A^{Self} \approx 0$. The hardness of H increases to a value of 0.24 if LiH is not included in the test set. (iv) We also observed that if the population of a particular atomic orbital increases, all valence orbital energies increase.

	A	Q_A	χ_A^{Self}			A	Q_A	χ_A^{Self}
$[CH_3]^-$	C	-1.30	-0.02		LiF	F	-0.97	0.42
$LiNH_2$	N	-1.62	0.01		HNO_2	O	-0.47	0.47
NH_3	N	-1.29	0.06		N_2	N	0.00	0.49
LiOH	O	-1.53	0.10		HONO	O	-0.34	0.50
CH_3NH_2	N	-1.08	0.10		HF	F	-0.70	0.52
H_2O	O	-1.20	0.19		HCl	Cl	-0.34	0.55
$[H_3O]^+$	O	-1.14	0.21		CH_3Cl	Cl	-0.25	0.59
CH_3OH	O	-0.95	0.28		CH_3F	F	-0.53	0.63
HCOOH	O	-0.93	0.28		HCF_3	F	-0.48	0.63
NaCl	Cl	-0.98	0.30		Cl_2	Cl	0.00	0.69
HCN	N	-0.41	0.32		O_2	O	0.00	0.70
HCOOH	O	-0.80	0.33		HONO	N	0.51	0.76
HONO	O	-0.76	0.35		F_2	F	0.00	0.91
CH_2O	O	-0.66	0.39					

Table 4.1: Atomic charges (Q_A) and IQA electronegativities (χ_A^{Self}) for atoms that contain a free electron pair in the molecule. Calculations are performed at the ROHF/Aug-cc-pVDZ level of theory. Nucleophilicity trends are reflected relatively well in the values for χ_A^{Self}.

In Table 4.1 atomic charges (Q_A) and IQA electronegativities (χ_A^{Self}) are tabulated for atoms that contain a free electron pair in the molecule. The species are sorted by their IQA electronegativity values. Some important observations: (i) Basic and nucleophilic species (small values) are located at the top of the table, acid and bad nucleophiles (large values) are listed at the bottom. The ordering of the electronegativity values ($[CH_3]^- < Li^+[NH_2]^- < NH_3 < Li^+[OH]^- < H_2O <$ HCOOH $< Na^+[Cl]^- < Li^+[F]^-$) roughly respect the pKa values of the conjugate acid[1](60, 36, 9, 16, -2, -6, -8, 3) [136]. (ii) Trends for the χ_A^{Self}values are in accordance with trends in electrophilicity.

[1]Note that for the ionic species, the conjugate acid of the anion is considered.

The atom containing the free electron pair in functional groups has a smaller IQA electronegativity and is less electrophilic (more nucleophilic) in salts than in the corresponding covalent compound ($[NH_2]^- < NH_3$, $[OH]^- < H_2O$, $F^- < HF$). The values also increase going to the right in the periodic table ($NH_3 < H_2O < HF$) and going up the periodic table (NaCl < LiF). The trend in IQA electronegativities also agrees with the trend in experimental nucleophilicity values cited by Jamarillo et al. [157]. Note that the nucleophilicity of Cl^- and F^- is somewhat underestimated by the trends in χ_A^{Self} values.

H-R	H-B	$\chi_B^{Self} - \chi_H^{Self}$		H-R	H-B	$\chi_B^{Self} - \chi_H^{Self}$
H-CF_3	H-C	0.63		H-CCH	H-C	-0.10
H-NO_2	H-N	0.44		$[H\text{-}H_2O]^+$	H-O	-0.12
H-COOH	H-C	0.44		H-OH	H-O	-0.12
H-Cl	H-Cl	0.28		H-CH-CH_2	H-C	-0.12
H-CHO	H-C	0.23		H-CH_2Cl	H-C	-0.14
H-F	H-F	0.18		H-CH_2CH_3	H-C	-0.15
H-CN	H-C	0.05		H-OLi	H-O	-0.22
H-ONO	H-O	0.04		H-CH_3	H-C	-0.22
H-BH_2	H-B	0.03		$[H\text{-}CH_2]^-$	H-C	-0.22
H-CH_2F	H-C	0.02		H-NH_2	H-N	-0.23
H-H	H-H	0.00		H-$(CH_2)_2CH_3$	H-C	-0.24
H-CH_2OH	H-C	0.00		$[H\text{-}NH_3]^+$	H-N	-0.25
H-CH_2NH_2	H-C	-0.05		H-CH_2Li	H-C	-0.25
H-CH-$(CH3)_2$	H-C	-0.09		H-NHLi	H-N	-0.27

Table 4.2: The IQA electronegativity difference ($\chi_B^{Self} - \chi_H^{Self}$) for H-B bonds in H-R molecules. Calculations are performed at the ROHF/Aug-cc-pVDZ level of theory. The electronegativity differences reflect relatively well the electron withdrawing or electron donating effects of the functional group R.

Table 4.2 lists the IQA electronegativity difference ($\chi_B^{Self} - \chi_H^{Self}$) for H-B bonds in H-R molecules. The electronegativity differences reflect relatively well the electron withdrawing/pushing effect of the functional group R. Electron withdrawing functional groups (large positive values) are located at the top of the table, electron pushing functional groups (large negative values) are listed at the bottom. Electrophilic aromatic substitutions are highly deactivated/activated by the presence of an electron withdrawing/pushing functional group. -CF_3 and -NO_2 are known as strongly deactivating functional groups, -COOH and -CHO are moderately deactivating, -F and -Cl are weakly deactivating, alkynes -CCH, alkenes -$CHCH_2$ and alkanes -CH_2CH_3 are weakly activating and -OH, -OLi, -NH2 and -NHLi are strongly activating

functional groups [133]. This is roughly in accordance with the ordering of the species according to the electronegativity difference values. Note that the electron withdrawing character of Cl^- and F^- would be somewhat overestimated by the trends in electronegativity differences. A curious case is the -NH_3^+ functional group. While -NR_3^+ (with R a hydrocarbon) is known to be a strong deactivating functional group, -NH_3^+ is classified to have a weak electron withdrawing effect. This is not in accordance with the large negative value for $\chi_B - \chi_H$. However, normal deactivating functional groups favor aromatic substitution at the meta position, while -NH_3^+ slightly favors substitution at the para position, like functional groups with an electron pushing inductive effect.

Note, while inspecting Tables 4.1 and 4.2, that it is possible to extend the electronegativity of the IQA (χ_A^{Self}) to functional groups of IQA. Such approach requires that the average valence electron energy is calculated for the entire functional group. The Fock matrix in Eq. (4.28) is then constructed from the summed atomic 1DMs ρ_A and an external potential generated by the nuclei of the functional group. It would be interesting to investigate the influence of this approach on the values presented in Tables 4.1 and 4.2.

5 General conclusions and perspectives

In this thesis the atom-in-molecule (AIM) concept has been explored. This concept reconciles the chemical approach to the molecule, based on atoms and functional groups, with the quantum mechanical framework. Whereas most conventional techniques depend on a partitioning of the electron density over the atoms in the molecule to define AIM properties, we investigated the possibility to base this elusive concept on the full one-electron density matrix (1DM) instead. In a 1DM-based approach to the AIM, all one-electron properties (and two-electron properties at the mean-field level) are evaluated with an explicit expression in terms of the atom-condensed 1DMs. In density-based approaches, precise and computationally tractable expressions in terms of the atomic densities are not available. In that case AIM values often rely on a partitioning of these properties themselves over the atoms. We pointed out in Chapter 1 that the latter approach is inherently ambiguous and problematic. For a non-local operator, e.g. the kinetic energy, density-based AIM values depend in general on the representation of the operator and the place where the partitioning of unity into atomic weight functions is introduced in the expectation value expression. In contrast, 1DM-based AIM values are uniquely determined.

In a first 1DM-based AIM approach, consistency was required with established density-based techniques. We developed and discussed a 1DM-based extension of the Hirshfeld-I atom-in-molecule model (see Chapter 2). The Hirshfeld-I scheme is known to produce reliable and transferable atomic charges for application in forcefield methods. It has been applied to study dispersion interactions and to generate atom-condensed reactivity descriptors. The success of this scheme relies notably on the local and positive semi-definite character of the AIM densities. However, as the Hirshfeld-I atom-in-molecule model is based on a partitioning of the electron density over atomic domains

with fuzzy boundaries, it is particularly subject to the problems described in the previous paragraph. We showed that the inherent ambiguity in the AIM values for the kinetic energy is more than 0.3 Hartree in some cases. We demonstrated that counterintuitive and potentially misleading values can be obtained for the energy terms associated with the AIM. The advantage of the 1DM-based extension of the Hirshfeld-I atom-in-molecule model, is that the energy terms are now fully consistent with the underlying partitioning of the electronic structure. In addition, this procedure can be implemented efficiently in one-electron Hilbert space to avoid the cumbersome numerical integrations that appear in the conventional methods. The full power of 1DM-based extension is shown in Chapter 4, where it is pointed out that the accurate calculation of substituent effects in the molecule actually requires (i) a 1DM-based approach to the AIM and (ii) a local and positive semi-definite character for the atomic densities.

We also discussed a second 1DM-based AIM technique. In this method we tried to create an optimal framework for the atom-in-molecule model at the 1DM level, by combining the advantages of the different existing techniques. Inspired by the Mulliken approach, all calculations are performed as matrix manipulations in one-electron Hilbert space, making the method efficient in terms of computation time, since no grid-based numerical integrations are needed. For interpretative convenience, orthogonal projection operators are used to define atomic subspaces (of one-electron Hilbert space), as in the Quantum Theory of Atoms In Molecules (QTAIM). These operators are determined by the fraction of an atomic 1DM with respect to the sum of all atomic 1DMs, as in the Hirshfeld method (that is firmly rooted in information theory). The operators can then be used to define density matrices assigned to atoms and bonds in the molecule. An iterative procedure ensures that the method is independent from the initial guess for the atomic 1DMs, as in the iterative version of the Hirshfeld method.

This new AIM technique has many appealing characteristics. We pointed out that it is an orbital specific approach, based on atomic weight matrices that are not diagonal in $\boldsymbol{r}$-space. Core electrons and lone pairs are therefore assigned to only one atom. The atomic weight matrices also have an expression that ensures good convergence properties in standard basis sets. Their chemical relevance originates from the conceptual resemblance to the Hirshfeld-I method. E.g. atomic charges correlate fairly well with the established Hirshfeld-I charges, and have a plausible magnitude (in contrast to e.g. QTAIM charges). The orthogonal projector property allows to set up an interesting interpretative framework. In this framework, the partitioning of the molecular 1DM over atoms and bonds defines a set of "excited" atoms

in the molecule that do not share electrons, but contribute to the full electron number. The bonds are described as electronic deformation processes, in line with the chemical notion of a bond. The properties associated with bonding, e.g. the covalent bond-order index, are chemically much more intuitive in the current framework than in the Hirshfeld-I model. It is also possible to assess the individual contributions of the atoms to the bonds, i.e. it is possible to define a "dative" character for these bonds. We also observed that the well-localized atomic 1DMs ρ_{AA} can be associated with a set of delocalized atomic 1DMs ρ_A, that has the same eigenvalue spectrum and sums to the molecular 1DM. The populations of the atoms are therefore well-defined an can be used to construct model atoms ("excited" isolated atoms). These model atoms were used e.g. to determine substituent effects in Chapter 4.

Apart from the introduction of two 1DM-based AIM techniques, we also tested an application that requires a 1DM-based approach to the AIM. Note that not all the properties of "molecular fragments" can be derived from the AIM densities or from a fragmentation of the molecular property. A simple example is the electronegativity of the Interacting Quantum Atom (IQA). The IQA is a combination of an AIM fragment of the electronic structure with a single nucleus. Practical calculation of the electronegativity of this "excited" and "deformed" atom requires orbital energies and thus the atom-condensed 1DM at the mean-field level. We showed that the electronegativity of the IQA is related to the inductive effects of atoms and functional groups in molecules. This demonstrates that the 1DM-approach to the atom in the molecule provides important information about the bonding mechanisms in molecules that can not be obtained from one of the many previously investigated AIM methods.

It would be interesting to consider several extensions to this work, that might e.g. contribute to the investigation of the catalytic activity of transition metal complexes. In these complexes, the spin state of the central metal atom determines the catalytic activity of the complex. The substituents model the reactivity and the selectivity of the complex. Determination of the spin-state of the central atom and the influence of the substituents could be examined with the 1DM-based AIM approaches presented in the current work. In this thesis, we considered simple molecules with a singlet spin state. An extension to other spin states is required for application in transition metal complexes. Also, an extension of the electronegativity of the IQA to functional groups of IQA would be valuable. Both extensions are rather straightforward and will be studied in the future.

Part II

Papers

Paper I

Partitioning of the molecular density matrix over atoms and bonds

D. Vanfleteren, D. Van Neck, P. Bultinck, P. W. Ayers, and M. Waroquier

The Journal of Chemical Physics 132, 164111 (2010)

Partitioning of the molecular density matrix over atoms and bonds

Diederik Vanfleteren,[1] Dimitri Van Neck,[1] Patrick Bultinck,[2] Paul W. Ayers,[3] and Michel Waroquier[1]

[1]*Ghent University, Center for Molecular Modeling, Technologiepark 903, B-9052 Zwijnaarde, Belgium*

[2]*Ghent University, Department of Inorganic and Physical Chemistry, Krijgslaan 281 (S3), B-9000 Gent, Belgium*

[3]*McMaster University, Department of Chemistry, Hamilton, Ontario L8S 4M1, Canada*

(Dated: March 25, 2010)

Abstract

A double-index atomic partitioning of the molecular first-order density matrix is proposed. Contributions diagonal in the atomic indices correspond to atomic density matrices, whereas off-diagonal contributions carry information about the bonds. The resulting matrices have good localization properties, in contrast to single-index atomic partitioning schemes of the molecular density matrix. It is shown that the electron density assigned to individual atoms, when derived from the density matrix partitioning, can be made consistent with well-known partitions of the electron density over AIM basins, either with sharp or with fuzzy boundaries. The method is applied to a test set of about 50 molecules, representative for various types of chemical binding. A close correlation is observed between the trace of the bond matrices and the SEDI (shared electron density index) bond index.

I. INTRODUCTION

The most common way chemists look at molecules is to consider them as composed of atoms held together by chemical bonds. Moreover, the chemical characteristics of atoms and functional groups of bonded atoms are highly transferable between different molecules. Although this picture predates quantum mechanics, it is so useful for rationalizing and even predicting experimental observations that it is still ubiquitous. However, the question remains of how to properly describe the atom in the molecule (AIM) in a quantum mechanical way. This question has been addressed by many people and various techniques have been developed to describe this elusive concept.

The techniques used thus far can largely be divided into only a few different categories. In one group, one uses the attachment of basis functions to atomic centers to extract the AIM. The best known method of this type is the Mulliken population analysis [1]. The second, and for the present paper the most important, group of techniques uses a three-dimensional (3D) splitting of space with either sharp boundaries between different AIM (e.g. Bader's Quantum Chemical Topology (QCT) [2–4]), or with more fuzzy boundaries (e.g. the original Hirshfeld method [5] and recent extensions [6–10], and Mayer's fuzzy atoms [11]). In the second group of methods, one uses the molecular electron density and its properties as the guide for obtaining the AIM. However, not all AIM properties can be directly expressed in terms of the electron density. A very simple example is the kinetic energy of an AIM, for which the full (nondiagonal) density matrix is needed rather than the electron density. For some quantities one even has to go up to the second order density matrix. This means that a more fundamental approach to the AIM should be based on density matrices [12]. In Bader's QCT only the electron density and its derivatives are required to arrive at the AIM energy via the AIM virial theorem [2], but several concerns remain. For instance, Cioslowski and Karwowski arrived at the conclusion that arbitrary choices in the Lagrangian density can have an important influence on the uniqueness of Bader's AIM [13].

In the present work, we describe how an AIM density matrix that is consistent with a 3D division of the molecular density in AIM's can be obtained. Several such methods have been explored previously, for instance, in the work of Alcoba et al [14]. These authors derive a QCT-based density matrix in the following way. First, the one density matrix is expressed in terms of an orthonormal molecular orbital set through the matrix $D = \{D_{i\sigma,j\sigma}\}$ where $i\sigma$ denotes a molecular orbital i with spin σ. Given the positive-definite character of this matrix, it can be factorized as

follows:

$$D_{i\sigma,j\sigma} = \sum_{k,l} \left(D^{1/2}\right)_{i\sigma,k\sigma} \delta_{k\sigma,l\sigma} \left(D^{1/2}\right)_{l\sigma,j\sigma} \tag{1}$$

The Kronecker delta $\delta_{k\sigma,l\sigma}$ can be very simply rewritten as:

$$\delta_{k\sigma,l\sigma} = \langle k\sigma|l\sigma\rangle = \sum_A \langle k\sigma|l\sigma\rangle_A = \sum_A \langle k\sigma|W_A|l\sigma\rangle \tag{2}$$

In other words, the Kronecker delta is written as a sum of atom-condensed overlap integrals $\langle k\sigma|l\sigma\rangle_A$ where the weight $W_A(\boldsymbol{r})$ acts as an operator to delineate the AIM domain in the molecule. QCT and Hirshfeld based methods differ mainly in the choice of the operator $W_A(\boldsymbol{r})$, which is either binary as in QCT, or fuzzy as in Hirshfeld and related methods. An AIM density matrix $D^A_{i\sigma,j\sigma}$ with eigenvalues constrained to the interval $[0,1]$ can then be obtained as:

$$D^A_{i\sigma,j\sigma} = \sum_{k,l} \left(D^{1/2}\right)_{i\sigma,k\sigma} \langle k\sigma|W_A|l\sigma\rangle \left(D^{1/2}\right)_{l\sigma,j\sigma} \tag{3}$$

This method, although shown to give interesting results when starting from QCT, has as the drawback of being inconsistent with the underlying AIM method. QCT starts from a strict, binary division of space in AIM domains. However, inspection of Eq. (3) shows that the electron density of the AIM is not confined to the AIM domain but spreads over the entire space. Extension of the above to a more fuzzy partitioning of space is straightforward, but any scheme along the lines of Eq. (3) will result in orbitals extending far outside the atomic basin assigned to A. It should be added that Alcoba et al. also introduced a different partitioning by distributing only the molecular occupation numbers in the density matrix over the different atoms, retaining the molecular natural orbitals. This obviously again does not lead to very well localized density matrices although the authors reduce this problem by first localizing the molecular natural orbitals [15, 16].

In the following we pursue a method that satisfies all the following requirements:

- The AIM density matrix should be derived starting from a 3D partitioning of space into atomic domains such that the AIM density matrix and electron density always remain mutually consistent.
- The sum of AIM density matrix eigenvalues should equal the electron occupancy of the AIM as obtained from the density analysis, and both the starting AIM density and that obtained from the coordinate space diagonal elements of the density matrix should be the same.
- The density matrices obtained should be localized.

- The scheme should involve a double atomic index partitioning with diagonal elements AA corresponding to twice the atom A and off-diagonal elements AB, whose eigenvectors correspond to chemical bonding between the atoms A and B.

The two-index approach is necessary because of the inherent non-local nature of the density matrix, as was recently also argued by Mayer and Salvador [17]. Introducing the two-index partitioning also provides an orbital perspective on the changes in the atoms when bonds are formed, by extracting bond orbitals with associated eigenvectors from the "bond" density matrices. In this sense, our method is reminiscent in philosophy to the so-called Natural Orbitals for Chemical Valence introduced by Nalewajski et al. [18] and used recently to describe chemical bonds by Ziegler and co-workers [19, 20].

II. THEORY

A. Double atom partitioning of the molecular density matrix

We use notation $\boldsymbol{x} = \boldsymbol{r}\sigma$ to specify the single-electron states in coordinate space, where σ represents the spin degrees of freedom. The first-order density matrix (1DM) for an N-electron molecule with wave function $\Psi(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$ is defined as

$$\rho(\boldsymbol{x}, \boldsymbol{x}') = N \int d\boldsymbol{x}_2 \ldots \int d\boldsymbol{x}_N \Psi^\dagger(\boldsymbol{x}, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N)\Psi(\boldsymbol{x}', \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N) \tag{4}$$

We restrict ourselves to molecules with a singlet ground state. In that case

$$\rho(\boldsymbol{x}, \boldsymbol{x}') = \frac{1}{2}\delta_{\sigma,\sigma'}\rho(\boldsymbol{r}, \boldsymbol{r}'), \tag{5}$$

and the electron spin can be discarded.

Successful partitioning schemes based on dividing the molecular electron density $\rho(\boldsymbol{r}) \equiv \rho(\boldsymbol{r}, \boldsymbol{r})$ in atomic parts rely on introducing positive atomic weight functions $W_A(\boldsymbol{r})$, which set up an atomic decomposition of space [21] in a sense that $\sum_A W_A(\boldsymbol{r}) = 1$. Then,

$$\rho_A(\boldsymbol{r}) = \rho(\boldsymbol{r})W_A(\boldsymbol{r}) \tag{6}$$

represents the fraction of the molecular electron density assigned to atom A. The various schemes differ mainly in the nature of $W_A(\boldsymbol{r})$. In Bader's QCT [2–4], $W_A(\boldsymbol{r})$ is a binary operator. In the Hirshfeld method [5] and the more recent Hirshfeld-I extension of it [6], $W_A(\boldsymbol{r})$ is given as:

$$W_A(\boldsymbol{r}) = \frac{\rho_A^0(\boldsymbol{r})}{\sum_B \rho_B^0(\boldsymbol{r})} \tag{7}$$

where $\rho_A^0(\boldsymbol{r})$ is the density of the isolated atom A. The difference between the regular Hirshfeld and the Hirshfeld-I method lies in the choice of the states and charges of the atoms used, as described in Bultinck et al. [5]. In the Iterated Stockholder Atoms [9, 10], a separate weight function is used for different spherical shells around the atom.

In this paper we want to extend the idea of dividing the molecular electron density to the (more complex) first-order density matrix. At first, a single atom partitioning scheme

$$\rho(\boldsymbol{r},\boldsymbol{r}') = \sum_A \rho_A(\boldsymbol{r},\boldsymbol{r}') \tag{8}$$

was sought that would fulfill the essential property of hermiticity with eigenvalues between 0 and 2. This single atom density matrix should also be strictly localized; that is, the 1DM assigned to an atom A should be only appreciably different from zero when both $\boldsymbol{r}$ and $\boldsymbol{r'}$ are near atom A. Equivalently, the occupied natural orbitals of the atomic density matrix should be strictly confined to the neighborhood of the atom. The transferability of the AIM scheme clearly benefits from this concept of localization. Unfortunately, a single atom partitioning scheme of the 1DM as in Eq. (8) can never fulfill this requirement because the molecular 1DM is (almost always) delocalized and has sizeable contributions for $\boldsymbol{r}$ and $\boldsymbol{r'}$ near two different atoms. To accomodate the localization requirement there should be a double atomic weighting, depending on both $\boldsymbol{r}$ and $\boldsymbol{r'}$. This was also recently argued by Mayer and Salvador [17]. Therefore it seems more natural to introduce a double-atom partitioning.

$$\rho(\boldsymbol{r},\boldsymbol{r}') = \sum_{AB} \rho_{AB}(\boldsymbol{r},\boldsymbol{r}'), \tag{9}$$

where

$$\rho_{AB}(\boldsymbol{r},\boldsymbol{r}') = \frac{1}{2}\left(w_A(\boldsymbol{r})w_B(\boldsymbol{r}') + w_B(\boldsymbol{r})w_A(\boldsymbol{r}')\right)\rho(\boldsymbol{r},\boldsymbol{r}') = \rho_{BA}(\boldsymbol{r},\boldsymbol{r}') \tag{10}$$

and the $w_A(\boldsymbol{r})$ are atomic weight functions obeying

$$0 \leq w_A(\boldsymbol{r}) \leq 1 \quad \text{and} \quad \sum_A w_A(\boldsymbol{r}) = 1. \tag{11}$$

We will use the lowercase notation $w_A(\boldsymbol{r})$ to indicate weight functions used in the double-index partitioning (9-10). In general, these will differ from the (uppercase) weight functions $W_A(\boldsymbol{r})$ used in a single-index electron density partitioning as in Eq. (6).

Provided the weight functions are properly localized, it is clear that $\rho_{AB}(\boldsymbol{r},\boldsymbol{r}')$ also will have suitable localization properties, i.e. being appreciably different from zero only when one of the arguments is near atom A and the other one near atom B.

B. Properties of the matrix partitioning

The individual contributions to the decomposition (9) are all hermitian matrices. They clearly come in two types, diagonal (AA) and off-diagonal (AB with $A \neq B$), which will be called atomic density matrices and bond matrices, respectively.

The diagonal terms ρ_{AA} indeed qualify as first order density matrices, having eigenvalues between 0 and 2. The lower bound is a trivial consequence of

$$\rho_{AA}(\boldsymbol{r}, \boldsymbol{r}') = w_A(\boldsymbol{r})\rho(\boldsymbol{r}, \boldsymbol{r}')w_A(\boldsymbol{r}') \tag{12}$$

and of the positivity of the molecular 1DM. The upper bound follows from the fact that for any electron wave function $\phi(\boldsymbol{r})$ one has

$$\begin{aligned} \int d\boldsymbol{r} \int d\boldsymbol{r}' \phi(\boldsymbol{r})\rho_{AA}(\boldsymbol{r}, \boldsymbol{r}')\phi(\boldsymbol{r}') &= \int d\boldsymbol{r} \int d\boldsymbol{r}' [w_A(\boldsymbol{r})\phi(\boldsymbol{r})]\rho(\boldsymbol{r}, \boldsymbol{r}')[w_A(\boldsymbol{r}')\phi(\boldsymbol{r}')] && (13) \\ &\leq 2 \int d\boldsymbol{r} w_A(\boldsymbol{r})^2 \phi(\boldsymbol{r})^2 && (14) \\ &\leq 2 \int d\boldsymbol{r} \phi(\boldsymbol{r})^2. && (15) \end{aligned}$$

The inequality (14) results from the upper bound for the eigenvalues of the molecular 1DM, which implies that $\int d\boldsymbol{r} \int d\boldsymbol{r}' \chi(\boldsymbol{r})\chi(\boldsymbol{r}')\rho(\boldsymbol{r}, \boldsymbol{r}') \leq 2$ for any normalized wave function $\chi(\boldsymbol{r})$, and by applying this property to $\chi(\boldsymbol{r}) = [w_A(\boldsymbol{r})\phi(\boldsymbol{r})]/\{\int d\boldsymbol{r}[w_A(\boldsymbol{r})\phi(\boldsymbol{r})]^2\}^{1/2}$. The inequality (15) follows from $w_A(\boldsymbol{r}) \leq 1$.

The summed atomic density matrices do not carry the total number of electrons, since

$$\sum_A \int d\boldsymbol{r} \rho_{AA}(\boldsymbol{r}, \boldsymbol{r}) = \sum_A \int d\boldsymbol{r} \rho(\boldsymbol{r}) w_A^2(\boldsymbol{r}) \leq \sum_A \int d\boldsymbol{r} \rho(\boldsymbol{r}) w_A(\boldsymbol{r}) \leq N; \tag{16}$$

here we used Eq. (11) and $w_A^2(\boldsymbol{r}) \leq w_A(\boldsymbol{r})$. The defect in the electron number must be in the bond matrices, as Eq. (9) implies that $N = \sum_{AB} \int d\boldsymbol{r} \rho_{AB}(\boldsymbol{r}, \boldsymbol{r})$. For each atom pair AB, the bond matrix is seen to carry a positive number of electrons, $\int d\boldsymbol{r} \rho_{AB}(\boldsymbol{r}, \boldsymbol{r}) = \int d\boldsymbol{r} \rho(\boldsymbol{r}) w_A(\boldsymbol{r}) w_B(\boldsymbol{r}) \geq 0$. However, the bond matrices are not positive by construction, and in general negative eigenvalues do occur.

The presence of negative eigenvalues in the bond matrices, disturbing at first, can be readily understood by rewriting Eq. (10) as

$$\rho_{AB}(\boldsymbol{r}, \boldsymbol{r}') = \sum_i [\psi_{ABi}^{(+)}(\boldsymbol{r})\psi_{ABi}^{(+)}(\boldsymbol{r}') - \psi_{ABi}^{(-)}(\boldsymbol{r})\psi_{ABi}^{(-)}(\boldsymbol{r}')], \tag{17}$$

where $\psi^{(\pm)}_{ABi}(\boldsymbol{r}) = \sqrt{d_i/2}[w_A(\boldsymbol{r}) \pm w_B(\boldsymbol{r})]\psi_i(\boldsymbol{r})$, and the $\psi_i(\boldsymbol{r})$ and d_i are the natural orbitals and corresponding occupancies of the molecular 1DM. The summed atomic density matrices can be rewritten similarly as

$$\rho_{AA}(\boldsymbol{r},\boldsymbol{r}') + \rho_{BB}(\boldsymbol{r},\boldsymbol{r}') = \sum_i [\psi^{(+)}_{ABi}(\boldsymbol{r})\psi^{(+)}_{ABi}(\boldsymbol{r}') + \psi^{(-)}_{ABi}(\boldsymbol{r})\psi^{(-)}_{ABi}(\boldsymbol{r}')], \tag{18}$$

To see what is going on, consider the most naive picture of covalent binding, with a fully occupied ($d_i = 2$) molecular orbital $\psi_i(\boldsymbol{r}) \approx [\phi_A(\boldsymbol{r}) + \phi_B(\boldsymbol{r})]/\sqrt{2}$ as the bonding combination of atomic orbitals $\phi_A(\boldsymbol{r})$ and $\phi_B(\boldsymbol{r})$. Assuming extreme localization properties for the weight functions (i.e. $w_A(\boldsymbol{r})\phi_B(\boldsymbol{r}) = 0$, $w_A(\boldsymbol{r})\phi_A(\boldsymbol{r}) = \phi_A(\boldsymbol{r})$) one can then interpret the corresponding $\psi^{(\pm)}_i(\boldsymbol{r}) \approx [\phi_A(\boldsymbol{r}) \pm \phi_B(\boldsymbol{r})]/\sqrt{2}$ as the bonding/antibonding combination. In this idealized situation, the summed atomic density matrices in Eq. (18) have equal occupancy in the bonding/antibonding orbitals, and the bond matrix in Eq. (17) needs to have negative eigenvalues in order to destroy the occupancy of the antibonding combination and enforce the occupation of the bonding combination in the total molecular 1DM. Note that negative eigenvalues for AB combinations also occur in the Natural Orbitals for Chemical Valence (NOCV) technique used recen tly by Ziegler and co-workers [19, 20].

C. Consistency of density matrix and electron-density partitioning

It is interesting to reflect on what would be the total electron density $\rho_A(\boldsymbol{r})$ assigned to a particular atom A in the double atomic partitioning scheme of Eq. (9). Apart from the diagonal AA density matrix, there are now also contributions from the AB bond matrices, and these can be distributed over the single-atom densities in different ways. We will study two of these, as they can be considered to be extreme cases.

In the nonweighted scheme (this could also be called a Mulliken-like method [22]), the atoms A and B receive an equal share of the density in the bond,

$$\rho^n_A(\boldsymbol{r}) = \rho_{AA}(\boldsymbol{r},\boldsymbol{r}) + \sum_{B(\neq A)} \rho_{AB}(\boldsymbol{r},\boldsymbol{r}) = \rho(\boldsymbol{r})w_A(\boldsymbol{r}). \tag{19}$$

In the weighted scheme, the atoms receive a share reflecting the balance of the weights $w_A(\boldsymbol{r})$ and

$w_B(\boldsymbol{r})$,

$$\begin{aligned}\rho_A^w(\boldsymbol{r}) &= \rho_{AA}(\boldsymbol{r},\boldsymbol{r}) + \sum_{B(\neq A)} \rho_{AB}(\boldsymbol{r},\boldsymbol{r}) \left(\frac{2w_A(\boldsymbol{r})}{w_A(\boldsymbol{r}) + w_B(\boldsymbol{r})} \right) \\ &= \rho(\boldsymbol{r}) w_A(\boldsymbol{r}) \left[\sum_B \left(\frac{2w_A(\boldsymbol{r}) w_B(\boldsymbol{r})}{w_A(\boldsymbol{r}) + w_B(\boldsymbol{r})} \right) \right]. \end{aligned} \tag{20}$$

Summing Eq. (20) over all atoms yields the molecular electron density, as it should be:

$$\begin{aligned}\sum_A \rho_A^w(\boldsymbol{r}) &= \rho(\boldsymbol{r}) \left[\sum_{AB} \left(\frac{2w_A^2(\boldsymbol{r}) w_B(\boldsymbol{r})}{w_A(\boldsymbol{r}) + w_B(\boldsymbol{r})} \right) \right] \\ &= \rho(\boldsymbol{r}) \frac{1}{2} \left[\sum_{AB} \left(\frac{2w_A^2(\boldsymbol{r}) w_B(\boldsymbol{r})}{w_A(\boldsymbol{r}) + w_B(\boldsymbol{r})} + \frac{2w_B^2(\boldsymbol{r}) w_A(\boldsymbol{r})}{w_A(\boldsymbol{r}) + w_B(\boldsymbol{r})} \right) \right] \\ &= \rho(\boldsymbol{r}) \left[\sum_{AB} \left(\frac{w_A(\boldsymbol{r}) w_B(\boldsymbol{r}) \left(w_A(\boldsymbol{r}) + w_B(\boldsymbol{r}) \right)}{w_A(\boldsymbol{r}) + w_B(\boldsymbol{r})} \right) \right] \\ &= \rho(\boldsymbol{r}) \left[\sum_{AB} w_A(\boldsymbol{r}) w_B(\boldsymbol{r}) \right] = \rho(\boldsymbol{r}). \end{aligned} \tag{21}$$

In both schemes, there is now the possibility of choosing the weights $w_A(\boldsymbol{r})$ in such a way that the single-atom density $\rho_A^{n,w}(\boldsymbol{r})$ coincides with an established AIM electron-density model of the form (6). In the nonweighted scheme, one simply takes $w_A(\boldsymbol{r}) \equiv W_A(\boldsymbol{r})$. In the weighted scheme, consistency requires that the weights obey a set of nonlinear equations,

$$\forall A: \quad w_A(\boldsymbol{r}) \left[\sum_B \left(\frac{2w_A(\boldsymbol{r}) w_B(\boldsymbol{r})}{w_A(\boldsymbol{r}) + w_B(\boldsymbol{r})} \right) \right] = W_A(\boldsymbol{r}). \tag{22}$$

In practice, we always found that the iterative sequence

$$w_A^{(0)}(\boldsymbol{r}) = W_A(\boldsymbol{r}); \quad w_A^{(i+1)}(\boldsymbol{r}) = \left\{ \frac{W_A(\boldsymbol{r})}{\left[\sum_B \left(\frac{2w_B^{(i)}(\boldsymbol{r})}{w_A^{(i)}(\boldsymbol{r}) + w_B^{(i)}(\boldsymbol{r})} \right) \right]} \right\}^{1/2} \tag{23}$$

converges rapidly (in at most 20 iterations) to a stable solution of Eq. (22) with

$$\sum_A |w_A^{(i+1)} - w_A^{(i)}| < 10^{-10} \tag{24}$$

as the convergence criterion.

It should be noted that in all numerical work in this paper the Hirshfeld-I weight functions $W_A(\boldsymbol{r})$ were taken as input; for other "fuzzy atom" prescriptions, like ISA, we expect similar

results. The Bader AIM concept is fundamentally different in that it has hard boundaries for the atomic basins. In the absence of non-nuclear attractors space is partitioned into the atomic basins, and the corresponding Bader weight function $W_A(\boldsymbol{r})$ is zero/one if $\boldsymbol{r}$ is outside/inside the basin of atom A. One can easily verify that for such binary weight functions the nonweighted and weighted schemes coincide, and there is no need to solve the nonlinear equations (22). On the other hand, the possible presence of non-nuclear attractors is problematic, as it cannot easily be reconciled with the underlying AIM picture. However, it deserves to be mentioned that with the present density matrix partitioning, the atomic electron densities from QCT remain localized to the AIM basin.

D. Expressions in a finite basis set

In practice, the partitioned density matrices are expressed in a finite basis set, as used in the molecular calculation. In the basis of the molecular natural orbitals $\psi_i(\boldsymbol{r})$, e.g., the partitioned 1DM becomes

$$\begin{aligned}(\rho_{AB})_{ij} &= \int d^3r d^3r' \psi_i(\boldsymbol{r}) \rho_{AB}(\boldsymbol{r},\boldsymbol{r'}) \psi_j(\boldsymbol{r'}) \\ &= \int d^3r d^3r' \psi_i(\boldsymbol{r}) \psi_j(\boldsymbol{r'}) \frac{1}{2}[w_A(\boldsymbol{r}) w_B(\boldsymbol{r'}) + w_B(\boldsymbol{r}) w_A(\boldsymbol{r'})] \sum_k d_k \psi_k(\boldsymbol{r}) \psi_k(\boldsymbol{r'}) \\ &= \sum_k d_k \frac{1}{2}[C^A_{ik} C^B_{jk} + C^B_{ik} C^A_{jk}]. \end{aligned} \tag{25}$$

Note that in the present calculation the $\psi_i(\boldsymbol{r})$ are just the molecular HF orbitals, and the natural occupations d_i are 2 (0) for the occupied (unoccupied) orbitals. It follows that the ρ_{AB} matrix is expressed in terms of the atomic overlap matrix elements (AOM) :

$$C^A_{ij} = \int d^3r \psi_i(\boldsymbol{r}) w_A(\boldsymbol{r}) \psi_j(\boldsymbol{r}). \tag{26}$$

The traces of the atomic and bond density matrices correspond exactly to the net and overlap populations for fuzzy atoms as defined in Eq. (6) of Ref. [11] (of course, when the same weight functions are used).

The matrix representation $(\rho_{AB})_{ij}$ in Eq. (25), in the finite single-particle space spanned by the basis set, is only equivalent to Eq. (10) in the limit of a complete basis set. But since the underlying SCF calculation is performed in the same finite basis set, it can be argued that consistent calculations actually require the use of the limited basis set expressions in Eq. (25). In a sense,

the single-particle space spanned by the finite basis set is "all there is". This holds in particular for subsequent AIM energy considerations: care should be taken that only matrix elements of the electronic Hamiltonian in the finite basis set are used, since these are the ones that fixed the electronic structure of the molecule.

For a reasonable large basis set, the differences between results obtained using Eq. (25) and Eq. (10) are small anyway, as can be suspected by the clean convergence behaviour in Table III.

III. COMPUTATIONAL METHODS

The partitioning scheme described in Sec. II was tested by partitioning the 1DM of a small set (listed in Table I) of ca. 50 simple molecules with a singlet ground state, representative of a diverse variety of chemical bonds. The 1DM was calculated at the Hartree-Fock level of theory using the Aug-cc-pVDZ basis set [23–25]. The geometry was taken from a B3LYP [26–29] /cc-pVDZ [24, 25, 30] optimization.

The scheme was implemented using atomic weight functions $W_A(\boldsymbol{r})$ from a Hirshfeld-I analysis. The weight functions $w_A(\boldsymbol{r})$ for the double atomic partitioning in Eq. (9) were constructed in both the nonweighted scheme ($w_A(\boldsymbol{r}) = W_A(\boldsymbol{r})$) and in the weighted scheme of Eq. (22). The iterative Hirshfeld weights and the adapted weights $w_A(\boldsymbol{r})$ are calculated on atom-centered grids, using a logarithmic radial grid of 100 points with $r_{\text{min}} = 10^{-6}$ Å and $r_{\text{max}} = 20$ Å , and Lebedev angular grids [31–36] with 170 points having randomized orientation on the different shells.

For CO and C_3H_3N additional calculations were performed in larger (Aug-cc-pVTZ and Aug-cc-pVQZ)[23–25] basis sets and with larger integration grids, in order to assess basis set and grid convergence.

IV. RESULTS AND DISCUSSION

A. Natural orbitals and populations

Diagonalisation of the matrices in Eq. (25) leads to the natural orbitals and occupations. As an example of the results obtained, Figs. 1-2 depict the dominant natural orbitals and occupations of the atomic density matrix ρ_{AA} of carbon and oxygen in CO, within the weighted scheme. The natural orbitals are slightly deformed versions of the typical atomic $1s$ (a), $2s$ (b), $2p_x$, $2p_y$ and $2p_z$ (c-d-e) orbitals (with the z-direction corresponding to the internuclear axis). They are clearly

$AlCl_3$	C_6H_6 (benzene)	$CHONH_2$	H_3O^+	LiF	OS_2
AlH_3	CF_4	CHOOH	HCl	LiH	SF_6
B_2H_6	CH_2NH	Cl_2	HCN	LiOH	
BeH_2	CH_2O	CO	$HCOOCH_3$	N_2	
BH_3	CH_2O_2 (dioxirane)	CO_2	HF	N_2O	
C_2H_2 (acetylene)	C_3H_8 (propane)	F_2	HNO_2	NaCl	
C_2H_4	CH_3NH_2	H_2	HNO_3	NaOH	
C_2H_6	CH_3OCH_3	H_2O	HOCl	NH_3	
C_3H_3N (acrylonitrile)	CH_3OH	H_2S	HOOH	O_2	
C_3H_4 (cyclopropene)	CH_4	H_2SO_4	Li_2	O_3	

TABLE I: List of molecules in the test set.

localized and have populations between 0 and 2. Note that for these main contributions the shape of the orbitals in the nonweighted scheme is visually indistinguishable from those presented in Figs. 1-2. The orbitals that represent the core $1s$ atomic orbitals have a population well below 2.00 (the expected value of an orbital not involved in bonding), resp. 1.84 in carbon and 1.90 in oxygen. The rather low eigenvalues for the essentially core orbitals are less pleasing from a chemical point of view. The occupations in the nonweighted scheme are quite different, and are indicated between brackets in Figs. 1-2. In the nonweighted scheme, the populations are more in accordance with the full occupation expected of core orbitals, i.e. 1.96 and 1.99 for the carbon and oxygen $1s$ orbitals. As discussed later, this difference between both schemes is typical, as the weighted scheme tends to assign a larger fraction of electrons to the bond matrices than the nonweighted scheme. Hence the nonweighted scheme has invariably larger populations for the atomic density matrix natural orbitals. The $2s$ orbitals have shifted away from the bonding region, to accomodate the free electron pair on C and O in the CO molecule. However, they are not fully occupied. The $2p_x$, $2p_y$ and $2p_z$ orbitals (not expected to be fully occupied) are shifted to the bonding region.

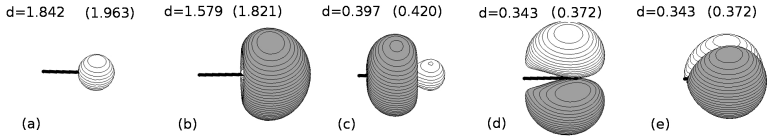


FIG. 1: The dominant natural orbitals, and the corresponding occupations, of the atomic density matrix of carbon in a CO molecule calculated in the weighted scheme, at the HF/Aug-cc-pVDZ level. The occupations in the nonweighted scheme are also indicated, between brackets.

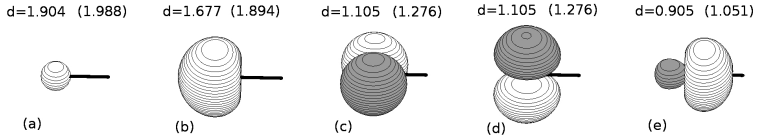


FIG. 2: The dominant natural orbitals, and the corresponding occupations, of the atomic density matrix of oxygen in a CO molecule calculated in the weighted scheme, at the HF/Aug-cc-pVDZ level. The occupations in the nonweighted scheme are also indicated, between brackets.

Figure 3 depicts the naturals and occupations of the CO bond matrix ρ_{AB} in the CO molecule. Only the main contributions are shown. Both schemes result in natural orbital shapes that are very similar. There are mainly typical bonding orbitals σ (a), π_1, π_2 (b-c) with positive eigenvalues, and antibonding orbitals σ^* (h), π_1^*, π_2^* (f-g) with negative eigenvalues. The negative eigenvalues delete antibonding contributions of the $\rho_{AA} + \rho_{BB}$ matrix, whereas the positive eigenvalues reinforce its bonding contributions. The picture also shows two rather "nonbonding" orbitals (d-e) with positive eigenvalues. These serve to reinforce the population of the free electron pairs on carbon and oxygen (the shifted $2s$ orbital on these atoms). Note that in the weighted scheme orbitals (i-j) appear which correspond to the $1s$ core orbitals on C and O and have a nonnegligible (0.17 - 0.10) occupation in the bond matrix.

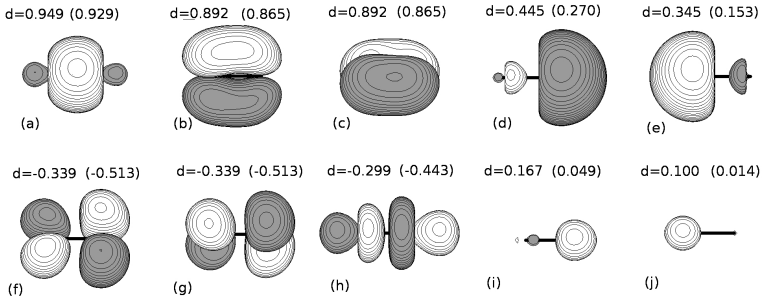


FIG. 3: The dominant eigenvectors and occupations of (twice) the CO bond matrix in a CO molecule calculated in the weighted scheme, at the HF/Aug-cc-pVDZ level. The occupations in the nonweighted scheme are also indicated, between brackets.

Apart from the dominant contributions shown in Figs. 1-3, various orbitals with much smaller populations are also present (as the Aug-cc-pVDZ molecular basis set has 46 basis functions). A complete overview is given in Table II. Note that for the atomic density matrices, apart from the five dominant natural orbitals, only two more have small ($\sim 10^{-2} - 10^{-3}$), while the remainder have vanishing ($< 10^{-13}$) populations. For the bond matrix, apart from the ten dominant orbitals, only four more have small occupations. This can be understood from the fact that the molecule is treated at the HF level, with only 7 (double occupied) spatial orbitals. One can rewrite the atomic density matrix as

$$\begin{aligned} \rho_{AA}(\boldsymbol{r},\boldsymbol{r'}) &= \sum_{i=1}^{N/2} \left[w_A(\boldsymbol{r})\psi_i(\boldsymbol{r})\right]\left[w_A(\boldsymbol{r'})\psi_i(\boldsymbol{r'})\right] \\ &= \sum_{jj'=1}^{N/2} T^A_{jj'}\widetilde{\psi_j}(\boldsymbol{r})\widetilde{\psi_{j'}}(\boldsymbol{r'}), \end{aligned} \tag{27}$$

where T^A is the overlapmatrix of the nonorthogonal basis functions $[w_A(\boldsymbol{r})\psi_i(\boldsymbol{r})]$

$$T^A_{ij} = \int w_A^2(\boldsymbol{r})\psi_i(\boldsymbol{r})\psi_j(\boldsymbol{r})\, d^3\boldsymbol{r} \tag{28}$$

and the $\widetilde{\psi_j}(\boldsymbol{r})$ form a set of orthonormal basis functions

$$\widetilde{\psi_j}(\boldsymbol{r}) = \sum_i (T^{A-\frac{1}{2}})_{ij}\left[w_A(\boldsymbol{r})\psi_i(\boldsymbol{r})\right]. \tag{29}$$

	C , C	O , O	C , O
label	eigenvalue	eigenvalue	eigenvalue
(a)	1.842 (1.963)	1.904 (1.988)	0.474 (0.464)
(b)	1.579 (1.821)	1.677 (1.894)	0.446 (0.432)
(c)	0.397 (0.420)	1.105 (1.276)	0.446 (0.432)
(d)	0.343 (0.372)	1.105 (1.276)	0.223 (0.135)
(e)	0.343 (0.372)	0.905 (1.051)	0.173 (0.076)
(f)	0.023 (0.008)	0.042 (0.021)	-0.170 (-0.256)
(g)	0.002 (0.000)	0.005 (0.001)	-0.170 (-0.256)
(h)	$< 10^{-13}$	$< 10^{-13}$	-0.149 (-0.221)
(i)	$\vdots$	$\vdots$	0.084 (0.024)
(j)			0.050 (0.007)
$\vdots$			-0.034 (-0.061)
			-0.007 (-0.007)
			-0.001 (-0.002)
			-0.000 (-0.000)
			$< 10^{-13}$
			$\vdots$
Sum	4.528 (4.955)	6.743 (7.508)	1.364 (0.768)

TABLE II: All natural populations in the atomic density matrices (CC and OO) and bond matrix (CO) in a CO molecule calculated in the weighted scheme, at the HF/Aug-cc-pVDZ level. The occupations in the nonweighted scheme are indicated between brackets. The labels in the first column correspond (for the dominant orbitals) to the labels in Figs. 1-2.

It is clear that $\rho_{AA}(\boldsymbol{r}, \boldsymbol{r'})$ is a matrix of rank $N/2$, since diagonalisation of the $N/2 \times N/2$ overlap-matrix T^A will yield $N/2$ eigenvalues. The same reasoning holds for the bond matrix $\rho_{AB}(\boldsymbol{r}, \boldsymbol{r'})$, but now starting from the 14 (linear independent) functions $\psi^{(\pm)}_{ABi}(\boldsymbol{r})$ defined in Eq. (17-18) with $d_i = 2$ for $i = 1, 2, ...N/2$.

The difference between the nonweighted and weighted scheme is again clear from Table II. The differences are sizeable, with consistently smaller populations in the weighted scheme, resulting in

traces for the atomic density matrices that are significantly (0.4 - 0.7) smaller. This is compensated for by larger eigenvalues in the bond matrix in the weighted scheme, for the orbitals not involved in bonding. The bonding orbitals (a-b-c) in the bond matrix, however, are about equally populated in both schemes.

CO	grid=100-170 Aug-cc-pVDZ	grid=500-590 Aug-cc-pVDZ	grid=100-170 Aug-cc-pVTZ	grid=100-170 Aug-cc-pVQZ
A , B	Tr($\rho_{A,B}$)	Tr($\rho_{A,B}$)	Tr($\rho_{A,B}$)	Tr($\rho_{A,B}$)
C , C	4.528 (4.955)	4.528 (4.955)	4.529 (4.956)	4.531 (4.958)
O , O	6.743 (7.508)	6.743 (7.508)	6.751 (7.516)	6.752 (7.518)
C , O	1.364 (0.768)	1.364 (0.768)	1.360 (0.764)	1.358 (0.762)

TABLE III: Number of electrons present in the atomic density and bond matrices for CO, in the non-weighted and weighted scheme. Bracketed values correspond to the nonweighted scheme. The second column contains the results of the default calculation. In the third column the number of radial and angular grid points is increased to 500 and 590, respectively. In the fourth column a triple-, rather than double-ζ basis is used, in the fifth colum a quadruple-ζ.

In Table III the stability of the proposed density matrix partitioning is examined. The results for the matrix traces are clearly converged as far as grid size is concerned, with deviations of less than 0.001 in electron number. The results also seem to be quite stable with respect to basis set size, with differences less than 0.01 going from DZ to TZ, and less than 0.002 going from TZ to QZ.

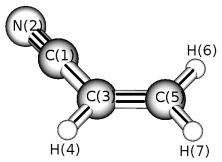


FIG. 4: Topology of the C_3H_3N (acrylonitrile) molecule.

As a second example we analyze some results obtained for the density matrix partitioning in C_3H_3N (acrylonitrile). The topology of the molecule is presented in figure 4. Figure 5 shows the

dominant eigenvectors and occupations of the CN bond matrix. The features are very similar to those of the CO bond matrix. The bond orbitals σ (a), π_1 (b) and π_2 (c) are about equally populated in both schemes. The nonbonding orbital (d) reinforces the population of the free electron pair on nitrogen. The antibonding orbitals π_1^* (e), π_2^* (f), and σ^* (g) have negative eigenvalues that delete antibonding contributions of the $\rho_{C(1)C(1)}+\rho_{N(2)N(2)}$ matrix. Small but nonnegligible contributions exist that are complementary to the 1s core electrons on C(1) and N(2) (h and j). All contributions that correct the sum of the atomic density matrices (by deleting antibonding parts and reinforcing nonbonding parts) are significantly larger within the weighted scheme.

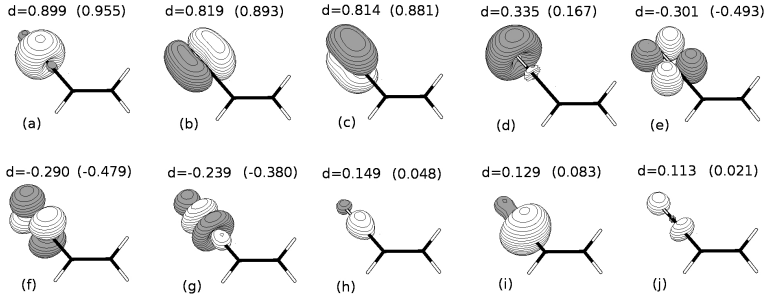


FIG. 5: The dominant eigenvectors and occupations of the CN bond matrix in a C_3H_3N molecule calculated in the weighted scheme, at the HF/Aug-cc-pVDZ level. The occupations in the nonweighted scheme are also indicated, between brackets.

Figure 6 shows the dominant eigenvectors and occupations of the C(3)-C(5) bond matrix. As it is formally a double bond, large bonding orbitals σ (a) and π_1 (b) are expected, next to some smaller antibonding contributions σ^* (d) and π_1^* (e). There are a lot of small contributions not shown here involving the $2s$ orbitals on both carbons, the σ_{C-H} bonds and the $1s$ orbitals on both carbons, that are complementary to the main orbitals of other atomic and bond matrices. However, beyond the many small contributions, there is one considerably larger than the others: a π_2 (c) bond. It's important to mention that this is a rather general feature, also noticed for example in the CH bond matrices of CH_4, where both π_1 and π_2 contributions are important, although there are no π bonds in the CH_4 molecule. Notice that the trace within the weighted scheme is still larger than the one within the nonweighted scheme, while the bonding orbitals of the weighted scheme have a significantly lower occupation than that of the nonweighted scheme.

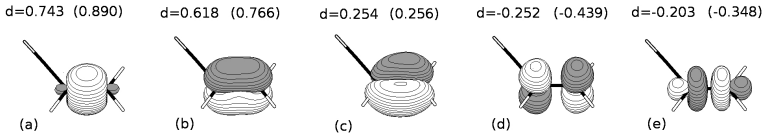


FIG. 6: The dominant eigenvectors and occupations of the C(3)-C(5) bond matrix in a C_3H_3N molecule calculated in the weighted scheme, at the HF/Aug-cc-pVDZ level. The occupations in the non-weighted scheme are also indicated, between brackets.

Figure 7 shows the dominant eigenvectors and occupations of the C(1)-C(3) bond matrix. As it is formally a single bond, a large bonding orbital σ (a) is expected, next to a smaller antibonding contribution σ^* (d). As in the previous case of the C(3)-C(5) double bond, there are a lot of smaller contributions, with two of them predominant: a π_1 (b) and a π_2 (c) orbital. Again, the trace within the weighted scheme is larger than the trace within the nonweighted scheme, while the bonding orbitals have a significantly lower population.

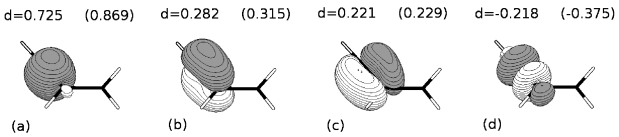


FIG. 7: The dominant eigenvectors and occupations of the C(1)-C(3) bond matrix in a C_3H_3N molecule calculated in the weighted scheme, at the HF/Aug-cc-pVDZ level. The occupations in the non-weighted scheme are also indicated, between brackets.

B. Correlation with shared electron density indices

A global test for the partitioning scheme is the evaluation of the total population in its bond matrices. This population should correlate somehow with the bond order. The classical definition of bond order equals it to the electronic occupancy of the bonding orbitals minus that of the antibonding orbitals divided by two. At the Hartree-Fock level of theory, there is a remarkable similarity between these classical bond orders and the results of so-called shared electron density indices

(SEDI) [11, 37–45]. SEDI are obtained from integration of the exchange density at the Hartree-Fock level or exchange-correlation density at correlated levels of theory. The exchange-correlation density reads

$$\rho^{xc}\left(\boldsymbol{r},\boldsymbol{r'}\right)=\rho\left(\boldsymbol{r}\right)\rho\left(\boldsymbol{r'}\right)-\sum_{\sigma\sigma'}\rho^{(2)}\left(\boldsymbol{r}\sigma,\boldsymbol{r'}\sigma';\boldsymbol{r}\sigma,\boldsymbol{r'}\sigma'\right) \tag{30}$$

where the diagonal elements of the second-order density matrix $\rho^{(2)}$ appear. At the single-determinant level, this simplifies to:

$$\rho^{xc}\left(\boldsymbol{r},\boldsymbol{r'}\right)=\frac{1}{2}\rho\left(\boldsymbol{r},\boldsymbol{r'}\right)\rho\left(\boldsymbol{r'},\boldsymbol{r}\right) \tag{31}$$

Integrating $\boldsymbol{r}$ and $\boldsymbol{r'}$ over the atomic domains of atoms A and B and multiplying by two to account for the symmetrical integration over B and A eventually results in:

$$SEDI\left(A,B\right)=4\sum_{i,j}S_{ij}^{A}S_{ji}^{B} \tag{32}$$

where i and j are occupied molecular orbitals, and S_{ij}^{A} and S_{ij}^{B} are the atom condensed overlaps. In the case that the atomic domains are defined by Hirshfeld-I, the atom condensed overlaps reduce to an expression formally very similar to Eq. (26).

$$S_{ij}^{A}=\int d^{3}r\psi_{i}(\boldsymbol{r})W_{A}(\boldsymbol{r})\psi_{j}(\boldsymbol{r}) \tag{33}$$

The main advantage of SEDI is that these can be computed also between non-covalently bonded atoms, where the classical expression can no longer be used. Given the classical expression for bond order that uses occupancies of bonding and antibonding orbitals, an intriguing question is whether the trace of the bond matrices would also give similar results. In order to answer this question not only for covalently bonded atoms but in general, this section examines the correlation between the trace of bond matrices and the SEDI, both computed starting from the Hirshfeld-I analysis. For the SEDI, the S_{ij}^{A} were constructed using the Hirshfeld-I weights $W_{A}(\boldsymbol{r})$ in Eq. (33). The Hirshfeld-I weights were also used to calculate the atomic overlap matrices C_{ij}^{A} of Eq.(26) for the nonweighted scheme. The weights $w_{A}(\boldsymbol{r})$ of the nonlinear equations of (22) were used to build the atomic overlap matrices for the weighted scheme.

	$\frac{1}{2}SEDI(A,A)$	Tr ρ^w_{AA}	Tr ρ^n_{AA}
C $_1$, C $_1$	3.415	3.635	4.296
N $_2$, N $_2$	5.826	5.822	6.604
C $_3$, C $_3$	3.881	3.829	4.692
H $_4$, H $_4$	0.272	0.392	0.480
C $_5$, C $_5$	4.038	3.909	4.798
H $_6$, H $_6$	0.279	0.400	0.487
H $_7$, H $_7$	0.282	0.405	0.490
	$SEDI(A,B)$	2 Tr ρ^w_{AB}	2 Tr ρ^n_{AB}
C $_1$, N $_2$	**2.871**	**2.397**	**1.615**
C $_1$, C $_3$	**1.222**	**1.294**	**0.907**
C $_1$, H $_4$	0.084	0.148	0.038
C $_1$, C $_5$	0.197	0.261	0.076
C $_1$, H $_6$	0.021	0.055	0.012
C $_1$, H $_7$	0.014	0.015	0.001
N $_2$, C $_3$	0.253	0.319	0.065
N $_2$, H $_4$	0.024	0.040	0.004
N $_2$, C $_5$	0.087	0.091	0.009
N $_2$, H $_6$	0.010	0.024	0.003
N $_2$, H $_7$	0.007	0.006	0.000
C $_3$, H $_4$	**0.914**	**0.852**	**0.663**
C $_3$, C $_5$	**1.927**	**1.644**	**1.228**
C $_3$, H $_6$	0.133	0.181	0.050
C $_3$, H $_7$	0.137	0.190	0.053
H $_4$, C $_5$	0.132	0.182	0.050
H $_4$, H $_6$	0.011	0.012	0.000
H $_4$, H $_7$	0.013	0.037	0.007
C $_5$, H $_6$	**0.942**	**0.875**	**0.673**
C $_5$, H $_7$	**0.947**	**0.883**	**0.676**
H $_6$, H $_7$	0.059	0.100	0.023
sum	28.000	28.000	28.000

TABLE IV: The SEDI index (second column) versus the traces of the ρ_{AB} matrices in the weighted scheme (third column) and the nonweighted scheme (fourth column) for the C_3H_3N (acrylonitrile) molecule. AB contributions corresponding to bonded atoms in the Lewis structure are in boldface.

Table IV shows a complete comparison between the SEDI index and the matrix traces within both the weighted and the nonweighted scheme for the C_3H_3N (acrylonitrile) molecule. It appears that for the triple, double and single bonds present in the molecule (indicated in boldface), the SEDI index is quite close to the classical bond order. The matrix traces of the weighted scheme are usually somewhat lower, while the matrix traces of the nonweighted scheme are much lower (in some cases more than 45 percent). For pairs of atoms that have no classical bond between them,

the matrix traces of the weighted scheme are noneglegible and generally somewhat larger than the SEDI indices, whereas the matrix traces of the nonweighted scheme are quite small (< 0.08). The largest SEDI indices and matrix traces for nonbonded atom pairs can be found in allylic places.

For the test set of table I it was investigated how well the nondiagonal Shared Electron Distribution Index (SEDI) correlates with the matrix traces within both the weighted and nonweighted scheme. A correlation conceptually analogous was first given by I. Mayer for the Mulliken case [46]. The correlation plots are shown in Fig. 8. For both schemes there is a strong linear correlation ($R^2 > 0.96$), but the proportionality factor is surprisingly close to unity for the weighted scheme (0.97). By comparison, the slope for the nonweighted scheme is much smaller (0.60). In the case of Hirshfeld-I and the weighted scheme, the SEDI index and the "overlap population" seem to give largely equivalent values $SEDI(A,B) \sim 2Tr(\rho^w_{AB})$. In view of the extent (50 molecules) and diversity of the test set, this can hardly be coincidental. One may argue that this is somewhat skewed because of the large number of classically nonbonded atom pairs, for which both measures are close to zero. However, restricting to the bonded atom pairs the proportionality factors hardly change (0.95 for the weighted, 0.59 for the nonweighted scheme). This is all the more surprising, as one would intuitively expect a closer relation between $SEDI(A,B)$ and $2Tr(\rho^n_{AB})$ since the weight functions are the same (i.e. $w_A(\boldsymbol{r} = W_A(\boldsymbol{r}))$ for the nonweighted scheme.

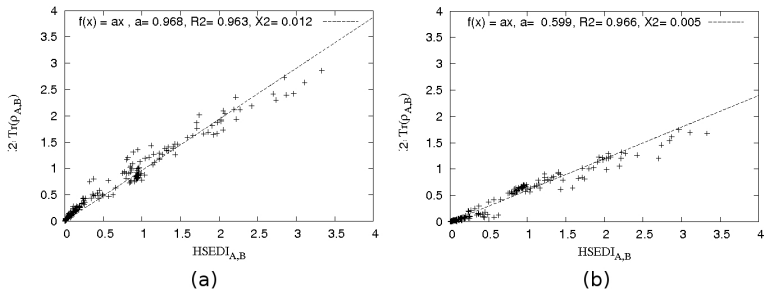


FIG. 8: Correlation between the SEDI index and the traces of the bond matrices for the weighted scheme (a) and the nonweighted scheme (b)

The fact that the slope of the correlation plot is larger in the weighted scheme, can be traced back to the fact that the bond matrices ρ_{AB} have a larger trace in the weighted than in the nonweighted scheme, as is evident from Table IV. The opposite holds for the trace of the atomic den-

sity matrices ρ_{AA}: these are larger in the nonweighted scheme than in the weighted scheme. Both observations can be explained by the Hirshfeld-I weight functions $W_A(\boldsymbol{r})$ to be more strongly localized around atom A than the weight function $w_A(\boldsymbol{r})$ resulting from the nonlinear equations (22). This is indeed what is found numerically, and it can also be understood by analyzing Eq. (22): in the vicinity of atom A, $w_A(\boldsymbol{r})$ dominates over the weight functions $w_B(\boldsymbol{r})$ of all other atoms. Replacing $w_B(\boldsymbol{r})$ in the denominator of Eq. (22) by $w_A(\boldsymbol{r}) > w_B(\boldsymbol{r})$ therefore leads to

$$W_A(\boldsymbol{r}) = \sum_B \frac{2w_A^2(\boldsymbol{r})w_B(\boldsymbol{r})}{w_A(\boldsymbol{r}) + w_B(\boldsymbol{r})} > \sum_B \frac{2w_A^2(\boldsymbol{r})w_B(\boldsymbol{r})}{w_A(\boldsymbol{r}) + w_A(\boldsymbol{r})} = w_A(\boldsymbol{r}). \tag{34}$$

So the Hirshfeld-I weights $W_A(\boldsymbol{r})$ are more localized on the individual atoms and will lead to a smaller number of shared electrons.

From the preceding discussion it's clear that there seems to be some freedom in choosing an appropriate scheme with corresponding atom weights: one can increase the number of electrons in the overlapmatrix $\rho_{AB}(\boldsymbol{r}, \boldsymbol{r'})$ and get it even close to the SEDI index, at the risk of including some parts (e.g. the core $1s$ electrons) that do not belong there.

V. CONCLUSIONS

We have introduced a succesful approach to partitioning of a molecular density matrix in constituent atomic and bond contributions. The partitioning is such that the density matrices and electron densities for the atoms in the molecule are mutually consistent, and follows from requiring that the atom and bond orbitals, as eigenfunctions of the corresponding density matrices, are strictly localized. This prevents the use of single atom density matrix partitioning as the molecular density matrix is inherently delocalized.

The atomic density matrices correspond to diagonal elements $\rho_{AA}(\boldsymbol{r}, \boldsymbol{r})$, whereas bond matrices correspond to the off-diagonal contributions $\rho_{AB}(\boldsymbol{r}, \boldsymbol{r})$. Only in the case of the diagonal elements are the eigenvalues restricted to the interval [0,2]. For the bond matrices, negative eigenvalues can and do occur.

The weight functions used for partitioning the molecular density matrix were constructed in two different ways: using either directly the weights produced from a regular atoms-in-molecules (AIM) density based theory (here Hirshfeld-I), or using a weighted scheme. In both cases the AIM density derived from the density matrix is equal to the one obtained directly from the AIM algorithm.

The trace of the bond density matrices was suggested as a useful source of bond indices, loosely related to classical bond orders. A remarkably good correlation with shared electron density indices (SEDI) was found, establishing the chemical relevance of the density matrix partitioning. The traces of the bond density matrices can in this way be used quite effectively to characterize chemical bonds without requiring second-order density matrices.

It should be mentioned that while the present analysis is restricted to spin singlet molecules, an extension to higher-spin ($S > 0$) molecular states seems entirely possible, be applying the decomposition in Eq. (10) to the spin up and down electron density separately,

$$\rho^{\sigma}_{AB}(\boldsymbol{r},\boldsymbol{r'}) = \frac{1}{2}\left(w_A(\boldsymbol{r})w_B(\boldsymbol{r'}) + w_B(\boldsymbol{r})w_A(\boldsymbol{r'})\right)\rho^{\sigma}(\boldsymbol{r},\boldsymbol{r'}) = \rho^{\sigma}_{BA}(\boldsymbol{r},\boldsymbol{r'}) \tag{35}$$

where the atomic weight functions are taken as spin independent. Certainly for Hirshfeld-like schemes this seems the most natural choice, as the spin-state of the isolated atoms building up the weight functions is washed away when these are considered as molecular constituents. Note that even for the simpler problem of molecular electron density partitioning, spin considerations have hardly been studied and should be explored further.

VI. ACKNOWLEDGEMENTS

We acknowledge support from Dr. Toon Verstraelen (CMM) who provided us with a Hirshfeld-I program and from Dieter Ghillemijn (Phd, Ghent University) for comparative calculations. DVF acknowledges support from the research council (BOF) of Ghent University and FWO-Vlaanderen. PWA acknowledges support from Sharcnet, NSERC, and the Canada Research Chairs.

[1] R. S. Mulliken. Electronic population analysis on lcao-mo molecular wave functions. 1. *Journal of Chemical Physics*, 23(10):1833–1840, 1955.

[2] Richard F. W. Bader. *Atoms in Molecules: A Quantum Theory (The International Series of Monographs on Chemistry, No 22)*. Oxford University Press, USA, 1994.

[3] Richard F. W. Bader. A quantum theory of molecular structure and its applications. *Chemical Reviews*, 91(5):893 – 928, 1991.

[4] P. Popelier. *Atoms In Molecules: An Introduction*. Prentice-Hall: Harlow, Great Britain, 2000.

[5] F. L. Hirshfeld. Bonded-atom fragments for describing molecular charge densities. *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)*, 44(2):129–138, 1977.

[6] Patrick Bultinck, Christian Van Alsenoy, Paul W. Ayers, and Ramon Carb-Dorca. Critical analysis and extension of the hirshfeld atoms in molecules. *Journal of Chemical Physics*, 126(14):144111, 2007.

[7] Patrick Bultinck. General discussion. *Faraday Discussions*, 135 - Chemical Concepts from Quantum Mechanics:244–246, 2007.

[8] Patrick Bultinck, Paul W. Ayers, Stijn Fias, Koen Tiels, and Christian Van Alsenoy. Uniqueness and basis set dependence of iterative hirshfeld charges. *Chemical Physics Letters*, 444(1-3):205 – 208, 2007.

[9] Timothy C. Lillestolen and Richard J. Wheatley. Redefining the atom: atomic charge densities produced by an iterative stockholder approach. *Chemical Communications*, 45:5909–5911, 2008.

[10] P. Bultinck, D. L. Cooper, and D. Van Neck. Comparison of the hirshfeld-i and iterated stockholder atoms in molecules schemes. *Physical Chemistry Chemical Physics*, 11(18):3424–3429, 2009.

[11] I. Mayer and P. Salvador. Overlap populations, bond orders and valences for 'fuzzy' atoms. *Chemical Physics Letters*, 383(3-4):368 – 375, 2004.

[12] Lemin Li and Robert G. Parr. The atom in a molecule: A density matrix approach. *Journal of Chemical Physics*, 84(3):1704–1711, 1986.

[13] J. Cioslowski and J. Karwowski. *Fundamentals of Molecular Similarity*. Kluwer Academic, Dordrecht, 2001.

[14] Diego R. Alcoba, Luis Lain, Alicia Torre, and Roberto C. Bochicchio. A study of the partitioning of the first-order reduced density matrix according to the theory of atoms in molecules. *Journal of Chemical Physics*, 123(14):144113, 2005.

[15] Diego R. Alcoba, Roberto C. Bochicchio, Alicia Torre, and Luis Lain. Decomposition of the first-order reduced density matrix: An isopycnic localization treatment. *Journal of Physical Chemistry A*, 110(29):9254–9260, 2006.

[16] Diego R. Alcoba, Roberto C. Bochicchio, Luis Lain, and Alicia Torre. Covalent bond indices and ionicities from similarity measures. *Chemical Physics Letters*, 442(1-3):157 – 163, 2007.

[17] I. Mayer and P. Salvador. Effective atomic orbitals for fuzzy atoms. *Journal of Chemical Physics*, 130(23):234106, 2009.

[18] R. F. Nalewajski and J. Mrozek. Modified valence indices from the two-particle density matrix. *Inter-*

national Journal of Quantum Chemistry, 51(4):187–200, 1994.

[19] Artur Michalak, Roger L. DeKock, and Tom Ziegler. Bond multiplicity in transition-metal complexes: Applications of two-electron valence indices. *Journal of Physical Chemistry A*, 112(31):7256–7263, 2008.

[20] Artur Michalak, Mariusz Mitoraj, and Tom Ziegler. Bond orbitals from chemical valence theory. *Journal of Physical Chemistry A*, 112(9):1933–1939, 2008.

[21] I. Mayer. A. Hamza Atomic decomposition of identity: General formalism for population analysis and energy decomposition. *International Journal of Quantum Chemistry*, 103:798, 2005.

[22] A. Martìn Pendàs, M.A. Blanco, and E Francisco. Chemical fragments in real space: definitions, properties, and energetic decompositions. *Journal of Computational Chemistry*, 28:161–184, 2007.

[23] R. A. Kendall, T.H. Dunning, and R. J. Harrison. Electron affinities of the first-row atoms revisited. systematic basis sets and wave functions. *Journal of Chemical Physics*, 96:6796–6806, 1992.

[24] D. E. Woon and T.H. Dunning. Gaussian basis sets for use in correlated molecular calculations. iii. the second row atoms, al-ar. *Journal of Chemical Physics*, 98:1358–1371, 1993.

[25] D.E. Woon, K.A. Peterson, and T.H. Dunning. Gaussian basis sets for use in correlated molecular calculations. vii. valence and core-valence basis sets for li, na, be, and mg. in preparation.

[26] A.D. Becke. Density-functional thermochemistry. iii. the role of exact exchange. *Journal of Chemical Physics*, 98:5648–5652, 1993.

[27] C. Lee, W. Yang, and R. G. Parr. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Physical Review B*, 37:785–789, 1988.

[28] S.H. Vosko, L. Wilk, and M. Nusair. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Canadian Journal of Physics*, 58:1200–1211, 1980.

[29] P.J. Stephens, F.J. Devlin, C.F. Chabalowski, and M.J. Frisch. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *Journal of Physical Chemistry*, 98:11623–11627, 1994.

[30] T.H. Dunning. Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen. *Journal of Chemical Physics*, 90:1007–1023, 1989.

[31] V. I. Lebedev. Values of the nodes and weights of ninth to seventeenth order gauss-markov quadrature formulae invariant under the octahedron group with inversion". *Computational Mathematics and Mathematical Physics*, 15:44–51, 1975.

[32] V. I. Lebedev. Quadratures on a sphere. *Computational Mathematics and Mathematical Physics*,

16:10–24, 1976.

[33] V. I. Lebedev. Spherical quadrature formulas exact to orders 25-29. *Siberian Mathematical Journal*, 18:99–107, 1977.

[34] V. I. Lebedev and A. L. Skorokhodov. Quadrature formulas of orders 41, 47, and 53 for the sphere. *Russian Academy of Sciences. Doklady. Mathematics*, 45:587–592, 1992.

[35] V. I. Lebedev. A quadrature formula for the sphere of 59th algebraic order of accuracy. *Russian Academy of Sciences. Doklady. Mathematics*, 50:283–286, 1995.

[36] V. I. Lebedev and D. N. Laikov. Quadrature formula for the sphere of 131-th algebraic order of accuracy. *Doklady Mathematics*, 59(3):477–481, 1999.

[37] K. B. Wiberg. Application of the pople-santry-segal cndo method to the cyclopropylcarbinyl and cyclobutyl cation and to bicyclobutane. *Tetrahedron*, 24(3):1083 – 1096, 1968.

[38] R.F.W. Bader, M.E. Stephens. Spatial localization of the electronic pair and number distributions in molecules. *Journal of the American Chemical Society*, 97:7391, 1975.

[39] M. Giambiagi, M. S. Giambiagi, D. R. Grempel, and C. D. J. Heymann. Definition of ave bond index with non-ortogonal basis - properties and applications. *Journal de Chimie Physique et de Physico-chimie Biologique*, 72:15–22, 1975.

[40] I. Mayer. Charge, bond order and valence in the ab initio scf theory. *Chemical Physics Letters*, 97(3):270 – 274, 1983.

[41] Robert L. Fulton. Sharing of electrons in molecules. *Journal of Physical Chemistry*, 97(29):7516–7529, 1993.

[42] J.G. Angyan, M. Loos, I. Mayer. Covalent bond orders and atomic valence indices in the topological theory of atoms in molecules. *Journal of Physical Chemistry*, 98:5244, 1994.

[43] Xavier Fradera, Maggie A. Austen, and Richard F. W. Bader. The lewis model and beyond. *Journal of Physical Chemistry A*, 103(2):304–314, 1999.

[44] I. Mayer. Bond order and valence indices: a personal account. *Journal of Computational Chemistry*, 28:204–221, 2007.

[45] Robert Ponec and David L. Cooper. Anatomy of bond formation. bond length dependence of the extent of electron sharing in chemical bonds. *Journal of Molecular Structure: THEOCHEM*, 727(1-3):133 – 138, 2005. Dedicated to Ramon Carbo-Dorca on the Occasion of his 65th Birthday.

[46] I. Mayer. Bond order and valence: relations to Mulliken's population analysis. *International Journal of Quantum Chemistry*, 26:151, 1984.

Paper II

Communication: Hilbert-space partitioning of the molecular one-electron density matrix with orthogonal projectors

D. Vanfleteren, D. Van Neck, P. Bultinck, P. W. Ayers, and M. Waroquier

The Journal of Chemical Physics 133, 231103 (2010)

HILBERT-SPACE PARTITIONING OF THE MOLECULAR ONE-ELECTRON DENSITY MATRIX WITH ORTHOGONAL PROJECTORS

Diederik Vanfleteren,[1,2] Dimitri Van Neck,[1,2] Patrick Bultinck,[1,3] Paul W. Ayers,[4] and Michel Waroquier[1,2]

[1]*Members of the Ghent Brussels quantum chemistry and molecular modelling alliance.*

[2]*Ghent University, Center for Molecular Modeling, Technologiepark 903, B-9052 Zwijnaarde, Belgium*

[3]*Ghent University, Department of Inorganic and Physical Chemistry, Krijgslaan 281 (S3), B-9000 Gent, Belgium*

[4]*McMaster University, Department of Chemistry, Hamilton, Ontario L8S 4M1, Canada*

(Dated: June 19, 2011)

Abstract

A double-atom partitioning of the molecular one-electron density matrix is used to describe atoms and bonds. All calculations are performed in Hilbert space. The concept of atomic weight functions (familiar from Hirshfeld analysis of the electron density) is extended to atomic weight matrices. These are constructed to be orthogonal projection operators on atomic subspaces, which has significant advantages in the interpretation of the bond contributions. In close analogy to the iterative Hirshfeld procedure, self-consistency is built in at the level of atomic charges and occupancies. The method is applied to a test set of about 67 molecules, representing various types of chemical binding. A close correlation is observed between the atomic charges and the Hirshfeld-I atomic charges.

I. INTRODUCTION

The concept of Atoms In the Molecule (AIM) has always played a central role in classifying and predicting chemical properties. Because this concept does not naturally show up in molecular orbital (MO) theory, there is a sustained interest in extracting chemical atoms and functional groups from MO-based calculations. Among the most popular techniques are Mulliken population analysis [1], Bader's Quantum Chemical Topology (QCT) [2–4], the Hirshfeld method in its original [5] and iterative version [6–9], the iterative stockholder approach [10], and Mayer's fuzzy atoms [11].

Most methods are restricted to the partitioning of the electron density, but not all properties of a quantum mechanical object can be explicitly expressed in terms of the electron density. A more fundamental approach to the AIM should be based on density matrices [12]. Because of the inherent non-locality of the density matrix, it was recently argued that a two-index partitioning into atomic (diagonal) and bond (off-diagonal) contributions, is necessary to guarantee the local nature and transferability of the atoms [13, 14]. The authors recently introduced such an approach [14], but the bond matrices had significant contributions from core electrons and free electron pairs, which somewhat blurred the interpretation in terms of chemical bonding. In this paper we attempt to improve the description of atoms and bonds in a molecule. The interpretive problems are overcome by defining atomic weight matrices to be orthogonal projection operators onto one-electron subspaces assigned to atoms. The proposed method bears a close resemblance to the iterative Hirshfeld procedure, as we introduce the requirement of consistency between weight matrices and atomic contributions.

II. DOUBLE ATOM PARTITIONING SCHEME

The spin-summed one-electron density matrix (1DM) for a singlet N-electron molecular wavefunction $\Psi(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$ is expressed as

$$\rho_{ij} = N \int d\boldsymbol{r} d\boldsymbol{r'} \varphi_i(\boldsymbol{r}) \varphi_j(\boldsymbol{r'}) \sum_\sigma \left[\int d\boldsymbol{x}_2 ... \int d\boldsymbol{x}_N \Psi^\dagger(\boldsymbol{r'}\sigma, \boldsymbol{x}_2, ...\boldsymbol{x}_N) \Psi(\boldsymbol{r}\sigma, \boldsymbol{x}_2, ...\boldsymbol{x}_N) \right], \tag{1}$$

where the matrix representation in an orthogonal basis set $\varphi_i(\boldsymbol{r})$ was used. A decomposition of the identity matrix into atomic weight matrices [15] $(w_A)_{ij}$

$$\delta_{ij} = \sum_A (w_A)_{ij}, \tag{2}$$

can be inserted on both sides of the molecular one-electron density matrix to arrive at a double-index atomic partitioning

$$\rho = \sum_{AB} \rho_{AB} = \sum_{AB} w_A \rho w_B. \tag{3}$$

The characteristics of such a partitioning are spatially localized contributions ρ_{AA} belonging to an atom A and some delocalized contributions $(\rho_{AB})_{B\neq A}$ that describe the bonding between atoms A and B [14].

If the atomic weight matrices are chosen as projectors onto orthogonal subspaces, which implies that they are both idempotent and orthogonal

$$w_A w_B = w_A \delta_{AB}, \tag{4}$$

then the bond matrices $(\rho_{AB})_{B\neq A}$ have a zero trace and only the atomic matrices ρ_{AA} will contribute to the number of electrons N in the molecule

$$N \;=\; Tr(\rho) = \sum_A Tr(\rho_{AA}). \tag{5}$$

Note that in general the orthogonality of the weight matrices does not imply a zero overlap between the atomic electron densities $\rho_{AA}(\boldsymbol{r}, \boldsymbol{r'})$ in coordinate space, as occurs in e.g. QCT.

III. CONSTRUCTING ORTHOGONAL ATOMIC PROJECTORS

The decomposition of one-electron space according to Eq. (4) can be done in myriad ways, so a way must be found that generates a chemically relevant decomposition. We found that, starting from the 1DM's $\rho_{AA}^{(0)}$ for isolated atoms, the following recursive scheme $(i = 0, 1, ...)$

$$\begin{aligned} \rho^{(i)} \;&=\; \sum_A \rho_{AA}^{(i)}\,; \quad w_A^{(i)} = (\rho^{(i)})^{-\frac{1}{2}} \rho_{AA}^{(i)} (\rho^{(i)})^{-\frac{1}{2}} \\ \rho_{AA}^{(i+1)} \;&=\; w_A^{(i)} \rho w_A^{(i)} \end{aligned} \tag{6}$$

converges and generates weight matrices $w_A^{(\infty)}$ that obey Eq. (4). In a forthcoming publication we will analyze this result and introduce alternative recursive schemes, including proofs. Here we

just mention three points: (1) it is understood that the molecular 1DM ρ used in Eq. (6) is positive semidefinite by construction and therefore the same holds for the $\rho_{AA}^{(i)}$ and $\rho^{(i)}$; (2) when the 1DM of the isolated atom $\rho_{AA}^{(0)}$ contains only basis functions centered on atom A, the subspaces spanned by the eigenvectors of $\rho_{AA}^{(0)}$ are linearly independent, and as a result the weight matrices $w_A^{(0)}$ of the 0th iteration already obey Eq. (4); (3) depending on the level of theory for the atomic calculation it is possible that a nullspace is generated for $\rho^{(i)}$, making the inverse square root in Eq. (6) singular. We therefore replace any eigenvalues of the $\rho_{AA}^{(0)}$ and ρ by a small (10^{-5}) positive value. This is sufficient to ensure that all $\rho^{(i)}$ remain nonsingular.

IV. SELF-CONSISTENCY REQUIREMENTS

As in ordinary Hirshfeld, the choice of the starting point (the isolated atom 1DM's $\rho_{AA}^{(0)}$) is crucial, since different results are obtained depending on the charge and electronic state of the isolated atom. Following the ideas behind the iterative Hirshfeld procedure [6], the result can be made independent of the starting point by building in self-consistency through an outer iterative scheme (where Eq. (6) would represent the inner iterative scheme). Note that the need for a fitted start point

In our simplest implementation (called 'charge equalization') we start from rotationally averaged 1DM's of the neutral isolated atoms. The recursive scheme in Eq. (6) generates effective electron numbers $N_A = Tr(\rho_{AA}^{\infty})$ for the atoms. These can be used to create a rotationally averaged 1DM of the charged isolated atom according to the linear interpolation between integer electron numbers ($k \leq N_A < k+1$)

$$\begin{aligned} \rho_{AA}^{(0)}[N_A] \;=\; & (k+1-N_A)\,\rho_{AA}^{(0)}[k] \\ & +\; (N_A-k)\,\rho_{AA}^{(0)}[k+1]. \end{aligned} \tag{7}$$

The charged $\rho_{AA}^{(0)}[N_A]$ can be used as the next starting point in Eq. (6) and the whole process is repeated until convergence for the effective electron numbers.

We noticed that, in contrast to the electron density, the 1DM is much more sensitive to a mismatch in the orientation of one-electron orbitals in the molecule and in the isolated atom used to set up the AIM. In some cases, this resulted in rather large AIM charges. In order to solve this problem we also implemented a more sophisticated scheme (called 'population equalization'). The simple implementation of the previous paragraph is followed up to the point where $\rho_{AA}^{(0)}[N_A]$

is obtained. We then use the eigenvalue decompositions

$$\rho_{AA}^{(\infty)} = \sum_k n_{AA,k}^{(\infty)} \varphi_{AA,k}^{(\infty)} \varphi_{AA,k}^{(\infty)} \tag{8}$$

$$\rho_{AA}^{(0)}[N_A] = \sum_l n_{AA,l}^{(0)} \varphi_{AA,l}^{(0)} \varphi_{AA,l}^{(0)} \tag{9}$$

to generate the new starting point to Eq. (6). This is given by Eq. (9), but with the rotationally averaged occupation numbers $n_{AA,l}^0$ replaced with the occupation numbers $n_{AA,k}^{(\infty)}$ of the partitioned atom. The correspondence is made on the basis of maximal orbital overlap $|\langle\varphi_{AA,k}^{(\infty)}|\varphi_{AA,l}^{(0)}\rangle|$. By this procedure, the rotationally averaged 1DM's are replaced by 1DM's containing information on the molecular geometry. The whole process is again repeated until convergence.

V. RESULTS AND DISCUSSION

The procedure described in Sec. IV was tested by partitioning the 1DM of a small set of ca. 67 simple molecules with a singlet ground state, representative of a variety of chemical bonds. The set of molecules comprises the species tested in [14], supplemented with some extra molecules with relatively high Hirshfeld-I charges like CF_4 and H_2SO_4. The molecular and atomic 1DM's were calculated at the restricted Hartree-Fock level of theory using the Aug-cc-pVDZ basis set [16–18]. The molecular geometry was taken from a B3LYP [20–23] /cc-pVDZ [17, 18, 24] optimization. For CO additional calculations were performed in larger (Aug-cc-pVTZ and Aug-cc-pVQZ) [16–18] basis sets in order to assess basis set convergence.

A. The CO molecule

Diagonalization of the matrices in Eq. (3) leads to natural orbitals and occupancies. Fig. 1 presents the dominant natural orbitals, and the corresponding occupancies, of the atomic density matrix of carbon (a-e) and oxygen (f-j) in a CO molecule using the population equalization scheme.

Similar to the double-index partitioning in 3D space based on the use of Hirshfeld-I weights [14], the natural orbitals are slightly deformed versions of the typical atomic $1s$ (a), $2s$ (b), $2p_z$, $2p_x$ and $2p_y$ (c-d-e) orbitals. However, in contrast to [14], the current scheme generates occupancies of the orbitals not involved in bonding (core orbitals or orbitals containing free electron pairs) that are very close to their expected integer value. Also note that the double occupied molecular σ -, π_1

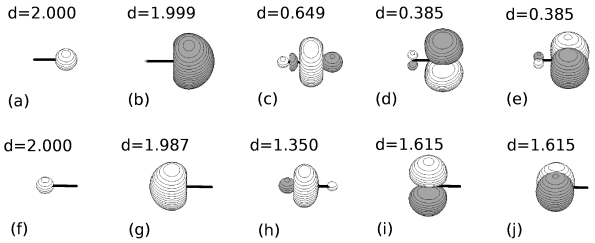


FIG. 1: The dominant natural orbitals, and the corresponding occupancies of the atomic density matrix of carbon (a-e) and oxygen (f-j) in a CO molecule. For details, see text.

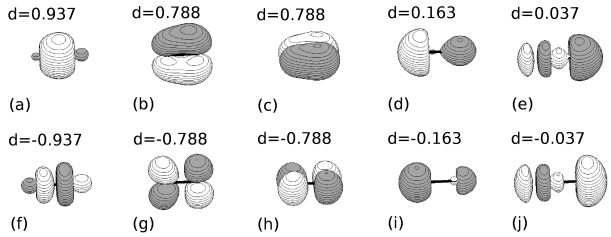


FIG. 2: The dominant eigenvectors and eigenvalues of (twice) the CO bond matrix in a CO molecule. For details, see text.

and π_2 Hartree-Fock orbitals simply divide their occupancy over the atomic p_z, p_x and p_y orbitals. The summed occupancy of (c) and (h) in Fig. 1 is exactly 2.

Fig. 2 depicts the dominant eigenvectors and their eigenvalues of the CO bond matrix ρ_{AB} in the CO molecule. Their shapes match those of the expected bonding and antibonding orbitals, but in contrast to [14], the corresponding bonding and antibonding orbitals have exactly opposite eigenvalues by construction. The $\sigma* \rightarrow \sigma$, $\pi_1* \rightarrow \pi_1$ and $\pi_2* \rightarrow \pi_2$ electron relocations are the main contributions to the bond matrix, though there are some smaller contributions (i $\rightarrow$ d , j $\rightarrow$ e) that have a mainly non-bonding character. Note that we avoid the term "occupancy" for the bond matrices, as the bond matrix reflects a change in density rather then a density itself.

Apart from the dominant contributions shown in Figs. 1-2, various orbitals with much smaller eigenvalues are also present (as the Aug-cc-pVDZ molecular basis set has 46 basis functions).

	C , C	O , O	C , O + O , C
(a)	2.000 (2.000)	2.000 (2.000)	$\pm$ 0.937 ($\pm$ 0.919)
(b)	1.999 (1.988)	1.987 (1.992)	$\pm$ 0.788 ($\pm$ 0.799)
(c)	0.649 (0.605)	1.615 (1.601)	$\pm$ 0.788 ($\pm$ 0.799)
(d)	0.385 (0.399)	1.615 (1.601)	$\pm$ 0.163 ($\pm$ 0.154)
(e)	0.385 (0.399)	1.350 (1.395)	$\pm$ 0.037 ($\pm$ 0.128)
(f)	0.013 (0.008)	0.001 (0.012)	$\pm$ 0.014 ($\pm$ 0.009)
(g)	0.000 (0.000)	0.000 (0.000)	$\pm$ 0.011 ($\pm$ 0.007)
(...)	$< 10^{-13}$	$< 10^{-13}$	$-10^{-13} >< 10^{-13}$
Sum	5.432 (5.400)	8.568 (8.600)	2.738 (2.816)

TABLE I: All natural populations in the atomic density matrices (CC and OO) and eigenvalues of the bond matrix (CO) in a CO molecule. Non-bracketed and bracketed values belong to the population equalization scheme and the charge equalization scheme respectively. For details, see text.

A complete overview is given in Table I for both population equalization and charge equalization schemes (bracketed values). Note that for the atomic density matrices, apart from the five dominant natural orbitals, only two more have small ($\sim 10^{-2} - 10^{-3}$) populations, while the remainder have vanishing ($< 10^{-13}$) populations. For the bond matrix, apart from the ten dominant orbitals, only four more have small eigenvalues, as was already noted in [14].

In Table II the stability of the proposed density matrix partitioning is examined. The results seem to be quite stable with respect to basis set size, with differences for the traces of atomic density matrices less than 0.01 going from DZ to TZ, and less than 0.001 going from TZ to QZ. For the positive and negative components of the bond matrices there are fluctuations of about 0.02 from DZ to TZ and from TZ to QZ. We checked that the stability holds even at the level of the individual orbital eigenvalues in Table I. In the charge equalization scheme, all eigenvalues differ less than 0.003 going from DZ to TZ, and less than 0.002 from TZ to QZ. In the population equalization scheme, the atomic matrices behave similarly (maximal deviation of 0.005 from DZ to TZ and of 0.003 from TZ to QZ), but the bond matrix eigenvalues deviate slightly more (maximal deviation of 0.015 from DZ to TZ and of 0.022 from TZ to QZ). We also checked that for almost all the orbitals in Figs. Figs. 1-2, the shapes are visually indistinguishable when changing basis set size from DZ to QZ. The exceptions are the bond matrix orbitals (e) and (j), corresponding to

CO	Aug-cc-pVDZ	Aug-cc-pVTZ	Aug-cc-pVQZ
A , B	Tr($\rho_{A,B}$)	Tr($\rho_{A,B}$)	Tr($\rho_{A,B}$)
C , C	5.432 (5.400)	5.425 (5.393)	5.426 (5.394)
O , O	8.568 (8.600)	8.575 (8.607)	8.574 (8.606)
C , O + O , C	$\pm$ 2.738 ($\pm$2.816)	$\pm$ 2.756 ($\pm$ 2.811)	$\pm$ 2.731 ($\pm$ 2.814)

TABLE II: Basis set convergence of the summed positive and negative eigenvalues in the atomic and bond matrices for CO. Non-bracketed and bracketed values belong to the population equalization scheme and the charge equalization scheme respectively. For details, see text.

small (in absolute value) eigenvalues, where a somewhat more distorted shape is observed.

B. Evaluation of the atomic charges

Fig. 3 displays the correlation between the charges obtained with the Hirshfeld-I method [6] (known to be a reliable AIM method) and the charges resulting from either the charge equalization scheme or the population equalization scheme applied to the entire set of 67 molecules. Both schemes show a strong linear correlation with Hirshfeld-I. It is remarkable that the linear correlation is more satisfying for the population equalization scheme (R^2=0.96, slope=1.18) than when only the atomic charges are made self-consistent (R^2=0.90, slope = 1.45). As mentioned, this is related to the inadequacy of the rotationally averaged 1DM's used as starting point in the charge equalization scheme. In some cases, the latter leads to unexpectedly high charges. This problem seems to be solved using the population equalization procedure, which is more adapted to the molecular geometry.

VI. CONCLUSIONS

We have implemented a two-index partitioning of the molecular density matrix into atomic and bond contributions using atomic weight matrices that are orthogonal projection operators in one-electron Hilbert space. The method is highly efficient in terms of computation time since no 3D integrals need to be computed. The resulting decomposition provides a rather physical description of the adaptations made by atoms in a molecule and the deformation caused by the bonding

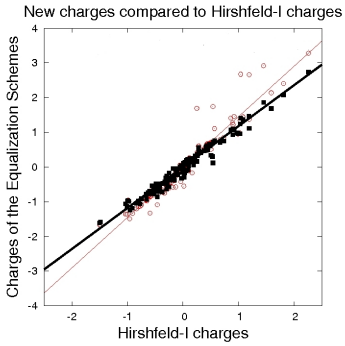


FIG. 3: Comparison of the Hirshfeld-I charges with those generated by the charge equalization scheme (red circles) and the population equalization scheme (black squares). For details, see text.

process. The bond matrices are traceless, i.e. the electrons are all in the atomic contributions and not in the bond matrices. Core orbitals and free electron pairs are fully assigned to the atoms. The trial atoms and the AIM atoms are required to have equal charges or equal orbital populations. The latter approach leads to a better correlation of the atomic charges with the Hirshfeld-I atomic charges.

[1] R. S. Mulliken, *J. Chem. Phys.* **23**, 1833 (1955).

[2] R. F. W. Bader, *Atoms in Molecules: A Quantum Theory.* The International Series of Mono-graphs on Chemistry Vol. 22 (Oxford University Press, Oxford, 1994).

[3] R. F. W. Bader, *Chem. Rev.* **91**, 893 (1991).

[4] P. Popelier, *Atoms In Molecules: An Introduction*. (Prentice-Hall, Harlow, Great Britain, 2000).

[5] F. L. Hirshfeld, *Theor. Chim. Acta* **44**, 129 (1977).

[6] P. Bultinck, C. Van Alsenoy, P. W. Ayers and R. Carbó-Dorca, *J. Chem. Phys.* **126**, 144111 (2007).

[7] P. Bultinck, *Faraday Discussions* **135**, 244 (2007).

[8] P. Bultinck, P. W. Ayers, S. Fias, K. Tiels, and C. Van Alsenoy, *Chem. Phys. Lett.* **444**, 205 (2007).

[9] P. Bultinck, D. L. Cooper, and D. Van Neck, *Phys. Chem. Chem. Phys.* **11**, 3424 (2009).

[10] T. C. Lillestolen and R. J. Wheatley, *Chem. Commun.* **45**, 5909 (2008).

[11] I. Mayer and P. Salvador, *Chem. Phys. Lett.* **383**, 368 (2004).

[12] L. Li and R. G. Parr, *J. Chem. Phys.* **84**, 1704 (1986).

[13] I. Mayer and P. Salvador, *J. Chem. Phys.* **130**, 234106 (2009).

[14] D. Vanfleteren, D. Van Neck, P. Bultinck, P. W. Ayers and M. Waroquier, *J. Chem. Phys.* **132**, 164111 (2010).

[15] I. Mayer and A. Hamza *Int. J. Quantum Chem.* **103**, 798 (2005).

[16] R. A. Kendall, T. H. Dunning and R. J. Harrison, *J. Chem. Phys.* **96**, 6796 (1992).

[17] D. E. Woon and T. H. Dunning, *J. Chem. Phys.* **98**, 1358 (1993).

[18] D. E. Woon, K. A. Peterson and T. H. Dunning, Gaussian basis sets for use in correlated molecular calculations. VII. Valence and core-valence basis sets for Li, Na, Be, and Mg, J.Chem. Phys. (in press).

[19] B. P. Prascher, D. E. Woon, K. A. Peterson, T. H. Dunning Jr and A. K. Wilson, *Theor. Chem. Acc.* DOI: 10.1007/s00214-010-0764-0

[20] A. D. Becke, *J. Chem. Phys.* **98**, 5648 (1993).

[21] C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B* **37**, 785 (1988).

[22] S. H. Vosko, L. Wilk, and M. Nusair, *Can. J. Phys.* **58**, 1200 (1980).

[23] P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.* **98**, 11623 (1994).

[24] T. H. Dunning, *J. Chem. Phys.* **90**, 1007 (1989).

Paper III

Fast density-matrix based partitioning of the energy over the atoms in a molecule consistent with the Hirshfeld-I partitioning of the electron density

D. Vanfleteren, D. Ghillemijn, D. Van Neck, P. Bultinck, M. Waroquier, and P. W. Ayers

Accepted by The Journal of Computational Chemistry, august (2011)

Fast density-matrix based partitioning of the energy over the atoms in a molecule consistent with the Hirshfeld-I partitioning of the electron density

Diederik Vanfleteren,[1,2,*] Dieter Ghillemijn,[1,3] Dimitri Van Neck,[1,2]
Patrick Bultinck,[1,3] Michel Waroquier,[1,2] and Paul W. Ayers[4]

[1]*Members of the Ghent Brussels Quantum Chemistry and Molecular Modeling alliance.*
[2]*Ghent University, Center for Molecular Modeling,*
Technologiepark 903, B-9052 Zwijnaarde, Belgium
[3]*Ghent University, Department of Inorganic and Physical Chemistry,*
Krijgslaan 281 (S3), B-9000 Gent, Belgium
[4]*McMaster University, Department of Chemistry,*
Hamilton, Ontario L8S 4M1, Canada

Abstract

For the Hirshfeld-I atom-in-molecule model, associated single-atom energies and interaction energies at the Hartree-Fock level are determined efficiently in one-electron Hilbert space. In contrast to most other approaches, the energy terms are fully consistent with the partitioning of the underlying one-electron density matrix. Starting from the Hirshfeld-I atom-in-molecule model for the electron density, the molecular one-electron density matrix is partitioned with a previously introduced double-atom scheme [Vanfleteren D. et al., J Chem Phys 2010, 132, 164111]. Single-atom density matrices are constructed from the atomic and bond contributions of the double-atom scheme. Since the Hartree-Fock energy can be expressed solely in terms of the one-electron density matrix, the partitioning of the latter over the atoms in the molecule leads naturally to a corresponding partitioning of the Hartree-Fock energy. When the size of the molecule or the molecular basis set does not grow too large, the method shows considerable computational advantages compared to other approaches that require cumbersome numerical integration of the molecular energy integrals weighted by atomic weight functions.

*Electronic address: diederik.vanfleteren@ugent.be

I. INTRODUCTION

In the last decades, several algorithms have been developed to identify the atom in the molecule (AIM). For example the Mulliken method [1] relies on the attachment of basis functions to atomic centers; Natural Population Analysis [2] relies on the analysis of blocks of the one-electron density matrix (1DM) expressed in some molecular orbital basis; some other methods rely on the partitioning of the molecular electron density in AIM parts. However, not all AIM properties can be directly expressed in terms of the electron density. Of course the density determines all the properties and there are even explicit formulas for them, but these formulas are often computationally impractical. A common example is the kinetic energy of an AIM, which is directly computable from the full 1DM $\rho(\mathbf{r}, \mathbf{r}')$, but not (trivially) from the electron density, the diagonal element of the 1DM.

A widely adopted solution, which is also the most common recipe employed within the Quantum Theory of Atoms in Molecules (QTAIM) [3], is to partition such molecular properties using the same atomic weight functions $w_A(\mathbf{r})$ that are used to partition the molecular density. QTAIM uses a zero flux condition of the electron density to define interatomic surfaces, resulting in nonoverlapping atomic regions. Atomic energies are obtained by partitioning the kinetic energy over the atomic domains and using the virial ratio to construct the corresponding atomic potential energy. Other approaches also rely on atomic regions to partition the molecular energy, but use them in a more general way to decompose the energy into one- and two-atom terms [4–7]. Recently, Mandado et al. [8] partitioned the Hartree-Fock energy in terms of the overlapping Hirshfeld atoms. Their scheme appears useful to investigate proton acidity, the anomeric effect and group transferability, it has also been used in studies of bonding and polarizability [9, 10]. However, partitioning of the molecular energy in atomic fragments can be ambiguous as it often relies on the introduction of an arbitrary number of partitionings of unity [11] into the energy expressions. Moreover, the exact place where the partition of unity is introduced in an expectation value expression can have an important influence on the resulting AIM condensed values [12]. This ambiguity is often circumvented by the convention to introduce the unity and its partitioning into weight functions ($1 = \sum_A w_A(\boldsymbol{r})$) before any operator [11].

To avoid these problems, we partition the Hartree-Fock energy starting from a partitioning of Hartree-Fock molecular 1DM. At the Hartree-Fock level of theory the energy can be expressed solely and directly in terms of a (partitioned) molecular 1DM. A partitioning of the Hartree-

Fock molecular 1DM therefore directly leads to a natural partitioning of the energy. In a previous paper [13] we introduced a double-atom partitioning scheme for the molecular 1DM that is consistent with existing partitioning schemes for the molecular density. In the current work we use molecular 1DM fragments from that scheme to calculate the energy terms naturally associated with these fragments. The strategy of this work is therefore to calculate "properties of the molecular fragments", $\int d\mathbf{r}\ \hat{\mathrm{A}}\rho_A(\mathbf{r},\mathbf{r}')\Big|_{\mathbf{r}=\mathbf{r}'}$, instead of "fragmenting molecular properties", $\int d\mathbf{r}\ w_A(\mathbf{r})\,\hat{\mathrm{A}}\rho(\mathbf{r},\mathbf{r}')\Big|_{\mathbf{r}=\mathbf{r}'}$. This is in line with the work of Bultinck et al.[12] for the derivative of the density function with respect to a change in the number of electrons. We also examine the correspondences and differences between both approaches.

From a practical point of view, a fragmentation of molecular properties requires that the expectation value integrals are computed numerically on a large spatial grid. In general there is no simple analytical expression for the atomic weight functions and these weight functions are often not well expressed in one-electron Hilbert space. In contrast, molecular 1DM fragments show a satisfying basis set convergence in one-electron Hilbert space [13]. Therefore, it is tempting to calculate the atomic and interaction energies of the 1DM fragments in one-electron Hilbert space, avoiding cumbersome numerical integrations in $\boldsymbol{r}$-space. As a consequence, we expect to find that significant computational advantages are a major asset of our current approach.

II. ATOMIC DENSITY MATRICES THAT ARE CONSISTENT WITH THE HIRSHFELD-I PARTITIONING OF THE ELECTRON DENSITY

We use the notation $\boldsymbol{x} = \boldsymbol{r}\sigma$ to specify the single-electron states in coordinate space, where σ represents the spin degrees of freedom. The one-electron density matrix (1DM) for an N-electron molecule with wave function $\Psi(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N)$ is defined as

$$\rho(\boldsymbol{x},\boldsymbol{x}') = N\int d\boldsymbol{x}_2\ldots\int d\boldsymbol{x}_N\Psi^\dagger(\boldsymbol{x},\boldsymbol{x}_2,\ldots,\boldsymbol{x}_N)\Psi(\boldsymbol{x}',\boldsymbol{x}_2,\ldots,\boldsymbol{x}_N). \tag{1}$$

We restrict ourselves to molecules with a singlet ground state. In that case

$$\rho(\boldsymbol{x},\boldsymbol{x}') = \frac{1}{2}\delta_{\sigma,\sigma'}\rho(\boldsymbol{r},\boldsymbol{r}'), \tag{2}$$

and the electron spin can be discarded. In a previous paper we introduced a double-atom partitioning scheme for the molecular spin-summed 1DM [13],

$$\rho(\boldsymbol{r},\boldsymbol{r}') = \sum_{AB}\rho_{AB}(\boldsymbol{r},\boldsymbol{r}'), \tag{3}$$

in terms of atomic ($A = B$) and diatomic contributions ($A \neq B$), where A and B label atoms. The individual contributions are defined as

$$\rho_{AB}(\boldsymbol{r}, \boldsymbol{r'}) = \frac{1}{2}\left[w_A(\boldsymbol{r})w_B(\boldsymbol{r'}) + w_B(\boldsymbol{r})w_A(\boldsymbol{r'})\right]\rho(\boldsymbol{r}, \boldsymbol{r'}), \tag{4}$$

with positive local weight functions $w_A(\boldsymbol{r})$ obeying

$$\sum_A w_A(\boldsymbol{r}) = 1. \tag{5}$$

At the Hartree-Fock level the energy is expressed solely in terms of the 1DM. Therefore, the partitioning of the latter over the atoms in the molecule leads naturally to a corresponding partitioning of the Hartree-Fock energy. However, for a simple energy partitioning we require that (1) it is based on single-atom density matrices (rather than double-atom density matrices) and (2) the electron density of the atoms is local and positive definite. Note that any single-atom density matrix $\rho_A(\boldsymbol{r}, \boldsymbol{r'})$ unavoidably has bad localization properties (in contrast to the double-atom density matrices) [13], and therefore we restrict the requirement of positive definiteness and locality to the electron density, the diagonal element of the 1DM. These requirements are needed to prevent some energy components to have the wrong sign, or to become unrealistically large on the chemical energy scale. Note that in another type of approach, domain-averaged Fermi-hole analysis [14], the positivity requirement of the density attached to an atom is also automatically fulfilled at the Hartree-Fock level.

When the single-atom density matrices are defined as

$$\rho_A(\boldsymbol{r}, \boldsymbol{r}') \;=\; \sum_B \rho_{AB}(\boldsymbol{r}, \boldsymbol{r}'), \tag{6}$$

it is clear that on the diagonal ($\boldsymbol{r} = \boldsymbol{r}'$) the following definition is obtained,

$$\rho_A(\boldsymbol{r}) = \sum_B \rho_{AB}(\boldsymbol{r}, \boldsymbol{r}) = w_A(\boldsymbol{r})\rho(\boldsymbol{r}) \tag{7}$$

that is familiar from existing partitioning schemes for the electron density (with good localization properties). Also the fundamental property

$$\sum_A \rho_A(\boldsymbol{r}, \boldsymbol{r}') = \rho(\boldsymbol{r}, \boldsymbol{r}') \tag{8}$$

is obeyed as a trivial consequence of Eq. (3). Note that these single atom density matrices may not be N-representable, an issue that, however, does not stand in the way of obtaining integrated quantities such as energy contributions.

In order to define a computationally efficient energy partitioning scheme, we assume real wavefunctions and express the partitioned density matrices in the finite basis set used for the molecular calculation. E.g., in the molecular Hartree-Fock basis set, the following spatial integrals

$$S_{ij}^{A} = \int d\boldsymbol{r}\phi_i(\boldsymbol{r})w_A(\boldsymbol{r})\phi_j(\boldsymbol{r}), \tag{9}$$

that represent elements of the regular atomic overlap matrix (AOM), are sufficient to determine the single-atom 1DM in this basis,

$$(\rho_A)_{ij} = \int d\boldsymbol{r}d\boldsymbol{r'}\phi_i(\boldsymbol{r})\rho_A(\boldsymbol{r},\boldsymbol{r'})\phi_j(\boldsymbol{r'}). \tag{10}$$

The precise expressions for the density matrices are:

$$(\rho_A)_{ij} = \sum_n d_n \frac{1}{2}\sum_B (S_{in}^A S_{jn}^B + S_{in}^B S_{jn}^A) = \frac{1}{2}(d_i + d_j)S_{ij}^A, \tag{11}$$

where the quantity d_n represents the occupation numbers of the natural orbitals resulting from the diagonalization of the molecular density matrix. A simple sum rule

$$\sum_B S_{jn}^B = \delta_{jn} \tag{12}$$

is used to simplify the expression in Eq. (11).

In order to ensure that the single-atom density $\rho_A(\boldsymbol{r})$ is local, it can be made to coincide with the AIM density of some well-established density partitioning scheme like Hirshfeld [15], iterative Hirshfeld [16], Iterated Stockholder Atoms [17, 18] or QTAIM [3, 19, 20],

$$\rho_A^{AIM}(\boldsymbol{r}) = h_A(\boldsymbol{r})\rho(\boldsymbol{r}) \tag{13}$$

in which the AIM density is obtained by multiplying the molecular density $\rho(\boldsymbol{r})$ with the characteristic weight function for that AIM technique $h_A(\boldsymbol{r})$. One simply takes

$$w_A(\boldsymbol{r}) \equiv h_A(\boldsymbol{r}). \tag{14}$$

An alternative to Eq. (7), consisting of distributing the diatomic contributions in a weighted manner, was also investigated in [13], but will be discarded here because it gave inferior results.

III. ENERGY DECOMPOSITION

A. Self-energies and interaction energies of the 1DM fragments

Following the ideas behind the Interacting Quantum Atoms (IQA) [21–23], but from the perspective of density matrix fragments in one-electron Hilbert space, each single-atom 1DM with

elements $(\rho_A)_{ij}$ constructed in section II can be thought of as forming an atomic subsystem [24] when combined with the nucleus on center A. The molecular energy can then be decomposed as the sum of the "self-energies" of the atomic subsystems and the interaction energies between them:

$$E^{mol} = \sum_A E_A^{self} + \sum_{A<B} E_{AB}^{int}. \tag{15}$$

At the closed-shell Hartree-Fock level, the self-energy of such an atomic subsystem corresponds to:

$$E_A^{self} = \sum_{ij}(t + V_A^{ext})_{ij}(\rho_A)_{ij} + \frac{1}{2}\sum_{ijkl} V_{ijkl}\left((\rho_A)_{ik}(\rho_A)_{jl} - \frac{1}{2}(\rho_A)_{il}(\rho_A)_{jk}\right) \tag{16}$$

where

$$\begin{aligned} t_{ij} &= \int \phi_i(\boldsymbol{r})\hat{t}\phi_j(\boldsymbol{r})d\boldsymbol{r} \\ (V_A^{ext})_{ij} &= \int \phi_i(\boldsymbol{r})\frac{Z_A}{|\boldsymbol{r}-\boldsymbol{R}_A|}\phi_j(\boldsymbol{r})d\boldsymbol{r} \\ V_{ijkl} &= \int \phi_i(\boldsymbol{r}_1)\phi_j(\boldsymbol{r}_2)\frac{1}{|\boldsymbol{r}_1-\boldsymbol{r}_2|}\phi_k(\boldsymbol{r}_1)\phi_l(\boldsymbol{r}_2)d\boldsymbol{r_1}d\boldsymbol{r_2} \end{aligned} \tag{17}$$

In Eq. (17), $\hat{t}$ is the kinetic energy operator and Z_A is the nuclear charge on an atom A with nuclear coordinate $\boldsymbol{R}_A$. V_{ijkl} are two-electron integrals that are not anti-symmetrized. Note that it is implied in Eq. (16) that the atomic subsystems are spin-averaged, consistent with Eq.(2-4). The atomic subsystems interact with each other, according to the following expressions:

$$\begin{aligned} E_{AB}^{int} &= \frac{Z_A Z_B}{r_{AB}} + \sum_{ij}(V_A^{ext})_{ij}(\rho_B)_{ij} + \sum_{ij}(V_B^{ext})_{ij}(\rho_A)_{ij} \\ &\quad + \sum_{ijkl} V_{ijkl}\left((\rho_A)_{ik}(\rho_B)_{jl} - \frac{1}{2}(\rho_A)_{il}(\rho_B)_{jk}\right) \end{aligned} \tag{18}$$

$$E^{int} = \sum_{A<B} E_{AB}^{int}. \tag{19}$$

E^{int} represents the total interaction energy of all subsystems within the molecule. E_{AB}^{int} quantifies all interactions between atom pairs in molecules, including the interaction between atom pairs that do not share a chemical bond. This quantity is also useful to assess the strength of the interactions in a ring structure. E_{AB}^{int} does not depend on the choice of reference atoms, it only depends on the AIM. The localized character of the Hirshfeld-I densities ensures that the promotion and interaction energies are within a reasonable range of values, although they are often significantly larger (in absolute value) than typical "bond energies".

B. Promotion energies

Since it is well known that the molecular environment induces only relatively small changes in atomic energy [24], atomic energies are usually referenced to the energy of the isolated atoms. Atoms within a molecular environment are slightly distorted compared to the isolated atoms. This implies that their energy, with respect to the Hamiltonian of the isolated atoms, has increased. This energy increase is called the atomic promotion energy E_A^{prom}. When the atom in a molecule is identified with the atomic subsystem defined above, its promotion energy is the difference between the (Hartree-Fock) atomic self-energy E_A^{self} in Eq. (16) and the Hartree-Fock energy E_A^0 of the isolated atom:

$$E_A^{prom} = E_A^{self} - E_A^0 \tag{20}$$

$$E^{prom} = \sum_A E_A^{prom}. \tag{21}$$

The atomic promotion energies can be considered to result from three successive processes: first a charge transfer step that accounts for the fact that the AIM has a different atomic charge from the isolated atom.

$$\Delta E_A^{CT} = E_A^0(Q_A) - E_A^0(0) \tag{22}$$

Classically, the reference for this step is the neutral isolated atom in its ground state $E_A^0(0)$. $E_A^0(Q_A)$ represents the charged isolated atom, and is approximated as the linearly interpolated value between the energies of the isolated atoms with an integer number of electrons $N < (Q_A = N + a) < N + 1$. In principle, the HF energy as a function of the number of electrons N is a concave curve between the integers [8, 25–28]. The assumption that it is linear is based on the fact that this holds for the exact energies [29–31]. This leads to $E_A^0(Q_A)$ computed as:

$$E_A^0(Q_A) = aE_A^0(N+1) + (1-a)E_A^0(N). \tag{23}$$

The interpolated state $E_A^0(Q_A)$ is characterized by a value of $\langle S^2 \rangle$ (where S is the spin angular momentum) that does not always correspond to a singlet whereas the AIM is always assumed to be spin averaged (see Eq. (2)). We therefore introduce a second step that accounts for the spin averaging. It corresponds to the energy difference between on the one hand the isolated atom with interpolated charge and averaged spin $E_A^{0,\overline{S}}(Q_A)$ and on the other hand $E_A^0(Q_A)$,

$$\Delta E_A^{\overline{S}} = E_A^{0,\overline{S}}(Q_A) - E_A^0(Q_A). \tag{24}$$

$E_A^{0,\overline{S}}(Q_A)$ is obtained as the interpolated value between the energies of the spin-averaged isolated atoms with an integer number of electrons,

$$E_A^{0,\overline{S}}(Q_A) = aE_A^{0,\overline{S}}(N+1) + (1-a)E_A^{0,\overline{S}}(N), \tag{25}$$

and the $E_A^{0,\overline{S}}(N)$ are calculated from the following formula:

$$E_A^{0,\overline{S}}(N) \;=\; \sum_A \sum_{ij} (t + V_A^{ext})_{ij} (\rho_A^0)_{ij} + \frac{1}{2} \sum_{ijkl} V_{ijkl} \left((\rho_A^0)_{ik} (\rho_A^0)_{jl} - \frac{1}{2} (\rho_A^0)_{il} (\rho_A^0)_{jk} \right), \tag{26}$$

where ρ_A is the spin-summed 1DM of the isolated atom with N electrons.

In the third step, charge reorganization takes place which corresponds to the difference between the self- energy of the of the atomic 1DM fragment and the energy of the isolated atom with interpolated charge and averaged spin,

$$\Delta E_A^{CR} = E_A^{self} - E_A^{0,\overline{S}}(Q_A). \tag{27}$$

Finally, the promotion energy is retrieved as the sum of three terms as shown below:

$$E_A^{prom} = \Delta E_A^{CT} + \Delta E_A^{\overline{S}} + \Delta E_A^{CR}. \tag{28}$$

Although one may point out that there is some arbitrariness in the order chosen for the processes, the decomposition in Eq. (28) might give a general idea about their relative importance.

C. Bond energies

Since the promotion and interaction energies depend on the AIM definition, they are not directly observable. The sum of all promotion and interaction energies

$$\sum_A E_A^{prom} + \sum_{A<B} E_{AB}^{int} = \Delta E^{at}, \tag{29}$$

should be equal the atomization energy ΔE^{at} of the molecule,

$$\Delta E^{at} = E^{mol} - \sum_A E_A^0, \tag{30}$$

and can be compared to experiment. In order to get a quantity that is more in line with the chemical concept of a "bond energy", it would be convenient to interpret the atomization of a molecule

strictly in terms of atom pairs. Therefore, the atomic promotion energies (E_A^{prom}) are attributed to the atom pairs ($A < B$),

$$E_{AB}^{prom} = E_A^{prom} \left(\frac{E_{AB}^{int}}{\sum_C E_{AC}^{int}} \right) + E_B^{prom} \left(\frac{E_{AB}^{int}}{\sum_C E_{BC}^{int}} \right), \tag{31}$$

and Hartree-Fock bond energies are derived,

$$E_{AB}^{bond} = E_{AB}^{int} + E_{AB}^{prom}, \tag{32}$$

that reproduce exactly the Hartree-Fock atomization energy

$$\Delta E_{(HF)}^{at} = \sum_{A<B} E_{AB}^{bond}. \tag{33}$$

These bond energies can then be compared to average bond dissociation energies.

In Eq. (31) the atomic promotion energy of atom A is partitioned over the atom pairs AB in proportion to the strength of the interaction between the AB fragments. We noticed that for diatomic molecules there is only a moderate correlation between the promotion energy ($E_A^{prom} + E_B^{prom}$) and the bond order index (R^2=0.84) but a strong correlation between the promotion energy and the interaction energy (R^2=0.98). So the heuristic foundation for Eq. (31) is the observation that strongly interacting 1DM fragments have in general high promotion energies as well. Although Eq. (31) is an empirical formula, it works remarkably well. Since the bond energies E_{AB}^{bond} in Eq. (32) are much smaller (in absolute value) than the contributions E_{AB}^{int} and E_{AB}^{prom} from which it is constructed, one might expect that this quantity is very sensitive and could easily have the wrong sign. However, using the prescription in Eq. (31) and for the data in the paper we did not observe any value that is unrealistic (compared to the atomization energy at the Hartree-Fock level).

IV. CONSISTENT DECOMPOSITION OF MOLECULAR PROPERTIES

In the introduction we already noted that the current approach to calculate self-energies and interaction energies for 1DM fragments is unambiguous. That is not the case for the widely adopted approach to partition the molecular energy using numerical integration of the molecular integrals weighted by the atomic weight functions. In this section we explore the exemplary cases of two components of the atomic self-energy (see Eq. (16)): the atom-condensed kinetic energy and the atom-condensed Fock energy.

A. Atom-condensed kinetic energy

The strategy adopted in the current paper is to calculate energy components of the molecular fragments, rather than to fragment the molecular energy. For the kinetic energy in particular, this can have important consequences. The kinetic energy t_A of the single-index atomic density matrix is calculated in Eqs. (16- 17) in a finite basis set, but the corresponding expression in $\boldsymbol{r}$-space is:

$$t_A = -\frac{1}{2}\int d\boldsymbol{r}\ \nabla^2 \rho_A(\boldsymbol{r},\boldsymbol{r'})\big|_{\boldsymbol{r}=\boldsymbol{r'}}\,, \tag{34}$$

where the notation $|_{\boldsymbol{r}=\boldsymbol{r'}}$ indicates that $\boldsymbol{r'}$ is replaced by $\boldsymbol{r}$ after the action of $\nabla^2_{(\boldsymbol{r})}$ on $\rho_A(\boldsymbol{r},\boldsymbol{r'})$ but before the integration is carried out. Note that different but equivalent representations of the kinetic energy operator exist, *i.e.*

$$t_A = \frac{1}{2}\int d\boldsymbol{r}\ \nabla\cdot\nabla' \rho_A(\boldsymbol{r},\boldsymbol{r'})|_{\boldsymbol{r}=\boldsymbol{r'}}\,, \tag{35}$$

should give the same result as in Eq. (34). This is indeed fulfilled for the present formulation in terms of atomic density matrices, as follows directly from partial integration by generalizing a well known relationship for the kinetic energy of the molecular 1DM,

$$\frac{1}{2}\int d\boldsymbol{r}\ \nabla\cdot\nabla' \rho(\boldsymbol{r},\boldsymbol{r'})|_{\boldsymbol{r}=\boldsymbol{r'}} = -\frac{1}{2}\int d\boldsymbol{r}\ \nabla^2 \rho(\boldsymbol{r},\boldsymbol{r'})\big|_{\boldsymbol{r}=\boldsymbol{r'}}\,, \tag{36}$$

to its fragments. In contrast, the expression for the fragmentation of the molecular kinetic energy with Hirshfeld-I weight functions clearly depends on the representation of the kinetic energy operator [32, 33], since in general

$$t_A^h = -\frac{1}{2}\int d\boldsymbol{r}\ w_A(\boldsymbol{r})\nabla^2 \rho(\boldsymbol{r},\boldsymbol{r'})\big|_{\boldsymbol{r}=\boldsymbol{r'}} \tag{37}$$

yields a result that differs from

$$t_A^{h'} = \frac{1}{2}\int d\boldsymbol{r}\ w_A(\boldsymbol{r})\nabla\cdot\nabla' \rho(\boldsymbol{r},\boldsymbol{r'})|_{\boldsymbol{r}=\boldsymbol{r'}}\,. \tag{38}$$

Only in special cases, e.g. when QTAIM weight functions are used, do Eq. (37) and Eq. (38) coincide. This shows that a naive fragmentation of the molecular energy by introducing weight functions in the molecular integrals is ambiguous by nature, in contrast to the fragmentation of the molecular 1DM and the calculation of the associated energy components.

It is shown in the appendix that the $\boldsymbol{r}$-space expressions for t_A (Eqs. (34) and (35)) and t_A^h (Eq. (37)) are mathematically identical. Since Eq. (37) does not contain the action of the kinetic energy operator on an atomic weight function it can be calculated quite easily using 3D numerical integration on a spatial grid. This provides an indication of the error induced by using the finite basis set expression in Eq. (16) instead of the full $\boldsymbol{r}$-space expression in Eq. (34).

B. Atom-condensed Fock energy

The Fock operator provides another example of a nonlocal operator that causes problems in the common approach, where atomic weight functions are inserted in the molecular Fock integrals. Indeed, there are several possibilities for the fragmentation of the Fock energy,

$$E^{Fock} = -\frac{1}{2}\int d\boldsymbol{r}d\boldsymbol{r'}\ \frac{[\rho(\boldsymbol{r},\boldsymbol{r'})]^2}{|\boldsymbol{r}-\boldsymbol{r'}|}. \tag{39}$$

Depending on the place where the decomposition of unity is inserted, one could have

$$\begin{aligned} E^{Fock} &= \sum_{AB} -\frac{1}{2}\int d\boldsymbol{r}d\boldsymbol{r'}\ w_A(\boldsymbol{r})w_B(\boldsymbol{r'})\ \frac{[\rho(\boldsymbol{r},\boldsymbol{r'})]^2}{|\boldsymbol{r}-\boldsymbol{r'}|} \\ &= \sum_{AB} -\frac{1}{4}\int d\boldsymbol{r}d\boldsymbol{r'}\ [w_A(\boldsymbol{r})w_B(\boldsymbol{r}) + w_A(\boldsymbol{r'})w_B(\boldsymbol{r'})]\frac{[\rho(\boldsymbol{r},\boldsymbol{r'})]^2}{|\boldsymbol{r}-\boldsymbol{r'}|} \\ &\ \ \vdots \end{aligned} \tag{40}$$

Only one fragmentation is consistent with the underlying partitioning in Eq. (8) of the molecular 1DM into single-index atomic density matrices,

$$\begin{aligned} E^{Fock} &= \sum_{AB} -\frac{1}{2}\int \frac{\rho_A(\boldsymbol{r},\boldsymbol{r'})\rho_B(\boldsymbol{r},\boldsymbol{r'})}{|\boldsymbol{r}-\boldsymbol{r'}|} \\ &= \sum_{AB} -\frac{1}{8}\int d\boldsymbol{r}d\boldsymbol{r'}\ [w_A(\boldsymbol{r})w_B(\boldsymbol{r}) + w_A(\boldsymbol{r'})w_B(\boldsymbol{r'}) \\ &\quad + w_A(\boldsymbol{r})w_B(\boldsymbol{r'}) + w_B(\boldsymbol{r})w_A(\boldsymbol{r'})]\ \frac{[\rho(\boldsymbol{r},\boldsymbol{r'})]^2}{|\boldsymbol{r}-\boldsymbol{r'}|}. \end{aligned} \tag{41}$$

Note that the Fock contributions with indices (A = B) are attributed to E_A^{self}, while the Fock contributions with indices (A $\neq$ B) are attributed to E_{AB}^{int}.

V. A FAST DECOMPOSITION IN HILBERT SPACE

In a previous paper [13], it was observed that the molecular 1DM fragments show a satisfactory basis set convergence when they are expressed in one-electron Hilbert space. To avoid cumbersome numerical integrations in $\boldsymbol{r}$-space, atomic self-energies and interaction energies are calculated efficiently in one-electron Hilbert space (see Eqs. (16) and (18)). The widespread alternative strategy to fragment the molecular energy (by inserting atomic weight functions in the molecular expressions) requires that the atom-condensed energy integrals are computed numerically on a large spatial grid. In this section we explore the computational consequences for two

components of the atomic self-energy (see Eq. (16)): the atom-condensed kinetic energy and the atom-condensed Fock energy.

A. Atom-condensed Kinetic energy

When the kinetic energy is calculated by numerical integration of the molecular energy integrals weighted by the atomic weight functions (see Eq. 37),

$$t_A^h \;=\; -\frac{1}{2}\sum_i d_i \int d\boldsymbol{r} w_A(\boldsymbol{r}) \left[\phi_i(\boldsymbol{r}) \left(\nabla^2 \phi_i(\boldsymbol{r}) \right) \right], \tag{42}$$

it is clear that only few integrals have to be computed since nonzero contributions are obtained only for occupied MO's. This is an important computational advantage over using an atomic density matrix as in Eq. (16-17), to which all orbitals in the basis contribute. However, experience shows that the numerical computation of the integrals above requires quite large integration grids whereas the atomic overlap integrals can be computed with sufficient precision using more modest grids. As will be shown, there is a trade-off between either computing numerically fewer integrals that require larger grids and more atomic overlap integrals that are individually computed much easier.

The strategy adopted in the current paper is to calculate energy components via matrix manipulations in one-electron Hilbert space (See Eqs. (16) and (18)). Using Eq. (11), the precise expression for the kinetic energy is,

$$t_A \;=\; \frac{1}{2}\sum_{ij}(d_i + d_j) t_{ij} S_{ij}^A. \tag{43}$$

A non-zero contribution is obtained if either d_i or d_j is nonzero. This implies that the summation essentially runs over all i and j and that all of the following integrals must be computed:

$$\begin{aligned} S_{ij}^A \;&=\; \int d\boldsymbol{r} \phi_i(\boldsymbol{r}) w_A(\boldsymbol{r}) \phi_j(\boldsymbol{r}) \\ t_{ij} \;&=\; -\frac{1}{2}\int \phi_i(\boldsymbol{r}) \left(\nabla^2 \phi_j(\boldsymbol{r}) \right) d\boldsymbol{r} \end{aligned}$$

However, the t_{ij} can be calculated analytically and moderate grids suffice to construct the S_{ij}^A.

B. Atom-condensed Fock energy

For the partitioning of e.g. the Fock-energy, the considerations of the previous subsection are slightly more outspoken than for the kinetic energy. Using the strategy of numerical integration,

$$E_{AB}^{Fock} = -\frac{1}{8}\sum_{ij} d_i d_j \int d\boldsymbol{r} d\boldsymbol{r'} \; [w_A(\boldsymbol{r}) w_B(\boldsymbol{r}) + w_A(\boldsymbol{r'}) w_B(\boldsymbol{r'}) + w_A(\boldsymbol{r}) w_B(\boldsymbol{r'}) + w_B(\boldsymbol{r}) w_A(\boldsymbol{r'})] \; \frac{\phi_i(\boldsymbol{r})\phi_i(\boldsymbol{r'})\phi_j(\boldsymbol{r})\phi_j(\boldsymbol{r'})}{|\boldsymbol{r} - \boldsymbol{r'}|}, \tag{44}$$

the number of integrals that has to be computed is limited to the square of the number of occupied orbitals. On the other hand, the cost of numerical integration is squared with respect to the situation for the kinetic energy since the integrals run over $\boldsymbol{r}$ and $\boldsymbol{r'}$.

When the Fock energy is partitioned via matrix manipulations in one-electron Hilbert space, the precise expression is:

$$E_{AB}^{Fock} = -\frac{1}{16}\sum_{ijkl}(d_i + d_l)(d_j + d_k) V_{ijkl} S_{il}^A S_{jk}^B, \tag{45}$$

where V_{ijkl} is calculated analytically. The summation runs over all ijkl, but the S_{il}^A are calculated on a moderate grid that runs over only $\boldsymbol{r}$. Except for large systems, the approach presented in this paper is more appealing from a computational point of view. Note that E_{AB}^{Fock} is part of E_A^{self} when $(A = B)$, while it is part of E_{AB}^{int} when $(A \neq B)$.

VI. COMPUTATIONAL METHODS

The formalism described in the sections II-III was applied to the set of 25 molecules listed in table I, representing a variety of chemical bonds. All molecules have a singlet ground state, apart from O_2, for which we consider the singlet (excited) state.

The molecular 1DM was calculated at the restricted open-shell Hartree-Fock (ROHF) [34–36] level of theory using the (cartesian) Aug-cc-pVTZ basis, including a geometry optimization. All molecular one-electron density matrices were calculated with the Gaussian 03 program [40].

The partitioning scheme was implemented using the atomic weight functions $h_A(\boldsymbol{r})$ in Eq.(13) from a Hirshfeld-I analysis. To be consistent with our previous work on the Hirshfeld-I compatible 1DM partitioning [13], the iterative Hirshfeld weights $h_A(\boldsymbol{r})$ and the S_{ij}^A coefficients derived from these weights were calculated on atom-centered grids, using a logarithmic radial grid of 100 points

H_2	N_2	HCl	CH_3OH	H_2O
F_2	LiH	CH_4	H_2CO	$(H_2O)_2$
Cl_2	LiF	CH_3CH_3	HCOOH	H_2O_2
Li_2	NaCl	CH_2CH_2	CO_2	NH_3
O_2	HF	CHCH	CO	N_2H_4

TABLE I: List of molecules in the test set.

with $r_{\text{min}} = 10^{-6}$ Å and $r_{\text{max}} = 20$ Å , and the 170-point Lebedev angular grids, [41–46] with each radial shell given a randomized orientation. The sum rule of Eq. (12) was fulfilled with a reasonable precision of 10^{-3} except for some very diffuse virtual orbitals (as we work in an augmented basis set), where the sum rule is fulfilled with less precision (10^{-2}) because the grid is too localized. Note that diffuse virtual orbitals do not pose a problem for the conventional strategy of numerical integration of the molecular energy integrals weighted by atomic weight functions, since a simple product of the molecular occupancies appears in those atom-condensed expressions (see Eqs. (42) and (44)). For the current approach, the diffuse virtual orbitals are of little relevance in the calculation of the energies and do not significantly affect the outcome of our analysis. In order to have an exact reproduction of the energy sum rule in Eq. (29), we have renormalized the S_{ij}^A, according to

$$S^A \rightarrow (\sum_B S^B)^{-\frac{1}{2}} S^A (\sum_B S^B)^{-\frac{1}{2}}. \tag{46}$$

If we do not perform this renormalization step then the deviation for the energy sum rule in Eq. (29) is typically on the order of 10^{-4} Hartree, though in a few cases (for CO_2 and the ionic compounds) deviations of the order of millihartrees are found. For these species the grid construction still has room for improvement, in order to increase the accuracy on the energy sum rule. Similar accuracy values were reported by Mandado et al. for their Hartree-Fock energy partitioning in terms of Hirshfeld atoms [8].

VII. RESULTS

A. Energy decomposition using the single-atom density matrices

Table II displays the total promotion and interaction energies belonging to the (single-index) density matrix fragments (see section III).

	$E^{prom}_{(HF)}$	$E^{int}_{(HF)}$	$-\Delta E^{at}_{(HF)}$	$-\Delta E^{at}_{CC}$		$E^{prom}_{(HF)}$	$E^{int}_{(HF)}$	$-\Delta E^{at}_{(HF)}$	$-\Delta E^{at}_{CC}$
H_2	0.178	-0.311	-0.133	-0.173	CH_2-CH_2	2.240	-2.932	-0.692	-0.881
F_2	0.753	-0.710	0.043	-0.053	CHCH	1.843	-2.321	-0.478	-0.628
Cl_2	0.559	-0.601	-0.041	-0.085	CH_3OH	2.254	-2.855	-0.601	-0.801
Li_2	0.113	-0.119	-0.006	-0.038	H_2CO	1.853	-2.276	-0.423	-0.581
O_2	1.367	-1.353	0.014	-0.131	HCOOH	2.936	-3.485	-0.548	-0.776
N_2	1.598	-1.791	-0.193	-0.344	CO_2	2.606	-3.020	-0.414	-0.598
LiH	0.347	-0.401	-0.054	-0.091	CO	1.292	-1.582	-0.290	-0.399
LiF	0.307	-0.456	-0.149	-0.215	H_2O	1.060	-1.316	-0.256	-0.363
NaCl	0.047	-0.164	-0.117	-0.152	$(H_2O)_2$	2.158	-2.675	-0.517	-0.733
HF	0.446	-0.606	-0.160	-0.221	H_2O_2	1.787	-2.018	-0.232	-0.413
Hcl	0.349	-0.478	-0.129	-0.167	NH_3	1.479	-1.802	-0.324	-0.464
CH_4	1.386	-1.913	-0.527	-0.660	N_2H_4	2.721	-3.156	-0.434	-0.677
CH_3CH_3	2.610	-3.498	-0.887	-1.118					

TABLE II: Total promotion and interaction energies ($E^{prom}_{(HF)}$ and $E^{int}_{(HF)}$) calculated at the ROHF/Aug-cc-pVTZ level of theory. For comparison with the HF atomization energies $\Delta E^{at}_{(HF)}$, the CCSD(T) atomization energies ΔE^{at}_{CC} are also presented. All energies are in Hartree.

The sum of the total promotion and interaction energies yields (minus) the Hartree-Fock atomization energy -$\Delta E^{at}_{(HF)}$; in most cases this represents about 3/4 of the atomization energy calculated at the CCSD(T) [47–52] level of theory -ΔE^{at}_{CC}. Note that in some cases the Hartree-Fock atomization energy is far from satisfying, e.g. it predicts that F_2 has a higher energy then two isolated F-atoms in their ground state. The magnitude of the promotion and interaction energies is within the range of typical IQA (Interacting Quantum Atoms) [21–23] values, studied by Pendas et al. [24] for different methods including the Hirshfeld AIM method.

	at	E_A^0	Q_A	ΔE_A^{CT}	$\Delta E_A^{\overline{S}}$	ΔE_A^{CR}	E_A^{prom}		at	E_A^0	Q_A	ΔE_A^{CT}	$\Delta E_A^{\overline{S}}$	ΔE_A^{CR}	E_A^{prom}
H_2	H	-0.500	0.000	0.000	0.156	-0.067	0.089	H_2CO	C	-37.69	0.47	0.184	0.245	0.642	1.071
F_2	F	-99.403	0.000	0.000	0.245	0.131	0.377		H	-0.50	0.01	0.006	0.154	-0.027	0.133
Cl_2	Cl	-459.481	0.000	0.000	0.148	0.131	0.280		O	-74.81	-0.49	0.009	0.319	0.187	0.515
Li_2	Li	-7.433	0.000	0.000	0.059	-0.002	0.056	HCOOH	C	-37.69	0.95	0.378	0.173	0.902	1.453
O_2	O	-74.807	0.000	0.000	0.447	0.236	0.684		O	-74.81	-0.78	0.015	0.243	0.334	0.592
N_2	N	-54.398	0.000	0.000	0.590	0.209	0.799		O	-74.81	-0.67	0.013	0.271	0.205	0.490
LiH	Li	-7.433	0.969	0.190	0.002	0.130	0.322		H	-0.50	0.00	0.000	0.157	-0.029	0.128
	H	-0.500	-0.969	0.012	0.005	0.008	0.024		H	-0.50	0.51	0.253	0.077	-0.057	0.273
LiF	Li	-7.433	0.977	0.192	0.001	0.175	0.369	CO_2	C	-37.69	1.23	0.602	0.019	1.112	1.733
	F	-99.403	-0.977	-0.047	0.006	-0.020	-0.062		O	-74.81	-0.62	0.012	0.286	0.139	0.437
NaCl	Na	-161.858	0.965	0.176	0.002	0.104	0.281	CO	C	-37.69	0.27	0.107	0.273	0.313	0.694
	Cl	-459.481	-0.965	-0.090	0.005	-0.150	-0.235		O	-74.81	-0.27	0.005	0.377	0.216	0.598
HF	H	-0.500	0.549	0.274	0.071	-0.068	0.276	H_2O	O	-74.81	-0.95	0.018	0.198	0.294	0.510
	F	-99.403	-0.549	-0.027	0.111	0.086	0.170		H	-0.50	0.48	0.238	0.082	-0.045	0.275
Hcl	H	-0.500	0.257	0.128	0.116	-0.042	0.203	$(H_2O)_2$	O	-74.81	-0.98	0.019	0.192	0.302	0.513
	Cl	-459.481	-0.257	-0.024	0.110	0.060	0.146		H	-0.50	0.49	0.244	0.080	-0.039	0.285
CH_4	C	-37.687	-0.474	-0.010	0.353	0.384	0.728		O	-74.81	-0.97	0.019	0.194	0.318	0.530
	H	-0.500	0.118	0.059	0.138	-0.032	0.165		H	-0.50	0.47	0.235	0.083	-0.045	0.274
CH_3CH_3	C	-37.687	-0.256	-0.005	0.335	0.490	0.819		H	-0.50	0.50	0.248	0.079	-0.048	0.278
	H	-0.500	0.085	0.043	0.143	-0.023	0.162	H_2O_2	O	-74.81	-0.43	0.008	0.335	0.289	0.632
	H	-0.500	0.085	0.043	0.143	-0.023	0.162		H	-0.50	0.43	0.215	0.089	-0.043	0.261
CH_2CH_2	C	-37.687	-0.213	-0.004	0.331	0.476	0.803	NH_3	N	-54.40	-1.02	0.082	0.315	0.386	0.783
	H	-0.500	0.107	0.053	0.140	-0.034	0.158		H	-0.50	0.34	0.170	0.103	-0.041	0.232
CHCH	C	-37.687	-0.213	-0.004	0.331	0.431	0.758	N_2H_4	N	-54.40	-0.58	0.043	0.437	0.432	0.912
	H	-0.500	0.213	0.106	0.123	-0.066	0.164		H	-0.50	0.28	0.140	0.112	-0.030	0.222
CH_3OH	C	-37.687	0.104	0.041	0.298	0.595	0.934		H	-0.50	0.30	0.151	0.109	-0.033	0.226
	H	-0.500	0.067	0.033	0.146	-0.025	0.154								
	H	-0.500	0.036	0.018	0.151	-0.018	0.151								
	O	-74.807	-0.683	0.013	0.269	0.317	0.599								
	H	-0.500	0.441	0.220	0.087	-0.042	0.266								

TABLE III: Charge transfer, spin-averaging and charge redistribution energies (ΔE_A^{CT}, $\Delta E_A^{\overline{S}}$ and ΔE_A^{CR}, in Hartree) for the molecular fragments of some small molecules calculated at the ROHF/Aug-cc-pVTZ level of theory. These are the components of the atomic promotion energies E_A^{prom}. The reference energy, E_A^0, is the energy of the neutral isolated atom calculated at the ROHF/Aug-cc-pVTZ level of theory. Q_A is the charge of the single-index atom.

Table III displays the charge transfer (CT), spin averaging (S) and the charge redistribution (CR) components of the individual atomic promotion energies. The current scheme is based on a partitioning of the 1DM. To calculate the energy components of the 1DM fragments, it is necessary to specify the electron spin of the fragments. For singlet molecules, the current scheme averages the electron spin of the 1DM fragments. In an attempt to avoid the spin-averaging step and get lower atomic promotion energies, one could implement an alternative 1DM partitioning scheme and request that e.g. the hydrogen atoms keep their doublet structure in a H_2 molecule. However, such requirement would lead to delocalized electron densities for the AIMs and would exclude consistency with the (well-localized) Hirshfeld-I model. Since the localization of the AIM densities is a necessary condition to produce chemically meaningful results within the IQA approach [24], it is clear that the spin must be averaged at least to some degree. Note that the problem related to the spin is not addressed by the common energy decomposition schemes that are based on a partitioning of the molecular expectation values using AIM weight functions. In these schemes, the underlying electronic structure is not explicitly treated (and possibly not consistent). Energetically, it appears that the spin averaging energy defined in Eq. (24) is of similar importance as the charge redistribution energy. E.g. the average of the absolute values in Table III for ΔE_A^{CT}, $\Delta E_A^{\overline{S}}$ and ΔE_A^{CR} is 0.09, 0.19 and 0.19 Hartree respectively. The charge redistribution (CR) energies are mostly positive, although small negative values are found for the hydrogen atoms. At first sight, one would expect these CR-energies to be positive for variational reasons, since it is based on the atomic HF-Hamiltonian, for which the isolated atom 1DM with fractional charge should be optimal. However, ρ_A is not necessarily N-representable [53], so it is variationally not sufficiently constrained. It is nevertheless remarkable that these negative values are rather small.

In Table IV the interaction energies between the molecular fragments are presented. The Fock component is listed separately,

$$F_{AB}^{int} = \sum_{ijkl} V_{ijkl} \left(\frac{1}{2} (\rho_A)_{il} (\rho_B)_{jk} \right), \tag{47}$$

as it seems to be a robust indicator of the covalent character of that interaction. It clearly singles out the ionic species with a small value, whereas for the (covalent) homonuclear diatomic molecules in the analysis, it correlates linearly with the interaction energy (R^2=0.997). In contrast to the SEDI (bond order) index [54–63] the value is also low for the covalent but weakly bound Li_2.

The bond energies are also listed in Table IV. They do depend on the choice of reference atoms, but chemists are more familiar with their typical magnitudes, and much chemical reasoning

is based on this type of quantity. The linear correlation between the interaction energies and the bond energies is not strong (R^2=0.6). Both quantities can single out chemically not bonded atoms (having very small values) and for both quantities similar values are obtained for chemically similar bonds. The C-H bond has interaction energies between 0.45 Hartree (in CH_4) and 0.40 Hartree in (HCOOH) and bond energies between 0.12 and 0.8 Hartree. Interaction energies for the single, double and triple C-C bonds in ethane, ethene and ethyn are respectively -0.58, -0.98 and -1.38 Hartree, while the corresponding bond energies are 0.10, 0.17 and 0.25 Hartree. The interaction energies are theoretically appealing in the sense that they follow immediately from the density matrix partitioning. They should not be confused with the bond energies, which are more in line with chemical intuition but require the introduction of isolated reference atoms. The analysis presented in this work is particularly useful to investigate substituent effects in molecules. When an atom or functional group is substituted in the molecule, the strength of all bonds in the molecule is affected. The bond energies defined in Eq. (32) are a measure for this effect. E.g. it is clear from the entries in Table IV that the C-H bond in ethene (CH_2CH_2) is stronger than the C-H bond in formaldehyde (CH_2O) and formic acid (HCOOH). It is also possible to compare the interactions between atoms that have no chemical bond in the molecule. Although we recognize that it is possible to get similar information from other schemes that were developed to partition the Hartree-Fock energy, we stress the consistency of this approach, where different energy terms are derived from the same, symmetrical, atom condensed density matrices. Without this mathematical consistency, the results are necessarily ambiguous.

	bond	$SEDI_{AB}$	F_{AB}^{int}	E_{AB}^{int}	E_{AB}^{prom}	E_{AB}^{bond}		bond	$SEDI_{AB}$	F_{AB}^{int}	E_{AB}^{int}	E_{AB}^{prom}	E_{AB}^{bond}
H_2	H-H	1.000	-0.256	-0.311	-0.178	-0.133	H_2CO	C-O	2.293	-0.769	-1.342	-1.147	-0.196
F_2	F-F	1.687	-0.484	-0.710	-0.753	0.043		C-H	0.940	-0.290	-0.414	-0.320	-0.094
Cl_2	Cl-Cl	1.836	-0.404	-0.601	-0.559	-0.041		O...H	0.200	-0.035	-0.043	-0.028	-0.016
Li_2	Li-Li	1.031	-0.097	-0.119	-0.113	-0.006	HCOOH	C-O	1.346	-0.473	-1.028	-0.916	-0.113
O_2	O-O	2.766	-0.853	-1.353	-1.367	0.014		C-O	1.983	-0.700	-1.403	-1.209	-0.194
N_2	N-N	3.361	-1.089	-1.791	-1.791	-1.791		C-H	0.854	-0.270	-0.398	-0.318	-0.079
LiH	Li-H	0.261	-0.062	-0.401	-0.347	-0.054		O-H	0.751	-0.264	-0.624	-0.506	-0.118
LiF	Li-F	0.273	-0.076	-0.456	-0.307	-0.149		O...HC	0.173	-0.038	-0.044	-0.027	-0.017
NaCl	Na-Cl	0.314	-0.058	-0.164	-0.047	-0.117		O...HC	0.199	-0.038	-0.040	-0.023	-0.016
HF	H-F	0.793	-0.272	-0.606	-0.446	-0.160		O...HO	0.036	-0.008	-0.089	-0.070	-0.019
Hcl	H-Cl	1.132	-0.305	-0.478	-0.349	-0.129	CO_2	C-O	2.076	-0.741	-1.522	-1.310	-0.212
CH_4	C-H	0.989	-0.303	-0.455	-0.331	-0.124		O...O	0.459	-0.069	0.024	0.014	0.010
	H...H	0.070	-0.013	-0.016	-0.010	-0.005	CO	C-O	2.883	-0.946	-1.582	-1.292	-0.290
CH_3-CH_3	C-C	1.226	-0.387	-0.585	-0.484	-0.101	H_2O	O-H	0.854	-0.289	-0.688	-0.556	-0.132
	C-H	0.948	-0.295	-0.432	-0.317	-0.115		H...H	0.035	-0.008	0.060	0.053	0.007
	C...H	0.102	-0.021	-0.033	-0.024	-0.009	$(H_2O)_2$	O-H	0.780	-0.274	-0.680	-0.547	-0.133
	$C_1H...HC_1$	0.075	-0.015	-0.019	-0.012	-0.007		O-H	0.861	-0.292	-0.693	-0.556	-0.137
CH_2-CH_2	C-C	2.055	-0.628	-0.984	-0.813	-0.171		O-H	0.824	-0.281	-0.691	-0.564	-0.126
	C-H	0.964	-0.301	-0.439	-0.322	-0.117		O...O	0.110	-0.019	0.151	0.112	0.039
	C...H	0.150	-0.026	-0.041	-0.030	-0.011		O—H	0.074	-0.019	-0.141	-0.112	-0.028
	$C_1H...HC_1$	0.065	-0.012	-0.014	-0.009	-0.005		O...H	0.003	0.000	-0.074	-0.061	-0.013
CH-CH	C-C	2.975	-0.884	-1.383	-1.130	-0.253		O...H	0.004	-0.001	-0.067	-0.054	-0.014
	C-H	0.957	-0.297	-0.433	-0.329	-0.104	H_2O_2	O-O	1.760	-0.530	-0.774	-0.692	-0.082
	C...H	0.146	-0.019	-0.038	-0.029	-0.009		O-H	0.848	-0.292	-0.559	-0.492	-0.067
	H...H	0.006	0.000	0.006	0.004	0.002		O..H	0.100	-0.023	-0.082	-0.072	-0.010
CH_3OH	C-O	1.438	-0.475	-0.785	-0.666	-0.120	NH_3	N-H	0.949	-0.310	-0.619	-0.507	-0.112
	C-H	0.943	-0.294	-0.421	-0.318	-0.103	N_2H_4	N-N	1.700	-0.539	-0.765	-0.699	-0.066
	C-H	0.926	-0.290	-0.414	-0.316	-0.099		N-H	0.936	-0.312	-0.542	-0.457	-0.085
	O-H	0.846	-0.293	-0.616	-0.505	-0.111		N-H	0.916	-0.306	-0.541	-0.457	-0.084
	O...H	0.136	-0.026	-0.039	-0.027	-0.012		N...H	0.109	-0.022	-0.070	-0.059	-0.011
	O...H	0.132	-0.029	-0.049	-0.034	-0.015		N...H	0.108	-0.024	-0.078	-0.066	-0.012

TABLE IV: Interaction energies (E_{AB}^{int}, in Hartree) for the molecular fragments of some small molecules calculated at the ROHF/Aug-cc-pVTZ level of theory. In the column labeled "bond", the notations "-", "−−" and "$\cdots$" indicate respectively a chemical bond, a hydrogen bond and the interaction between a pair of non-bonded atoms. For comparison, the Hirshfeld-I SEDI-index is included as a bond order indicator and the Fock part of the interaction energy (F_{AB}^{int}) seems to be a robust indicator of covalent interaction energy. Atomic promotion energies are attributed to the bonds (E_{AB}^{prom}) to derive Hartree-Fock bond energies (E_{AB}^{bond}).

B. Consistent decomposition of molecular properties

Table V displays the kinetic energies t_A from the single-index atomic density matrices (see Eq. (16)) calculated with the matrix elements of the finite Aug-cc-pVTZ basis set. They are compared to the atom-condensed kinetic energies t_A^h [Eq. (37)] and $t_A^{h'}$ [Eq. (38)], obtained by numerical computation in $\boldsymbol{r}$-space of the molecular kinetic energy integral weighted by the Hirshfeld-I atomic weight functions. We confirmed that all entries in table V are converged with respect to grid size. Note that for the homonuclear diatomic case, all values coincide, since there is only one way to partition the molecular kinetic energy over equivalent atoms.

As discussed in section IV the atom-condensed kinetic energies t_A^h and $t_A^{h'}$ differ significantly (e.g. a difference of 0.3 Hartree for C in CO) pointing to the inherent ambiguity of combining atomic weight functions with the kinetic energy operator. These ambiguities are absent when the molecular 1DM itself is partitioned. For infinite basis sets, values for the kinetic energy of the 1DM fragments t_A should equal values for t_A^h. It is clear from the table that for the present Aug-cc-pVTZ basis set significant differences still occur. For example, the difference $|t_C^h - t_C|$ is more than 0.06 Hartree in CO. We checked for CO that this difference vanishes in the Aug-cc-pVQZ basis set ($|t_C^h - t_C|$ = 0.002 Hartree).

	A	t_A^0	t_A-t_A^0	t_A^h-t_A^0	$t_A^{h'}$-t_A^0		A	t_A^0	t_A-t_A^0	t_A^h-t_A^0	$t_A^{h'}$-t_A^0
H_2	H	0.499	0.066	0.066	0.066	H_2CO	C	37.676	-0.003	-0.080	-0.133
F_2	F	99.397	-0.025	-0.025	-0.025		H	0.499	0.115	0.120	0.247
Cl_2	Cl	459.443	0.010	0.009	0.011		O	74.797	0.193	0.259	0.060
Li_2	Li	7.432	0.003	0.003	0.003	HCOOH	C	37.676	-0.209	-0.335	-0.378
O_2	O	74.797	-0.038	-0.038	-0.038		O	74.797	0.453	0.518	0.275
N_2	N	54.385	0.101	0.102	0.102		O	74.797	0.307	0.370	0.302
LiH	Li	7.432	-0.039	-0.034	-0.225		H	0.499	0.126	0.132	0.292
	H	0.499	0.091	0.087	0.277		H	0.499	-0.139	-0.147	0.047
LiF	Li	7.432	0.007	-0.056	-0.223	CO_2	C	37.676	-0.159	-0.329	-0.329
	F	99.397	0.137	0.200	0.367		O	74.797	0.284	0.369	0.369
NaCl	Na	161.870	0.095	-0.071	-0.201	CO	C	37.676	0.057	0.000	0.285
	Cl	459.443	0.026	0.184	0.323		O	74.797	0.233	0.289	0.004
HF	H	0.499	-0.186	-0.197	0.075	H_2O	O	74.797	0.477	0.502	0.123
	F	99.397	0.337	0.347	0.076		H	0.499	-0.115	-0.127	0.062
Hcl	H	0.499	-0.033	-0.039	0.094	$(H_2O)_2$	O	74.797	0.485	0.510	0.133
	Cl	459.443	0.169	0.174	0.043		H	0.499	-0.127	-0.141	0.059
CH_4	C	37.676	0.264	0.207	-0.119		O	74.797	0.512	0.536	0.155
	H	0.499	0.066	0.080	0.162		H	0.499	-0.111	-0.123	0.067
CH_3CH_3	C	37.676	0.187	0.162	-0.094		H	0.499	-0.126	-0.138	0.046
	H	0.499	0.086	0.095	0.180	H_2O_2	O	74.797	0.187	0.204	-0.028
	H	0.499	0.086	0.095	0.180		H	0.499	-0.079	-0.096	0.136
CH_2CH_2	C	37.676	0.208	0.191	0.015	NH_3	N	54.385	0.422	0.428	0.058
	H	0.499	0.071	0.079	0.167		H	0.499	-0.033	-0.035	0.089
CHCH	C	37.676	0.238	0.229	0.140	N_2H_4	N	54.385	0.211	0.227	-0.065
	H	0.499	0.005	0.014	0.103		H	0.499	0.009	0.001	0.147
CH_3OH	C	37.676	0.051	-0.002	-0.130		H	0.499	-0.002	-0.010	0.135
	H	0.499	0.100	0.103	0.211						
	H	0.499	0.111	0.115	0.224						
	O	74.797	0.313	0.368	-0.038						
	H	0.499	-0.089	-0.102	0.106						

TABLE V: Kinetic energy (t_A) of the molecular Hirshfeld-I fragments compared to the Hirshfeld-I fragmented molecular kinetic energies (t_A^h) and ($t_A^{h'}$). Values are presented relative to the kinetic energy (t_A^0) of the neutral isolated atom. Computations were performed at the ROHF/Aug-cc-pVTZ level of theory.

C. Considerations about computational efficiency

In this section, the computational efficiency of the approach to calculate atomic self-energies and the interaction energies in one-electron Hilbert-space (see Eqs. (16) and (18)) is compared with the computational cost of the approach to compute these quantities using numerical integration in $\boldsymbol{r}$-space (see Eqs. (37) and (41)). When relatively small molecules and basis set sizes are considered, the matrix approach is computationally much less demanding than the $\boldsymbol{r}$-space approach. Figure 1 displays the ratio of the computational costs of both approaches as a function of the number of atoms. In order to obtain sufficiently accurate energy integrals (condensed to atoms and atom pairs) in the $\boldsymbol{r}$-space approach, one must go to very large grids, making the calculations quite time consuming. On the other hand, in the matrix approach, moderate grids suffice to construct the atom condensed atomic overlap matrices and the kinetic energy integrals over the molecular orbital basis are easily computed analytically.

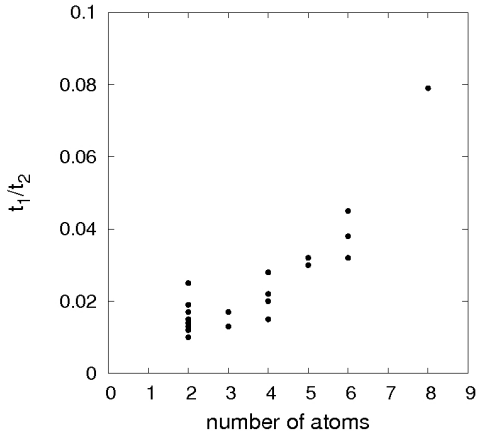


FIG. 1: Calculation times for the energy decomposition (at the ROHF/Aug-cc-pVTZ level of theory) using the matrix approach (t_1) relative to the computational cost of using an r-space approach (t_2) as a function of the number of atoms in the molecule.

Going to larger molecules and very large basis sets (e.g. Aug-cc-pVQZ), at one point the matrix approach will get computationally more demanding (at least at the Hartree Fock level) than the approach of Eq. (37). In the former method all virtual molecular orbitals are included in the

calculation, while in the latter method one only needs the occupied molecular orbitals. However, the crossover point is far from being reached in the present set of molecules.

VIII. CONCLUSIONS

In this paper, we applied a method for decomposing the one-electron molecular density matrix over the atoms in the molecule to calculate atom-condensed energy contributions at the Hartree-Fock level.

In our opinion, this method is preferable to other approaches because it determines (Hartree-Fock) energy terms naturally associated with molecular fragments, whereas most other approaches fragment the molecular energy without assuring that there is an underlying electronic structure from which these energy fragments can be calculated explicitly. Since the Hartree-Fock energy can be expressed directly in terms of the molecular 1DM, energy terms are derived from the molecular 1DM fragments. For the current study, the 1DM fragments are consistent with the local and positive definite Hirshfeld-I partitioning of the electron density. We have shown that without this mathematical consistency, the results are necessarily ambiguous.

Unlike in most cases where numerical integration is used intensively, the new scheme requires only the Hamiltonian matrix elements (one- and two-electron integrals) in Hilbert space, that are routinely computed in ab initio programs, and the atomic overlap integrals. Analysis of the computational efficiency of the new approach versus the more common one based on numerical integration of the molecular energy integrals weighted by atomic weight functions, shows that there is a trade-off between on the one hand the number of integrals that need to be computed and the size of the grids required to reach an acceptable level of accuracy. When using numerical integration for, e.g., the kinetic energy, relatively fewer integrals need to be computed but these are found to require large integration grids. When using the new density matrix formulation, more atomic overlap integrals are required but these can be computed with sufficient accuracy already for moderate grids. Analysis of the computational efficiency shows that the density matrix approach is computationally much more efficient.

It is clear that an energy partitioning at the Hartree-Fock level has a limited applicability and will break down whenever a correct description requires the inclusion of electron correlation. A full energy decomposition at the correlated level actually requires a partitioning of the two-electron density matrix (2DM). This will be the subject of future research.

IX. ACKNOWLEDGEMENTS

We acknowledge support from Dr. Toon Verstraelen (CMM), who provided us with a Hirshfeld-I program. DVF, DVN, PB and DG acknowledge support from the research council (BOF) of Ghent University and FWO-Vlaanderen. PWA acknowledges support from Sharcnet, NSERC, and the Canada Research Chairs.

X. APPENDIX

When the molecular density matrix $\rho(\boldsymbol{r},\boldsymbol{r'}) = \sum_i d_i\psi_i(\boldsymbol{r})\psi_i(\boldsymbol{r'})$ is partitioned, the atom-condensed kinetic energy based on $\rho_A(\boldsymbol{r},\boldsymbol{r'})$ can be written as:

$$\begin{aligned}
t_A &= -\frac{1}{2}\int d\boldsymbol{r}d\boldsymbol{r'}\delta(\boldsymbol{r}-\boldsymbol{r'})\left[\nabla^2\rho_A(\boldsymbol{r},\boldsymbol{r'})\right] \\
&= -\frac{1}{2}\int d\boldsymbol{r}d\boldsymbol{r'}\delta(\boldsymbol{r}-\boldsymbol{r'})\left[\nabla^2\rho(\boldsymbol{r},\boldsymbol{r'})\frac{1}{2}[w_A(\boldsymbol{r})+w_A(\boldsymbol{r'})]\right] \\
&= -\frac{1}{2}\int d\boldsymbol{r}d\boldsymbol{r'}\delta(\boldsymbol{r}-\boldsymbol{r'})\left[\left(\nabla^2\rho(\boldsymbol{r},\boldsymbol{r'})\right)\cdot\frac{1}{2}[w_A(\boldsymbol{r})+w_A(\boldsymbol{r'})]\right. \\
&\quad \left.+\rho(\boldsymbol{r},\boldsymbol{r'})\cdot\frac{1}{2}\nabla^2 w_A(\boldsymbol{r})+\left(\nabla\rho(\boldsymbol{r},\boldsymbol{r'})\right)\cdot\nabla w_A(\boldsymbol{r})\right] \\
&= -\frac{1}{2}\sum_i d_i\int d\boldsymbol{r}\left[\psi_i(\boldsymbol{r})\left(\nabla^2\psi_i(\boldsymbol{r})\right)w_A(\boldsymbol{r})\right. \\
&\quad \left.+\left(\psi_i(\boldsymbol{r})\right)^2\frac{1}{2}\left(\nabla^2 w_A(\boldsymbol{r})\right)+\psi_i(\boldsymbol{r})\left(\nabla\psi_i(\boldsymbol{r})\right)\cdot\nabla w_A(\boldsymbol{r})\right].
\end{aligned} \tag{48}$$

Removing $\nabla^2 w_A(\boldsymbol{r})$ and $\nabla w_A(\boldsymbol{r})$ using partial integration, Eqs. (34) and (37) are shown to be equivalent:

$$\begin{aligned}
t_A &= -\frac{1}{2}\sum_i d_i\int d\boldsymbol{r}w_A(\boldsymbol{r})\left[\psi_i(\boldsymbol{r})\nabla^2\psi_i(\boldsymbol{r})\right. \\
&\quad \left.+\frac{1}{2}\nabla^2\left(\psi_i(\boldsymbol{r})^2\right)-\nabla\left(\psi_i(\boldsymbol{r})\cdot\nabla\psi_i(\boldsymbol{r})\right)\right] \\
&= -\frac{1}{2}\sum_i d_i\int d\boldsymbol{r}w_A(\boldsymbol{r})\left[\psi_i(\boldsymbol{r})\left(\nabla^2\psi_i(\boldsymbol{r})\right)\right] \\
&= -\frac{1}{2}\int d\boldsymbol{r}d\boldsymbol{r'}\delta(\boldsymbol{r}-\boldsymbol{r'})w_A(\boldsymbol{r})\left[\nabla^2\rho(\boldsymbol{r},\boldsymbol{r'})\right] = t_A^h.
\end{aligned} \tag{49}$$

To demonstrate that the atom-condensed kinetic energy based on $\rho_A(\boldsymbol{r},\boldsymbol{r'})$ does not depend on the representation of the kinetic energy operator, we repeat the former exercise starting from a

different respresentation of the kinetic energy operator:

$$\begin{aligned} t_A &= \frac{1}{2}\int d\boldsymbol{r}d\boldsymbol{r'}\delta(\boldsymbol{r}-\boldsymbol{r'})\left[\nabla\cdot\nabla'\rho_A(\boldsymbol{r},\boldsymbol{r'})\right] \\ &= \frac{1}{2}\int d\boldsymbol{r}d\boldsymbol{r'}\delta(\boldsymbol{r}-\boldsymbol{r'})\left[\nabla\cdot\nabla'\rho(\boldsymbol{r},\boldsymbol{r'})\frac{1}{2}[w_A(\boldsymbol{r})+w_A(\boldsymbol{r'})]\right] \\ &= \frac{1}{2}\int d\boldsymbol{r}d\boldsymbol{r'}\delta(\boldsymbol{r}-\boldsymbol{r'})\left[(\nabla\cdot\nabla'\rho(\boldsymbol{r},\boldsymbol{r'}))\cdot\frac{1}{2}[w_A(\boldsymbol{r})+w_A(\boldsymbol{r'})]\right. \\ &\quad \left.+(\nabla\rho(\boldsymbol{r},\boldsymbol{r'}))\cdot\frac{1}{2}\nabla'w_A(\boldsymbol{r'})+(\nabla'\rho(\boldsymbol{r},\boldsymbol{r'}))\cdot\frac{1}{2}\nabla w_A(\boldsymbol{r})\right] \\ &= \frac{1}{2}\sum_i d_i\int d\boldsymbol{r}\left[(\nabla\psi_i(\boldsymbol{r}))^2 w_A(\boldsymbol{r})+\psi_i(\boldsymbol{r})\left(\nabla\psi_i(\boldsymbol{r})\right)\cdot\nabla w_A(\boldsymbol{r})\right]. \end{aligned} \tag{50}$$

Removing $\nabla w_A(\boldsymbol{r})$ using partial integration, Eqs. (35) and (37) are seen to be equivalent as well:

$$\begin{aligned} t_A &= \frac{1}{2}\sum_i d_i\int d\boldsymbol{r}w_A(\boldsymbol{r})\left[(\nabla\psi_i(\boldsymbol{r}))^2-\nabla\cdot(\psi_i(\boldsymbol{r})\nabla\psi_i(\boldsymbol{r}))\right] \\ &= -\frac{1}{2}\sum_i d_i\int d\boldsymbol{r}w_A(\boldsymbol{r})\left[\psi_i(\boldsymbol{r})\nabla^2\psi_i(\boldsymbol{r})\right] \\ &= -\frac{1}{2}\int d\boldsymbol{r}d\boldsymbol{r'}\delta(\boldsymbol{r}-\boldsymbol{r'})w_A(\boldsymbol{r})\left[\nabla^2\rho(\boldsymbol{r},\boldsymbol{r'})\right] = t_A^h. \end{aligned} \tag{51}$$

[1] Mulliken, R. S., J Chem Phys 1955, 23, 1833.

[2] Reed, A. E.; Weinstock, R. B.; Weinhold, F., J Chem Phys 1985, 83, 735.

[3] Bader, R. F. W., Atoms in Molecules: A Quantum Theory, The International Series of Monographs on Chemistry Vol. 22; Oxford University Press: Oxford, 1994.

[4] Mayer, I.; Hamza, A., Theor Chem Acc 2001, 105, 360.

[5] Salvador, P.; Duran, M.; Mayer, I., J Chem Phys 2001, 115, 1153.

[6] Pendás, A. M.; Blanco, M. A.; Francisco, E., J Chem Phys 2004, 120, 4581.

[7] Pendás, A. M.; Francisco, E.; Blanco, M. A., J Comput Chem 2005, 26, 344.

[8] Mandado, M.; Van Alsenoy, C.; Geerlings, P.; De Proft, F.; Mosquera, R. A., Chem Phys Chem 2006, 7, 1294.

[9] Mandado, M.; Krishtal, A.; Van Alsenoy, C.; Bultinck, P.; Hermida-Ramon, J. M., J Phys Chem A 2007, 111, 11885.

[10] Krishtal, A.; Senet, P.; Van Alsenoy, C., J Chem Theory Comput 2008, 4, 426.

[11] Mayer, I.; Hamza, A., Int J Quantum Chem 2005, 103, 798.

[12] Bultinck, P.; Fias, S.; Van Alsenoy, C.; Ayers, P. W.; Carbó-Dorca, R., J Chem Phys 2007, 127, 034102.

[13] Vanfleteren, D.; Van Neck, D.; Bultinck, P.; Ayers, P. W.; Waroquier, M., J Chem Phys 2010, 132, 164111.

[14] Ponec, R.; Cooper, D. L., Faraday Discuss 2007, 135, 31.

[15] Hirshfeld, F. L., Theor Chim Acta 1977, 44, 129.

[16] Bultinck, P.; Van Alsenoy, C.; Ayers, P. W.; Carbó-Dorca, R., J Chem Phys 2007, 126, 144111.

[17] Lillestolen, T. C.; Wheatley, R. J., Chem Commun 2008, 45, 5909.

[18] Bultinck, P.; Cooper, D. L.; Van Neck, D., Phys Chem Chem Phys 2009, 11, 3424.

[19] Bader, R. F. W., Chem Rev 1991, 91, 893.

[20] Popelier, P., Atoms In Molecules: An Introduction; Prentice-Hall: Harlow, 2000.

[21] Blanco, M. A.; Pendás, A. M.; Francisco, E., J Chem Theory Comput 2005, 1, 1096.

[22] Pendás, A. M.; Francisco, E.; Blanco, M. A., J Phys Chem A 2006, 110, 12864.

[23] Pendás, A. M.; Francisco, E.; Blanco, M. A., J Chem Phys 2006, 125, 184112.

[24] Pendás, A. M.; Blanco, M. A.; Francisco, E., J Comput Chem 2007, 28, 161.

[25] Cohen, M. H.; Wasserman, A., Isr J Chem 2003, 43, 219.

[26] Perdew, J. P.; Ruzsinszky, A.; Csonka, G. I.; Vydrov, O. A.; Scuseria, G. E.; Staroverov, V. N.; Tao, J. M., Phys Rev A 2007, 76, 040501.

[27] Cohen, A. J.; Mori-Sanchez, P.; Yang, W. T., Science 2008, 321, 792.

[28] Mori-Sanchez, P.; Cohen, A. J.; Yang, W. T., J Chem Phys 2006, 125, 201102.

[29] Perdew, J. P.; Parr, R. G.; Levi, M.; Balduz, J. L. Jr., Phys Rev Lett 1982, 49, 1691.

[30] Yang, W.; Zhang, Y.; Ayers, P. W., Phys Rev Lett 2000, 84, 5172.

[31] Ayers, P. W., J Math Chem 2008, 43, 285.

[32] Ayers, P. W.; Parr, R. G.; Nagy, A., Int J Quantum Chem 2002, 90, 309.

[33] Anderson, J. S. M.; Ayers, P. W.; Rodriguez Hernandez, J. I., J Phys Chem A, 2010, 114, 8884.

[34] Roothaan, C. C. J., Rev Mod Phys 1951, 23, 69.

[35] Pople, J. A.; Nesbet, R. K., J Chem Phys 1954, 22, 571.

[36] McWeeny, R.; Dierksen, G., J Chem Phys 1968, 49, 4852.

[37] Pople, J. A.; Seeger, R.; Krishnan, R., Int J Quantum Chem 1977, 12, 149.

[38] Raghavachari, K.; Schlegel, H. B.; Pople, J. A., J Chem Phys 1980, 72, 4654.

[39] Raghavachari, K.; Pople, J. A., Int J Quantum Chem 1981, 20, 1067.

[40] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A. Jr., Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A., Gaussian 03, Revision C.02 (Gaussian, Inc., Wallingford CT, 2004).

[41] Lebedev, V. I., Comput Math Math Phys 1975, 15, 44.

[42] Lebedev, V. I., Comput Math Math Phys 1976, 16, 10.

[43] Lebedev, V. I., Sib Math J 1977, 18, 99.

[44] Lebedev, V. I.; Skorokhodov, A. L., Dokl Math 1992, 45, 587.

[45] Lebedev, V. I., Dokl Math 1995, 50, 283.

[46] Lebedev, V. I.; Laikov, D. N., Dokl Math 1999, 59, 477.

[47] Bartlett, R. J.; Purvis, G. D., Int J Quantum Chem 1978, 14, 561.

[48] Pople, J. A.; Krishnan, R.; Schlegel, H. B.; Binkley, J. S., Int J Quantum Chem 1978, 14, 545.

[49] Cizek, J.; Advances in Chem Phys 1969, 14, 35.

[50] Purvis, G. D.; Bartlett, R. J., J Chem Phys 1982, 76, 1910.

[51] Scuseria, G. E.; Janssen, C. L.; Schaefer, H. F., J Chem Phys 1988, 89, 7382.

[52] Scuseria, G. E.; Schaefer, H. F., J Chem Phys 1989, 90, 3700.

[53] Verstichel, B.; van Aggelen, H., Van Neck, D.; Ayers, P. W.; Bultinck, P., Phys Rev A 2009, 80, 032508.

[54] Wiberg, K. B., Tetrahedron 1968, 24, 1083.

[55] Bader, R. F. W.; Stephens, M. E., J Am Chem Soc 1975, 97, 7391.

[56] Giambiagi, M.; Giambiagi, M. S.; Grempel, D. R.; Heymann, C. D. J., J Chim Phys Phys-Chim Biol 1975, 72, 15.

[57] Mayer, I., Chem Phys Lett 1983, 97, 270.

[58] Fulton, R. L., J Phys Chem 1993, 97, 7516.

[59] Angyan, J. G.; Loos, M.; Mayer, I., J Phys Chem 1994, 98, 5244.

[60] Mayer, I.; Salvador, P., Chem Phys Lett 2004, 383, 368.

[61] Fradera, X.; Austen, M. A.; Bader, R. F. W., J Phys Chem A 1999, 103, 304.

[62] Mayer, I., J Comput Chem 2007, 28, 204.

[63] Ponec, R.; Cooper, D. L., J Mol Struct (Theochem) 2005, 727, 133.

Paper IV

Hilbert-space partitioning of the molecular one-electron density matrix with orthogonal projectors: a detailed investigation

D. Vanfleteren, D. Van Neck, P. Bultinck, P. W. Ayers and M. Waroquier

Submitted to The Journal of Chemical Physics, june (2011)

HILBERT-SPACE PARTITIONING OF THE MOLECULAR ONE-ELECTRON DENSITY MATRIX WITH ORTHOGONAL PROJECTORS: A DETAILED INVESTIGATION

Diederik Vanfleteren,[1,2] Dimitri Van Neck,[1,2] Patrick Bultinck,[1,3] Paul W. Ayers,[4] and Michel Waroquier[1,2]

[1]*Members of the Ghent Brussels Quantum Chemistry and Molecular Modelling alliance.*

[2]*Ghent University, Center for Molecular Modeling, Technologiepark 903, B-9052 Zwijnaarde, Belgium*

[3]*Ghent University, Department of Inorganic and Physical Chemistry, Krijgslaan 281 (S3), B-9000 Gent, Belgium*

[4]*McMaster University, Department of Chemistry, Hamilton, Ontario L8S 4M1, Canada*

(Dated: June 17, 2011)

Abstract

A previously introduced partitioning of the molecular one-electron density matrix over atoms and bonds [D. Vanfleteren *et al.*, J. Chem. Phys. **133**, 231103 (2010)] is investigated in detail. Orthogonal projection operators are used to define atomic subspaces, as in the Quantum Chemical Topology method. All calculations are performed as matrix manipulations in one-electron Hilbert space. The orthogonal projection operators are chemically relevant and can be constructed with a recursive scheme, using concepts reminiscent of the Hirshfeld-I partitioning of the electron density. Mathematical proofs and numerical evidence concerning this recursive scheme are provided in the present paper. The advantages associated with the use of orthogonal projection operators are examined with respect to bond orders, bond polarization, covalent bond strengths, and energetics.

I. INTRODUCTION

Many successful theories in chemistry are based on the concept of atoms and bonds in molecules (AIM). Numerous attempts were made to get a quantum mechanical description of these AIM subsystems. Early techniques focussed on determining electron densities for atoms in molecules. This can be done in Hilbert space (the Mulliken population analysis [1]) or in 3D space, using either nonoverlapping atomic basins (Bader's Quantum Chemical Topology [2–4]) or overlapping atomic regions (the original Hirshfeld method [5]). To avoid some shortcomings in the latter method, alternative approaches were developed, like Hirshfeld-I (the iterative version of the Hirshfeld method [6–9]), ISA (the iterative stockholder approach [10]), and Mayer's fuzzy atoms [11].

Since expectation values for one-electron operators are naturally expressed in terms of the one-electron density matrices (1DMs) [12], we recently proposed 1DMs for atoms and bonds in molecules [14]. Because molecular 1DMs are inherently nonlocal, a two-index partitioning into atomic (diagonal) and bond (off-diagonal) contributions is necessary to guarantee the local nature and transferability of the atoms [13, 14]. In a recently introduced scheme [15] for defining the atomic 1DMs from the molecular 1DM, we tried to combine advantages of the different existing techniques. All calculations are performed as matrix manipulations in one-electron Hilbert space, making it highly efficient in terms of computation time, since no grid-based numerical integrations are needed. For interpretative convenience, orthogonal projection operators are used to define atomic subspaces of one-electron Hilbert space. As in the Hirshfeld method, these operators are determined by the fraction of an atomic 1DM with respect to the sum of all atomic 1DMs. The operators can then be used to define density matrices assigned to atoms and bonds in the molecule. An iterative procedure ensures that the method is independent from the initial guess for the atomic 1DMs. The operators can then be used to define density matrices assigned to atoms and bonds in the molecule.

Preliminary results obtained with that scheme are promising [15]: atomic 1DMs are well localized and atomic charges correlate linearly with well-established Hirshfeld-I charges. The 1DM assigned to bonds can be soundly interpreted in terms of depromoting electrons from antibonding to bonding orbitals, and have no contributions from core electrons and free electron pairs. Atomic populations and bond occupancies convergence efficiently with respect to basis set size.

The present paper is a continuation of that previous work [15]. In Sec. II the theoretical back-

ground of the method is established. We provide solid proof for the assumption that chemically relevant orthogonal projection operators can be constructed using a recursive scheme. In Secs. III-IV, we examine the advantages associated with the use of orthogonal projection operators with respect to bond orders, bond polarization, covalent bond strengths and the energetics.

II. THE CONSTRUCTION OF ORTHOGONAL PROJECTORS: RECURSIVE SCHEMES AND PROOFS

In a previous paper [15] we introduced a double-atom partitioning scheme for the molecular spin-summed one-electron density matrix

$$\rho = \sum_{AB} \rho_{AB} \tag{1}$$

in terms of atomic (AA) and bond ($A \neq B$) contributions. All calculations are performed in the one-electron Hilbert space, i.e. all ρ_{AB} are matrices with elements $(\rho_{AB})_{ij}$ where indices i, j refer to an orthogonal basis set. The individual contributions are defined as

$$\rho_{AB} = \frac{1}{2}\left(w_A \rho w_B + w_B \rho w_A\right) \tag{2}$$

in which the concept of atomic weight functions (familiar from Hirshfeld analysis of the electron density) is extended to atomic weight matrices w_A obeying:

$$\delta_{ij} = \sum_{A} (w_A)_{ij}. \tag{3}$$

The atomic weight matrices are constructed as orthogonal projection operators on one-electron atomic subspaces, i.e. they fulfill the relation:

$$w_A w_B = w_A \delta_{AB}. \tag{4}$$

As a consequence, the ρ_{AA} can be regarded as 1DMs of atoms that are excited by the molecular environment but still contain the full number of electrons N,

$$N = Tr(\rho) = \sum_{A} Tr(\rho_{AA}). \tag{5}$$

No electrons are lost to the bonds ($Tr(\rho_{AB}) = 0$ for $A \neq B$), which is a significant advantage in the interpretation of the atomic and bond contributions. We found that, starting from the 1DM's

$\rho_{AA}^{(0)}$ of isolated atoms, the following recursive scheme ($i = 0, 1, ...$)

$$\begin{aligned} \rho^{(i)} &= \sum_A \rho_{AA}^{(i)}\,; \quad w_A^{(i)} = (\rho^{(i)})^{-\frac{1}{2}} \rho_{AA}^{(i)} (\rho^{(i)})^{-\frac{1}{2}} \\ \rho_{AA}^{(i+1)} &= w_A^{(i)} \rho w_A^{(i)} \end{aligned} \tag{6}$$

converges and generates weight matrices $w_A^{(\infty)}$ that both obey Eq. (4) and can be applied in Eq. (2) to produce a chemically relevant decomposition. Upon convergence, $\rho^{(\infty)}$ contains the summed atomic 1DMs and therefore represents the molecular 1DM stripped of its bond contributions.

As in standard Hirshfeld, the choice of the starting point in the recursive scheme of Eq. (6) (the isolated atom 1DM's $\rho_{AA}^{(0)}$) appears to be crucial, since different results are obtained depending on the charge and electronic state of the isolated atom. Therefore, in close analogy to the iterative Hirshfeld procedure [6–9], an outer iterative scheme passes the atomic charges ("charge equalization") or even the populations of the individual orbitals ("population equalization") of $\rho_{AA}^{(\infty)}$ to $\rho_{AA}^{(0)}$ in order to eliminate the dependency on the starting point and to build in self-consistency.

The recursive scheme of Eq. (6) can be generalized and rewritten more compactly,

$$\begin{aligned} w_A^{(i)} &= (\rho^{(i)})^{-\frac{1}{2}} w_A^{(i-1)} M w_A^{(i-1)} (\rho^{(i)})^{-\frac{1}{2}} \\ \rho^{(i)} &= \sum_A w_A^{(i-1)} M w_A^{(i-1)} \end{aligned} \tag{7}$$

where M is in general any positive-definite matrix. To obtain a chemically meaningful decomposition, we found that M can be chosen as either the full molecular 1DM itself ($M = \rho$ as in Eq.(6)), or the recursive estimate for $\rho^{(\infty)}$ (with $M = \rho^{(i-1)}$ in Eq.(7)).

The recursion is applied to transform initial weight matrices to orthogonal projectors on one-electron subspaces (see Eq. (4)). It can be proven that every set of orthogonal projectors is a solution to the convergence limit of Eq.(7)

$$\begin{aligned} w_A^{(\infty)} &= (\rho^{(\infty)})^{-\frac{1}{2}} w_A^{(\infty)} M w_A^{(\infty)} (\rho^{(\infty)})^{-\frac{1}{2}} \\ \rho^{(\infty)} &= \sum_A w_A^{(\infty)} M w_A^{(\infty)} \end{aligned} \tag{8}$$

If M equals $\rho^{(\infty)}$, then these are also the only solutions to Eq.(8). If M equals ρ, then the sets of orthogonal projectors are only part of the entire solution set. This follows from the following lemmas, the proof of which is contained in the appendix:

Lemma II.1 *If $M = \rho^{(\infty)}$ then the converged weight matrices $w_A^{(\infty)}$ are idempotent.*

Lemma II.2 *If the converged weight matrices $w_A^{(\infty)}$ are idempotent, then they are orthogonal.*

Lemma II.3 *When idempotent weight matrices enter the recursion formula, immediate convergence is obtained.*

For $M = \rho$, it is in principle not excluded that particular non-idempotent solutions to Eq.(8) result from the recursion. However, in our investigation only idempotent solutions were obtained for any choice of the starting point. We included numerical evidence in the results section of the paper, demonstrating that for every molecule in the test set the recursion of Eq. (7) always transformed the initial weight matrices into perfect orthogonal projectors.

III. INTERPRETATIVE ADVANTAGES OF PARTITIONING WITH ORTHOGONAL PROJECTORS

A. Non-bonded atoms

1DM partitioning with orthogonal projectors facilitates considerably the interpretation of the atomic and bond matrices. It was already mentioned that the ρ_{AA} can be regarded as 1DMs of atoms that are excited by the molecular environment without loss of electrons to the bonds (Eq. 5). The bond matrices describe the electronic rearrangements that occur when a covalent bond is formed between atoms A and B. From Eq. (4), it is clear that no electrons are shared between different atomic matrices ρ_{AA}:

$$Tr\left(\rho_{AA}^{(\infty)}\rho_{BB}^{(\infty)}\right) = Tr\left((\rho_{AA}^{(\infty)})^2\right)\delta_{AB}. \tag{9}$$

Since the sharing of electrons between the atoms in a molecule is identified with the process of chemical bonding, Eq. (9) is the basis for the interpretation of $\rho_{AA}^{(\infty)}$ as the 1DM of a "non-bonded" atom in a molecular environment. Note that this interpretation holds here for the normal molecular geometry, and has nothing to do with dissociation processes. Also, no contributions of the $\rho_{AA}^{(\infty)}$ can be assigned to atom pairs

$$w_B^{(\infty)}(\rho_{AA}^{(\infty)})w_C^{(\infty)} = \rho_{AA}^{(\infty)}\delta_{AB}\delta_{AC}. \tag{10}$$

Only when orthogonal projectors are used for the 1DM-partitioning, Eqs. (9-10) are trivially satisfied.

B. Bonded atoms

A very different interpretation is attached to the single-index atomic 1DM's $\rho_A^{(\infty)}$, that are derived by single insertion of the orthogonal weight matrices $w_A^{(\infty)}$ into the molecular 1DM

$$\rho_A^{(\infty)} \;=\; \rho^{\frac{1}{2}} w_A^{(\infty)} \rho^{\frac{1}{2}} \qquad\qquad \sum_A \rho_A^{(\infty)} = \rho \tag{11}$$

In contrast to the double-index atomic matrices $\rho_{AA}^{(\infty)}$, these N-representable matrices do share electrons. The number of electrons that is shared between the single-index atoms is equivalent to the number of electrons shared by the bond matrices:

$$Tr\left(\rho_A^{(\infty)} \rho_B^{(\infty)}\right) = Tr\left(\rho_{AB}^{(\infty)} \rho\right) = \frac{1}{2} Tr\left(\rho_{AB}^{(\infty)} \rho_{BA}^{(\infty)}\right) \neq 0. \tag{12}$$

Therefore, the single-index atomic matrices $\rho_A^{(\infty)}$ can be considered as effectively "bonded" within the molecular geometry. The property in Eq. (12) is only trivially satisfied when orthogonal projectors are used.

C. Electron delocalization in "bonded" and "non-bonded" atoms

There are intimate connections between the single-atom (Eq. (11)) and double-atom 1DMs (Eq.(11)). Defining $C = \rho^{\frac{1}{2}} w_A^{(\infty)}$, one can rewrite

$$\rho_A^{(\infty)} = CC^T \tag{13}$$

$$\rho_{AA}^{(\infty)} = C^T C, \tag{14}$$

which implies that both sets of 1DMs share the same eigenvalue spectrum. Numerical evidence suggests that while the orbitals of the double-atom 1DMs are localized, the orbitals of the single-atom 1DM atoms have the shape of typical molecular orbitals (e.g. σ and π). This is consistent with the interpretation of "non-bonded" and "bonded" atoms in the molecule. As an example, Fig. 1 displays the main orbitals of the single-atom and double-atom 1DMs of N_2.

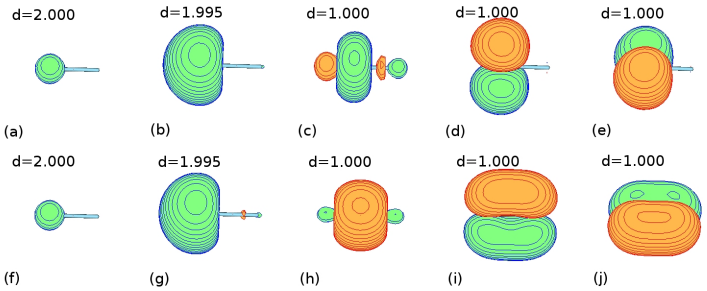


FIG. 1: Dominant natural orbitals and populations of the "non-bonded" $\rho_{AA}^{(\infty)}$ (a-e) and the "bonded" $\rho_{A}^{(\infty)}$ (f-j) in a N_2 molecule at the ROHF/Aug-cc-pVDZ level of theory.

The following reasoning can be useful to understand the localization properties in both sets of atoms. As noted in [14], any single-atom 1DM definition suffers from bad localization properties. For the "bonded" single-atom 1DMs in Eq. (11), the population d_i of a natural orbital φ_i in the molecular 1DM is simply divided over the occupancies $d_{A,i}$ for that orbital in the single-atom 1DM's:

$$\begin{aligned} d_{A,i} = (\rho_A^{(\infty)})_{ii} \; &= \; d_i \langle \varphi_i | w_A^{(\infty)} | \varphi_i \rangle \\ \sum_A d_{A,i} &= d_i. \end{aligned} \tag{15}$$

This implies that only orbitals φ_i that are occupied in the molecule are also occupied in the atoms. Typical chemical compounds have large populations for the bonding orbitals and smaller populations for the antibonding orbitals. Since good localization properties for the atoms are obtained only if the bonding and antibonding molecular orbitals have about equal occupancies in the atom, we conclude that the degree of electron delocalization in the single-index atoms is similar to the degree of electron delocalization in the molecule.

In contrast, the double-index atoms in Eq. (2) have good localization properties. The occupancy $d_{AA,i}$ of each (bonding and antibonding) molecular orbital φ_i in the double-atom 1DM's gets contributions from all populations in the molecular 1DM:

$$d_{AA,i} = (\rho_{AA}^{(\infty)})_{ii} \; = \; \sum_k d_k \langle \varphi_i | w_A^{(\infty)} | \varphi_k \rangle^2. \tag{16}$$

Therefore, similar occupancies are obtained for bonding and antibonding molecular orbitals. This results in excellent localization properties for the double-index atoms.

D. Bond orders and bond polarization

In the orthogonal projector framework, the number of electrons pairs shared between bonded atoms (or bond matrices) may be proposed as the most natural quantum mechanical bond-order index,

$$B_{AB}^{(OP)} = \frac{1}{2} Tr\left(\rho_A^{(\infty)} \rho_B^{(\infty)}\right) = \frac{1}{4} Tr\left(\rho_{AB}^{(\infty)} \rho_{AB}^{(\infty)}\right). \tag{17}$$

It can be compared to a well-established quantum mechanical bond order index like the iterative Hirshfeld SEDI-index [11, 16–24] (see Table II in Sec. IV):

$$B_{AB}^{(HI)} = \int d\boldsymbol{r} d\boldsymbol{r'} w_A^{(HI)}(\boldsymbol{r}) \rho(\boldsymbol{r}, \boldsymbol{r'}) w_B^{(HI)}(\boldsymbol{r'}) \rho(\boldsymbol{r'}, \boldsymbol{r}). \tag{18}$$

In Eq. (18), $w_A^{(HI)}(\boldsymbol{r})$ represents an atomic Hirshfeld-I weight function. For homonuclear diatomic molecules calculated at the Hartree-Fock level, Eq.(17) is extremely close to the classical bond order, as could be expected from Fig. 1.

The deformation matrices $\Delta\rho_A^{(\infty)}$ that transform the "non-bonded" atoms $\rho_{AA}^{(\infty)}$ into the corresponding "bonded" atoms $\rho_A^{(\infty)}$,

$$\Delta\rho_A^{(\infty)} \; = \; \rho_A^{(\infty)} - \rho_{AA}^{(\infty)}, \tag{19}$$

are traceless and in addition may be considered to be atom-condensed bond matrices:

$$\sum_A \Delta\rho_A^{(\infty)} \; = \; \sum_{A,B \neq A} \rho_{AB}^{(\infty)}. \tag{20}$$

Using these atom-condensed bond matrices, it is possible to define the contribution of atom A to the covalent AB bond order as

$$B_{A_{(B)}}^{(OP)} = \frac{1}{2} Tr\left(\Delta\rho_A^{(\infty)} \rho_{AB}^{(\infty)}\right). \tag{21}$$

In case of a diatomic, it is easy to see (using the orthogonal projector relationship) that

$$B_{AB}^{(OP)} = \frac{1}{4} Tr\left(\rho_{AB}^{(\infty)} \rho_{AB}^{(\infty)}\right) = \frac{1}{2} Tr\left((\Delta\rho_A^{(\infty)} + \Delta\rho_B^{(\infty)}) \rho_{AB}^{(\infty)}\right) = B_{A_{(B)}}^{(OP)} + B_{B_{(A)}}^{(OP)}. \tag{22}$$

Numerically, we found that Eq. (22) is also satisfied in the general case. We refer to the Sec. (IV) for more details. From Eq. (21), it is possible to assess the share of each atom in the creation of the bond. Classical chemistry distinguishes between normal covalent bonding and dative covalent bonding when a bond is formed between neutral atoms. In a normal covalent bond, each atom

has an equal share to the bond order. In a dative covalent bond, only one atom provides the electrons for bonding. In the AIM framework, a bond is formed between partially charged atoms and covalent bonds display a continuous spectrum between "normal" and "dative" bonding. From the splitting of the bond order in Eq. (22), it is then possible to assess the "normal" and "dative" character of the covalent bond.

IV. RESULTS

A. Computational methods

For the 1DM partitioning method described in [15] and in section II, numerical tests were performed on the recursive scheme used to construct orthogonal projectors. The molecular 1DMs used in these tests originated from the small set of ca. 67 simple molecules with a singlet ground state representative of a variety of chemical bonds. The list of molecules, that was also used in [15], is presented in Table I. Molecular geometries were obtained from a B3LYP [25–28] /cc-pVDZ [29–31] optimization. The molecular 1DMs were calculated at the restricted Hartree-Fock level of theory using the Aug-cc-pVDZ basis set [29, 32]. To eliminate the starting point dependence of the 1DM partitioning scheme, self-consistency was enforced for the populations of the atomic orbitals [15].

To assess the interpretative advantages associated with the use of orthogonal projectors for the 1DM partitioning, we also calculated covalent bond orders and bond energies for a small subset of mainly diatomic molecules. For that purpose additional calculations were performed at the full-valence CASSCF level of theory, for varying bond lengths.

$AlCl_3$	C_4H_4	CH_3F	CO	$HCOOCH_3$	LiOH	P_2
AlH_3	C_4H_6	CH_3NH_2	CO_2	HF	N_2	P_2H_4
B_2H_6	C_6H_6	CH_3OCH_3	F_2	HNO_2	N_2H_2	PH_3
BeH_2	CCl_4	CH_3OH	H_2	HNO_3	N_2O	S_2
BH_3	CF_4	CH_4	H_2O	HOCl	NaCl	SF_6
C_2H_2	CH_2F_2	CHF_3	H_2S	HOOH	NaOH	Si_2H_6
C_2H_4	CH_2NH	$CHONH_2$	H_2SO_4	HSSH	NH_3	SiH_4
C_2H_6	CH_2O	CHOOH	H_3O^+	Li_2	O_2	
C_3H_3N	CH_2O_2	Cl_2	HCl	LiF	O_3	
C_3H_4	$CH_3CH_2CH_3$	ClF_3	HCN	LiH	OS_2	

TABLE I: List of molecules in the test set.

B. The construction of orthogonal projectors: numerical evidence

In the preceding theoretical discussion, it was assumed that the weight matrices obtained from the recursive scheme Eq. (7) are orthogonal projectors. It was shown that this should be inevitably the case when $M = \rho^{(i)}$. However, for the more elegant implementation of the recursive scheme (when $M = \rho$) , we were unable to prove that the resulting weight matrices are indeed binary, since orthogonal projectors are only a subset of the entire solution set.

Numerical results seem to indicate that the recursive scheme with $M = \rho$ outcompetes the implementation with $M = \rho^{(i)}$, when the starting points are close to orthogonal projectors. In Fig. 2 the numerical accuracy is shown for the relationship that defines orthogonal projectors. For each molecule in the entire test set, all atom pairs AB contribute to a matrix C_{AB} that reduces to the zero matrix for perfect orthogonal projectors:

$$C_{AB}^{(1)} = w_A^{(\infty)} w_B^{(\infty)} - \delta_{AB} w_A^{(\infty)}. \tag{23}$$

The matrix element with a maximal deviation from zero in C_{AB}, for all atom pairs, is chosen to represent the accuracy attained for that molecule. A histogram is created from the integer logarithms of these accuracy measures.

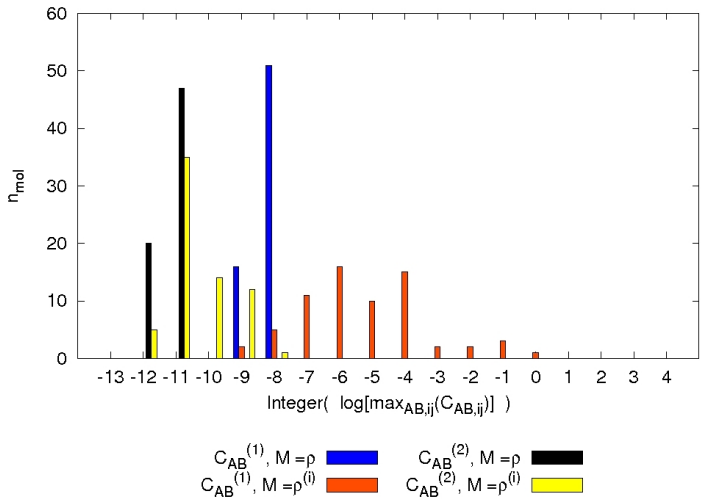


FIG. 2: Attained numerical accuracy in obtaining weight matrices that are orthogonal projectors, through the use of the recursive scheme. For details, see text.

When $M = \rho$ in Eq. (7), we indeed find that the relationship associated with orthogonal projectors is extremely well fulfilled by the weight matrices from the recursive scheme (the largest nonzero element of the C$_{AB}$ matrices being smaller than 10^{-8}). Therefore, we consider it as proven that even with this implementation (for which there is no airtight theoretical proof) the initial weight matrices were transformed into perfect orthogonal projectors for every molecule in the test set. For the other possible implementation of the recursive scheme (where $M = \rho^{(i)}$), it is observed that for only half of the molecules in the test set the largest nonzero element of the C_{AB} matrices is smaller than 10^{-5}. There are even cases for which an element larger than 10^{-1} was found, which is seemingly in contrast to the theoretical prediction that for this implementation all the obtained weight matrices should be orthogonal projectors. The analytical proof assumes, however, the absence of nullspace issues. Although the nullspace of the molecular 1DM was eliminated (by replacing zero eigenvalues with a small value like 10^{-5}), this was not the case for the $\rho^{(i)}$ derived from it. A repeated elimination of the nullspace of $\rho^{(i)}$ is required to make the numerical data coincide with the analytically derived predictions. We note that this nullspace issue hardly affects the atomic 1DM fragments. Multiplying Eq. (23) on both sides with $(\rho^{(\infty)})^{-\frac{1}{2}}$, one

obtains:

$$C_{AB}^{(2)} = \rho_{AA}^{(\infty)}(\rho^{(\infty)})^{-\frac{1}{2}}\rho_{BB}^{(\infty)} - \delta_{AB}\rho_{AA}^{(\infty)}, \tag{24}$$

where $\rho_{AA}^{(\infty)}$ is a shorthand notation for $w_A^{(\infty)} M w_A^{(\infty)}$. Expression 24 is an equivalent orthogonal-projector relationship, but is restricted to the relevant or "occupied" part of space. It is clear from Fig. 2 that the weight matrices from both recursive schemes satisfy these relationships almost perfectly (the largest nonzero element of the C_{AB} matrices being consistently smaller than 10^{-8}).

In Fig. 3 we examined which accuracy was attained for the zero bond traces, a direct consequence of the use of orthogonal projectors. For each molecule in the test set, the largest bond trace is smaller than 10^{-11} when $M = \rho$, and smaller than 10^{-5} when $M = \rho^{(i)}$. The implementation with $M = \rho$ performs consistently better. To show that there are considerable bond matrices for all molecules in the test set (also for the ionic compounds), the largest element in the bond matrices is also presented. There is no molecule for which all bond matrix elements are smaller than 10^{-2}.

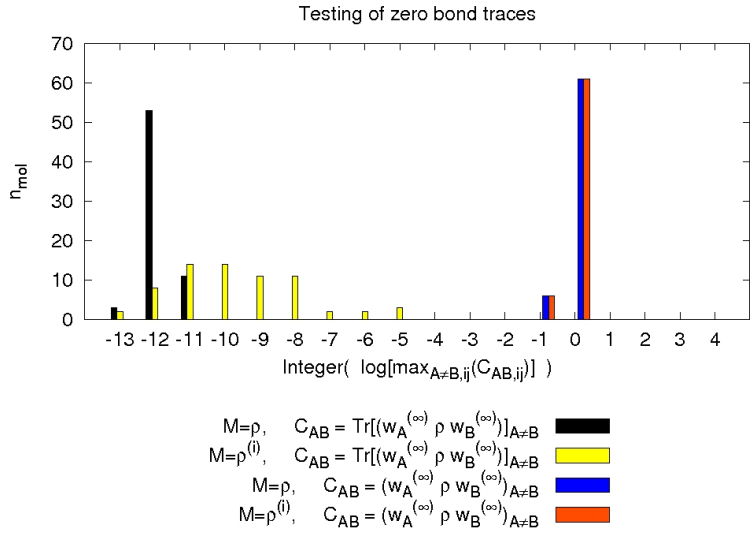


FIG. 3: Attained numerical accuracy in obtaining nonzero bond matrices with a zero trace.

C. Interpretative advantages: examination of bond electrons

Hartree-Fock covalent bond orders are trivial in many cases (e.g. homonuclear diatomic molecules), and their reproduction is an excellent test for any AIM scheme. In Table. II the number of electron pairs shared between "bonded atoms" $B_{AB}^{(OP)}$ (Eq. 17) is compared "to the well-established Hirshfeld-I SEDI-index $B_{AB}^{(HI)}$ at the Hartree-Fock level of theory. For homonuclear diatomic molecules, it is confirmed that values for $B_{AB}^{(OP)}$ are quite close to the integer values of the classical bond orders. In particular for the halogen gases, the expected values are obtained. The Hirshfeld-I SEDI-index seems to overestimate bond orders for these molecules. For ionic compounds, it is satisfying to observe that covalent bond orders in the orthogonal projector framework deviate only a few percent (0.03-0.08) from zero. Hirshfeld-I SEDI-indices for these compounds are much larger (0.26-0.32). For polar covalent compounds, values for the bond order indices are inversely correlated with the charges on the atoms.

	$B_{AB}^{(HI)}$	$B_{AB}^{(OP)}$		$B_{AB}^{(HI)}$	$B_{AB}^{(OP)}$
H_2	1.00	1.00	LiH	0.26	0.08
F_2	1.58	1.01	LiF	0.26	0.05
Cl_2	1.75	1.04	NaCl	0.32	0.03
Li_2	1.03	1.00	HF	0.80	0.48
O_2	2.69	2.02	CO	2.84	2.15
N_2	3.33	3.02	HCl	1.12	0.91
S_2	2.77	2.04	H_2O	0.88	0.65
P_2	3.29	3.01			

TABLE II: Comparison of the covalent bond orders obtained in the orthogonal projector framework ($B_{AB}^{(OP)}$) with the well established iterative-Hirshfeld SEDI-index ($B_{AB}^{(HI)}$). Values, calculated at the ROHF/Aug-cc-pVDZ level, are shown for the diatomic molecules in the testset.

	Q_A	Q_B	$B_{AB}^{(OP)}$	$\frac{B_{A_{(B)}}^{(OP)}}{B_{AB}^{(OP)}}$	$\frac{B_{B_{(A)}}^{(OP)}}{B_{AB}^{(OP)}}$	$\frac{\widetilde{B_{A_{(B)}}^{(OP)}}}{B_{AB}^{(OP)}}$	$\frac{\widetilde{B_{B_{(A)}}^{(OP)}}}{B_{AB}^{(OP)}}$
LiH	0.96	-0.96	0.08	0.02	0.97	0.02	0.98
LiF	0.97	-0.97	0.05	0.01	0.98	0.01	0.99
NaCl	0.98	-0.98	0.03	0.01	0.99	0.01	0.99
HF	0.70	-0.70	0.48	0.12	0.87	0.14	0.86
CO	0.57	-0.57	2.15	0.24	0.75	0.24	0.76
HCl	0.34	-0.34	0.91	0.33	0.67	0.34	0.66

TABLE III: The contribution of the individual atoms to the covalent bond, for the diatomic molecules in the testset calculated at the Hartree-Fock level. The covalent bond order $B_{AB}^{(OP)}$ and its contributions from the atoms $B_{A_{(B)}}^{(OP)}$ are determined within the orthogonal projector framework. The atomic contributions can also be estimated with a simple formule ($\widetilde{B_{A_{(B)}}^{(OP)}}$) from the atomic charges Q_A.

In Table III, the contribution of the individual atoms to the bond (order) is examined (see Eq. 21). The presented values for $B_{A_{(B)}}^{(OP)}$ indicate that in the AIM framework, the atoms do not contribute equally to the covalent bond. The contribution is mainly determined by the atomic charge and can be approximated (in the diatomic case) with the simple formula:

$$\widetilde{B_{A_{(B)}}^{(OP)}} = \frac{1}{2}\Big((b.o.)_{AB} - Q_A + F_A \Big) B_{AB}^{(OP)}. \tag{25}$$

where $(b.o.)_{AB}$ is the classical number of bonds between the atoms (covalent or ionic), Q_A is the atomic charge determined from the AIM and F_A is the formal charge. Note that for zero charges on the atoms, the classical number of dative bonds is obtained as $|\widetilde{B_{A_{(B)}}^{(OP)}} - \widetilde{B_{B_{(A)}}^{(OP)}}|$.

At the post-Hartree-Fock level, the number of electron pairs shared between the 1DMs ${\rho_A}^{(\infty)}$of the bonded atoms is -in general- lower than in Hartree-Fock (see Table IV). This can be roughly understood by considering that previously unoccupied orbitals in these atoms now have a small occupation, but contribute less efficiently to the bond order index than the previously fully occupied orbitals (with e.g. σ and π shapes) whose occupation is now somewhat reduced. For the polar bonds, it is observed that some covalent bond order indices are larger with respect to the Hartree-Fock case. This is related to the less polar character of the bonds in these compounds at the correlated level (smaller atomic charges).

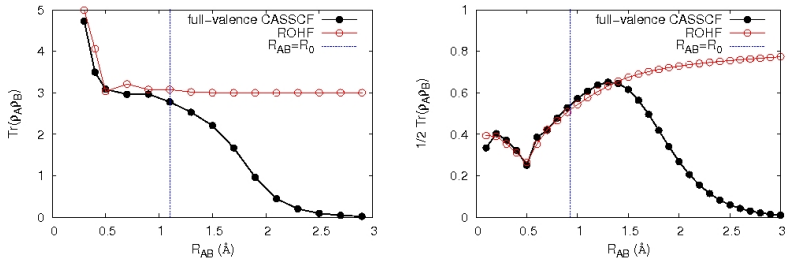
	$B_{AB}^{(OP)}$		$B_{AB}^{(OP)}$
H_2	0.95	LiH	0.13
F_2	0.77	LiF	0.08
Cl_2	0.93	NaCl	0.05
Li_2	0.79	HF	0.54
O_2	1.35	CO	2.15
N_2	2.78	HCl	0.87
S_2	1.44		
P_2	2.64		

TABLE IV: Covalent bond orders obtained in the orthogonal projector framework ($B_{AB}^{(OP)}$) for the diatomic molecules in the testset. Values are calculated at the full-valence CASSCF level of theory.

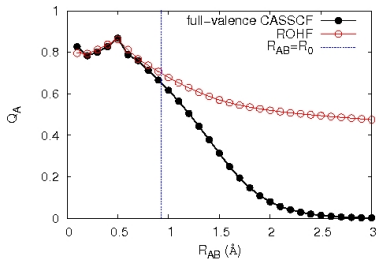
Calculations at the post-Hartree-Fock level are required to obtain reasonable indices for the covalent bond orders of molecules undergoing bond breaking and formation. As illustrated in Fig. 4 for N_2 and HF, Eq. (17) has the correct dissociation behavior at the full-valence CASSCF level, whereas the bond order index incorrectly evolves towards a constant nonzero value in the Hartree-Fock dissociation limit. Fig. 4(b) displays the covalent bond order of the polar HF molecule for bond lengths R_{AB} varying from the equilibrium bond length R_0. Fig. 4(c) displays the associated charge on the H-atom. Values for very short internuclear distances are chemically not relevant. It appears that covalent bond orders initially rise when the molecule starts dissociating, because atomic charges decrease and the molecule becomes less polar. At the dissociation limit, the CASSCF bond order and the atomic charges correctly drop to zero. The ROHF bond order and atomic charges remain nonzero for large internuclear distances.

D. Interpretive advantages: examination of bond matrix energy

Within the Hartree-Fock approach, the molecular energy can be written in terms of the molecular 1DM or its fragments. A meaningful fragmentation scheme generates predictable trends for



(a)Covalent bond order in N_2 (b)Covalent bond order in HF



(c)Charge on the H-atom in HF

FIG. 4: Covalent bond order indices $B_{AB}^{(OP)}$ in the N_2 and HF molecules for bond lengths R_{AB} varying from the equilibrium (B3LYP/Aug-cc-pVDZ) bond length R_0. The covalent bond order index was determined at the ROHF/Aug-cc-pVDZ and full-valence CASSCF/Aug-cc-pVDZ levels of theory.

the energy of its fragments. We introduce the following notations:

$$\begin{aligned} t_{ij} &= \int \phi_i(\boldsymbol{r})\hat{t}\phi_j(\boldsymbol{r})d\boldsymbol{r} \\ (V_A^{ext})_{ij} &= \int \phi_i(\boldsymbol{r})\frac{Z_A}{|\boldsymbol{r}-\boldsymbol{R}_A|}\phi_j(\boldsymbol{r})d\boldsymbol{r} \\ V_{ijkl} &= \int \phi_i(\boldsymbol{r}_1)\phi_j(\boldsymbol{r}_2)\frac{1}{|\boldsymbol{r}_1-\boldsymbol{r}_2|}\phi_k(\boldsymbol{r}_1)\phi_l(\boldsymbol{r}_2)d\boldsymbol{r_1}d\boldsymbol{r_2} \end{aligned} \tag{26}$$

In the latter equations, the ϕ_i represent real functions of an orthogonal basis set, $\hat{t}$ is the kinetic energy operator and Z_A is the nuclear charge on an atom A with nuclear coordinate $\boldsymbol{R}_A$. We

identify the total energy associated with the bond contributions as the energy difference between the molecular 1DM with and without its bond contributions,

$$\sum_{A<B} E_{AB} = E_{mol} - E\left[\sum_{A} \rho_{AA}^{(\infty)}\right]. \tag{27}$$

For the closed-shell Hartree-Fock case, the energy of the molecular 1DM without its bond contributions is readily expressed in one-electron Hilbert space as:

$$\begin{aligned} E\ \left[\sum_{A} \rho_{AA}^{(\infty)}\right] &= \sum_{ij} \left(t_{ij} + (\sum_{C} V_C^{ext})_{ij}\right) \left(\sum_{A} \rho_{AA}^{(\infty)}\right)_{ij} \\ &+\ \frac{1}{2}\sum_{ijkl} V_{ijkl} \left(\left(\sum_{A} \rho_{AA}^{(\infty)}\right)_{ik} \left(\sum_{A} \rho_{AA}^{(\infty)}\right)_{jl} - \frac{1}{2}\left(\sum_{A} \rho_{AA}^{(\infty)}\right)_{il} \left(\sum_{A} \rho_{AA}^{(\infty)}\right)_{jk} \right), \end{aligned} \tag{28}$$

where t_{ij}, $(V_C^{ext})_{ij}$ and V_{ijkl} are defined in Eq. (26). We introduce $\rho_{AB}^{\prime(\infty)} = \rho_{AB}^{(\infty)} + \rho_{BA}^{(\infty)}$ as a shorthand notation for the bond contributions of the molecular 1DM, to define the individual bond matrix energies E_{AB}:

$$\begin{aligned} E_{AB} \ &= \ \sum_{ij} \left(t_{ij} + (\sum_{C} V_C^{ext})_{ij}\right) (\rho_{AB}^{\prime(\infty)})_{ij} \\ &+\ \sum_{ijkl} V_{ijkl} \sum_{C} \left((\rho_{AB}^{\prime(\infty)})_{ik} (\rho_{CC}^{(\infty)})_{jl} - \frac{1}{2} (\rho_{AB}^{\prime(\infty)})_{il} (\rho_{CC}^{(\infty)})_{jk}) \right) \\ &+\ \frac{1}{2}\sum_{ijkl} V_{ijkl} \sum_{C<D} \left((\rho_{AB}^{\prime(\infty)})_{ik} (\rho_{CD}^{\prime(\infty)})_{jl} - \frac{1}{2} (\rho_{AB}^{\prime(\infty)})_{il} (\rho_{CD}^{\prime(\infty)})_{jk} \right). \end{aligned} \tag{29}$$

The bond matrix energy E_{AB} includes all interactions of $\rho_{AB}^{\prime(\infty)}$ with the atomic contributions and half of the interactions with other bond contributions.

In Table V we examine trends for the bond matrix energy (E_{AB}). These energies are quite large, as they can be interpreted as the energy related to the breaking of a (covalent) bond at the normal molecular geometry (without dissociation of the atoms). Table V also contains values for the exchange interaction f_{AB} between the bond matrices $\rho_{AB}^{(\infty)}$ and $\rho_{BA}^{(\infty)}$. This quantity is closely related to the covalent bond order $B_{AB}^{(OP)}$ and seems to be a robust indicator of the covalent bond strength

$$f_{AB} = -\frac{1}{4}\sum_{ijkl} V_{ijkl} \left((\rho_{AB}^{(\infty)})_{il} (\rho_{BA}^{(\infty)})_{jk} + (\rho_{BA}^{(\infty)})_{il} (\rho_{AB}^{(\infty)})_{jk} \right). \tag{30}$$

Both for the bond matrix energy E_{AB} and the indicator of the covalent bond strength f_{AB} trends are in accordance with chemical intuition. For ionic compounds, it is satisfying to observe that

values amount to only a few percent of the values for typical covalent molecules. For homonuclear diatomic molecules, values are clearly dependent on the bond order ($Li_2, F_2, H_2, Cl_2 < S_2, O_2 < P_2, N_2$) and bonding is less efficient between third row elements than between their second row equivalents ($Cl_2 < F_2$; $S_2 < O_2$; $P_2 < N_2$). When values are weighted by the bond order, it is clear that polar compounds have relatively strong covalent bonds (e.g. $H_2, F_2 < HF$) and the surplus depends on the polarity.

	E_{AB}	f_{AB}		E_{AB}	f_{AB}
H_2	-0.88	-0.24	LiH	-0.08	-0.01
F_2	-0.85	-0.25	LiF	-0.04	-0.01
Cl_2	-0.71	-0.19	NaCl	-0.02	0.00
Li_2	-0.18	-0.09	HF	-0.78	-0.17
O_2	-2.24	-0.56	CO	-2.99	-0.63
N_2	-3.96	-0.88	HCl	-0.96	-0.26
S_2	-1.37	-0.38	H_2-O	-1.00	-0.21
P_2	-1.84	-0.52			

TABLE V: Bond matrix energy E_{AB} (a.u.) and energetic covalency measures f_{AB} for the diatomic molecules at the Hartree-Fock level of theory

V. CONCLUSIONS

Recently, we developed a double-atom partitioning of the molecular 1DM to describe atoms and bonds. All calculations are performed in Hilbert space. The concept of atomic weight functions (familiar from Hirshfeld analysis of the electron density) is extended to atomic weight matrices. These are constructed to be orthogonal projection operators on atomic subspaces, which has significant advantages in the interpretation of the bond contributions.

In the current paper, we provide evidence for the assumption that chemically relevant orthogonal projection operators can be constructed using a recursive scheme. Numerical evidence was supplied for the recursive scheme used in [15] and solid theoretical proof could be obtained for a slightly adapted scheme.

We also examined the advantages associated with the use of orthogonal projection operators

with respect to bond orders, bond polarization, covalent bond strengths and energetics. The Hartree-Fock (HF) level of theory was applied to test the ability of the scheme to reproduce well-established information and known trends. In contrast to HF bond orders from a Hirshfeld-I analysis, HF covalent bond orders from the current scheme are small for ionic compounds and integer values are reproduced for homonuclear diatomic molecules. At the full-valence CASSCF level, covalent bond orders approximate typical Hartree-Fock values for equilibrium bond lengths, but (correctly) fall to zero at the dissociation limit. In the framework of orthogonal projectors, it is possible to assess the contributions from the individual atoms to the covalent bond (order). These contributions provide a measure for the "dative" character of covalent bonds. Since the Fock energy between bond matrices is closely related to the covalent bond order, this Fock energy was identified as a robust indicator of the covalent bond strength. Predictable trends were observed in the values of this covalent bond strength indicator, as well as for the bond matrix energies.

VI. ACKNOWLEDGEMENTS

DVF, DVN, PB and DG acknowledge support from the research council (BOF) of Ghent University and FWO-Vlaanderen. PWA acknowledges support from Sharcnet, NSERC, and the Canada Research Chairs.

VII. APPENDIX

Lemma II.1 *If M=$\rho^{(i)}$ in Eq.(7), then the converged weight matrices $w_A^{(\infty)}$ are idempotent.*

Proof: If M=$\rho^{(i)}$ in Eq.(7), then we can use the shorthand notation,

$$w_A^{(\infty)}\rho^{(\infty)}w_A^{(\infty)} = \rho_{AA}^{'(\infty)}, \tag{31}$$

to write down the converged solution of the recursive scheme in Eq.(7):

$$w_A^{(\infty)} = (\rho^{(\infty)})^{-\frac{1}{2}}\rho_{AA}^{'(\infty)}(\rho^{(\infty)})^{-\frac{1}{2}} \tag{32}$$

$$\rho^{(\infty)} = \sum_A \rho_{AA}^{'(\infty)}. \tag{33}$$

By substituting Eq. (32) in Eq. (44) we obtain an equation,

$$(\rho^{(\infty)})^{-\frac{1}{2}}\rho_{AA}^{'(\infty)}\rho_{AA}^{'(\infty)}(\rho^{(\infty)})^{-\frac{1}{2}} = \rho_{AA}^{'(\infty)}, \tag{34}$$

that can be multiplied on both sides by $\rho_{AA}^{'(\infty)}$ so that the square root of the resulting expression becomes:

$$\rho_{AA}^{'(\infty)}(\rho^{(\infty)})^{-\frac{1}{2}}\rho_{AA}^{'(\infty)} = (\rho_{AA}^{'(\infty)})^{\frac{3}{2}}. \tag{35}$$

After multiplying the left side of Eq. (34) by $\rho_{AA}^{'(\infty)}(\rho^{(\infty)})^{-\frac{1}{2}}$ and the right side by $(\rho^{(\infty)})^{-\frac{1}{2}}\rho_{AA}^{'(\infty)}$, the square root of that expression can be simplified on the right hand side by twice substituting Eq. (35) into it:

$$\begin{aligned}
\rho_{AA}^{'(\infty)}(\rho^{(\infty)})^{-1}\rho_{AA}^{'(\infty)} &= \left(\rho_{AA}^{'(\infty)}(\rho^{(\infty)})^{-\frac{1}{2}}\rho_{AA}^{'(\infty)}(\rho^{(\infty)})^{-\frac{1}{2}}\rho_{AA}^{'(\infty)}\right)^{\frac{1}{2}} \\
&= \left((\rho_{AA}^{'(\infty)})^{\frac{3}{2}}(\rho^{(\infty)})^{-\frac{1}{2}}\rho_{AA}^{'(\infty)}\right)^{\frac{1}{2}} \\
&= \left((\rho_{AA}^{'(\infty)})^{\frac{1}{2}}(\rho_{AA}^{'(\infty)})^{\frac{3}{2}}\right)^{\frac{1}{2}} \\
&= \rho_{AA}^{'(\infty)}.
\end{aligned} \tag{36}$$

A simple multiplication of Eq. (36) on both sides with $(\rho^{(\infty)})^{-\frac{1}{2}}$ retrieves

$$(w_A^{(\infty)})^2 = w_A^{(\infty)}, \tag{37}$$

establishing that the atomic weight matrices are idempotent or binary in Hilbert space.

Lemma II.2 *If the converged weight matrices $w_A^{(\infty)}$ are idempotent, then they are orthogonal.*

Proof:

Suppose that $x_{B,j}$ is an normalized eigenvector of $w_B^{(\infty)}$ with unit eigenvalue. Then

$$x_{B,j}^T \left(\sum_A w_A^{(\infty)} \right) x_{B,j} = x_{B,j}^T I x_{B,j} = x_{B,j}^T x_{B,j} = 1. \tag{38}$$

The left hand side of Eq. (38) can be rewritten,

$$\begin{aligned} x_{B,j}^T \left(w_B^{(\infty)} \right) x_{B,j} + x_{B,j}^T \left(\sum_{A \neq B} w_A^{(\infty)} \right) x_{B,j} &= 1 \\ 1 + x_{B,j}^T \left(\sum_{A \neq B} w_A^{(\infty)} \right) x_{B,j} &= 1, \end{aligned} \tag{39}$$

and therefore we have

$$\sum_{A \neq B} x_{B,j}^T \left(w_A^{(\infty)} \right) x_{B,j} = 0. \tag{40}$$

However, since the atomic weight matrices are positive semidefinite by construction, this implies that for every A $\neq$ B:

$$\begin{aligned} x_{B,j}^T \left(w_A^{(\infty)} \right) x_{B,j} &= 0 \\ \left(w_A^{(\infty)} \right) x_{B,j} &= \mathbf{0}. \end{aligned} \tag{41}$$

And so any eigenvector of $w_B^{(\infty)}$ with unit eigenvalue is an eigenvector of all the other weight matrices with zero eigenvalue. An explicit way to show that binary weight matrices are orthogonal is to expand an arbitrary trial vector in the eigen-basis of $w_B^{(\infty)}$. Then

$$w_A^{(\infty)} w_B^{(\infty)} \left(\sum_j c_j x_{B,j} \right) = \mathbf{0} \tag{42}$$

because either ($w_B^{(\infty)} x_{B,j} = \mathbf{0}$) or ($w_B^{(\infty)} x_{B,j} = x_{B,j}$ and $w_A^{(\infty)} x_{B,j} = \mathbf{0}$).
Since Eq. (42) is true for every trial vector, this implies that the weight matrices must be orthogonal:

$$w_A^{(\infty)} w_B^{(\infty)} = \mathbf{0}. \tag{43}$$

Lemma II.3 *When idempotent weight matrices enter the recursive scheme of Eq.(7), immediate convergence is obtained.*

Proof:

With the shorthand notation,

$$\rho_{AA}^{'(i)} = w_A^{(i-1)} M w_A^{(i-1)}, \tag{44}$$

and the assumption that for iteration (i-1) the weight matrices are binary [$w_A^{(i-1)} = (w_A^{(i-1)})^2$], the following equation is obtained:

$$\rho_{AA}^{'(i)} = w_A^{(i-1)} \rho_{AA}^{'(i)} w_A^{(i-1)}. \tag{45}$$

When both sides of this equation are multiplied by $(\rho_{AA}^{'(i)})^{\frac{1}{2}}$, and the square root of the resulting expression is multiplied once more on both sides by $(\rho_{AA}^{'(i)})^{\frac{1}{2}}$, we arrive at:

$$(\rho_{AA}^{'(i)})^2 = \rho_{AA}^{'(i)} w_A^{(i-1)} \rho_{AA}^{'(i)}. \tag{46}$$

Since the weight matrices are binary and therefore orthogonal, if follows from $\rho^{(i)} = \sum_A \rho_{AA}^{'(i)}$that,

$$(\rho_{AA}^{'(i)})^2 = \rho^{(i)} w_A^{(i-1)} \rho^{(i)}, \tag{47}$$

and the result can be rewritten in terms of the weight matrices for iteration (i-1) and (i):

$$\begin{aligned} w_A^{(i-1)} &= (\rho^{(i)})^{-1} (\rho_{AA}^{'(i)})^2 (\rho^{(i)})^{-1} \\ w_A^{(i-1)} &= (\rho^{(i)})^{-\frac{1}{2}} \left((\rho^{(i)})^{-\frac{1}{2}} \rho_{AA}^{'(i)} (\rho^{(i)})^{-\frac{1}{2}} \right) \rho^{(i)} \left((\rho^{(i)})^{-\frac{1}{2}} \rho_{AA}^{'(i)} (\rho^{(i)})^{-\frac{1}{2}} \right) (\rho^{(i)})^{-\frac{1}{2}} \\ w_A^{(i-1)} &= (\rho^{(i)})^{-\frac{1}{2}} w_A^{(i)} \rho^{(i)} w_A^{(i)} (\rho^{(i)})^{-\frac{1}{2}}. \end{aligned} \tag{48}$$

A similar relationship between the weight matrices for iteration (i-1) and (i) can be obtained, starting from the definitions of $w_A^{(i)}$ and $\rho_{AA}^{'(i)}$:

$$\begin{aligned} w_A^{(i)} &= (\rho^{(i)})^{-\frac{1}{2}} \rho_{AA}^{'(i)} (\rho^{(i)})^{-\frac{1}{2}} \\ w_A^{(i)} &= (\rho^{(i)})^{-\frac{1}{2}} w_A^{(i-1)} M w_A^{(i-1)} (\rho^{(i)})^{-\frac{1}{2}}. \end{aligned} \tag{49}$$

Matrix M can be substituted, using the idempotency and orthogonality for the weight functions,

$$w_A^{(i-1)} M w_A^{(i-1)} = (w_A^{(i-1)})^2 M (w_A^{(i-1)})^2 = (w_A^{(i-1)}) \rho_{AA}^{'(i)} (w_A^{(i-1)}) = (w_A^{(i-1)}) \rho^{'(i)} (w_A^{(i-1)}), \tag{50}$$

after which the relationship between the weight matrices for iteration (i-1) and (i) (Eq. (49)) becomes:

$$w_A^{(i)} = (\rho^{(i)})^{-\frac{1}{2}} w_A^{(i-1)} \rho^{(i)} w_A^{(i-1)} (\rho^{(i)})^{-\frac{1}{2}}. \tag{51}$$

Eq. (48) and Eq. (51) are equal apart from interchanging $w_A^{(i-1)}$ and $w_A^{(i)}$. Therefore, when the system of these equations is solved with the substitution method, it appears that

$$w_A^{(i-1)} = w_A^{(i)}. \tag{52}$$

Lemma II.4 *If M=ρ in the recursive scheme of Eq.(7), then it is in principle not excluded that particular non-idempotent solutions $w_A^{(\infty)}$ result from the recursion.*

Proof:

The convergence limit of the iterative scheme is:

$$\begin{aligned} w_A^{(\infty)} &= (\rho^{(\infty)})^{-\frac{1}{2}} w_A^{(\infty)} M w_A^{(\infty)} (\rho^{(\infty)})^{-\frac{1}{2}} \\ \rho^{(\infty)} &= \sum_A w_A^{(\infty)} M w_A^{(\infty)}. \end{aligned} \tag{53}$$

We have, then, that

$$\left(\rho^{(\infty)}\right)^{\frac{1}{2}} w_A^{(\infty)} \left(\rho^{(\infty)}\right)^{\frac{1}{2}} = w_A^{(\infty)} M w_A^{(\infty)}. \tag{54}$$

Consider a system with k fragments. Suppose that all the fragments had the same weight matrix,

$$w_A^{(\infty)} = \frac{1}{k} I. \tag{55}$$

Then

$$\rho^{(\infty)} = \sum_A \left(\frac{1}{k} I\right) M \left(\frac{1}{k} I\right) = \frac{1}{k} M. \tag{56}$$

Inserting into Eq. (54) reveals that this is also a particular solution,

$$\begin{aligned} \left(\rho^{(\infty)}\right)^{\frac{1}{2}} w_A^{(\infty)} \left(\rho^{(\infty)}\right)^{\frac{1}{2}} &= w_A^{(\infty)} M w_A^{(\infty)} \\ \left(\frac{1}{k} M\right)^{\frac{1}{2}} \left(\frac{1}{k} I\right) \left(\frac{1}{k} M\right)^{\frac{1}{2}} &= \left(\frac{1}{k} I\right) M \left(\frac{1}{k} I\right) \\ \frac{1}{k^2} M &= \frac{1}{k^2} M. \end{aligned} \tag{57}$$

[1] R. S. Mulliken, *J. Chem. Phys.* **23**, 1833 (1955).

[2] R. F. W. Bader, *Chem. Rev.* **91**, 893 (1991).

[3] R. F. W. Bader, *Atoms in Molecules: A Quantum Theory.* The International Series of Mono-graphs on Chemistry Vol. 22 (Oxford University Press, Oxford, 1994).

[4] P. Popelier, *Atoms In Molecules: An Introduction*. (Prentice-Hall, Harlow, Great Britain, 2000).

[5] F. L. Hirshfeld, *Theor. Chim. Acta* **44**, 129 (1977).

[6] P. Bultinck, C. Van Alsenoy, P. W. Ayers and R. Carbó-Dorca, *J. Chem. Phys.* **126**, 144111 (2007).

[7] P. Bultinck, *Faraday Discussions* **135**, 244 (2007).

[8] P. Bultinck, P. W. Ayers, S. Fias, K. Tiels, and C. Van Alsenoy, *Chem. Phys. Lett.* **444**, 205 (2007).

[9] P. Bultinck, D. L. Cooper, and D. Van Neck, *Phys. Chem. Chem. Phys.* **11**, 3424 (2009).

[10] T. C. Lillestolen and R. J. Wheatley, *Chem. Commun.* **45**, 5909 (2008).

[11] I. Mayer and P. Salvador, *Chem. Phys. Lett.* **383**, 368 (2004).

[12] L. Li and R. G. Parr, *J. Chem. Phys.* **84**, 1704 (1986).

[13] I. Mayer and P. Salvador, *J. Chem. Phys.* **130**, 234106 (2009).

[14] D. Vanfleteren, D. Van Neck, P. Bultinck, P. W. Ayers and M. Waroquier, *J. Chem. Phys.* **132**, 164111 (2010).

[15] D. Vanfleteren, D. Van Neck, P. Bultinck, P. W. Ayers and M. Waroquier, Hilbert-space partitioning of the molecular one-electron density matrix with orthogonal projectors *J. Chem. Phys.* accepted for publication.

[16] K. B. Wiberg, *Tetrahedron* **24**, 1083 (1968).

[17] R. F. W. Bader and M. E. Stephens, *J. Am. Chem. Soc.* **97**, 7391 (1975).

[18] M. Giambiagi, M. S. Giambiagi, D. R. Grempel and C. D. J. Heymann, *J. Chim. Phys. Phys.-Chim. Biol.* **72**, 15 (1975).

[19] I. Mayer, *Chem. Phys. Lett.* **97**, 270 (1983).

[20] R. L. Fulton, *J. Phys. Chem.* **97**, 7516 (1993).

[21] J. G. Angyan, M. Loos, I. Mayer, *J. Phys. Chem.* **98**, 5244 (1994).

[22] X. Fradera, M. A. Austen and R. F. W. Bader, *J. Phys. Chem. A* **103**, 304 (1999).

[23] R. Ponec and D. L. Cooper, *J. Mol. Struct. (Theochem)* **727**, 133 (2005).

[24] I. Mayer, *J. Comput. Chem.* **28**, 204 (2007).

[25] A. D. Becke, *J. Chem. Phys.* **98**, 5648 (1993).

[26] C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B* **37**, 785 (1988).

[27] S. H. Vosko, L. Wilk, and M. Nusair, *Can. J. Phys.* **58**, 1200 (1980).

[28] P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.* **98**, 11623 (1994).

[29] D. E. Woon and T. H. Dunning, *J. Chem. Phys.* **98**, 1358 (1993).

[30] B. P. Prascher, D. E. Woon, K. A. Peterson, T. H. Dunning Jr and A. K. Wilson, *Theor. Chem. Acc.* DOI: 10.1007/s00214-010-0764-0 (2010)

[31] T. H. Dunning, *J. Chem. Phys.* **90**, 1007 (1989).

[32] R. A. Kendall, T. H. Dunning and R. J. Harrison, *J. Chem. Phys.* **96**, 6796 (1992).

(Paper V)

Exact ionization potentials from wavefunction asymptotics: The extended Koopmans' theorem, revisited

D. Vanfleteren, D. Van Neck, P. W. Ayers, R. C. Morrison and P. Bultinck

The Journal of Chemical Physics 130, 194104 (2009)

THE JOURNAL OF CHEMICAL PHYSICS **130**, 194104 (2009)

Exact ionization potentials from wavefunction asymptotics: The extended Koopmans' theorem, revisited

Diederik Vanfleteren,[1] Dimitri Van Neck,[1] Paul W. Ayers,[2,a)] Robert C. Morrison,[3] and Patrick Bultinck[4]
[1]*Center for Molecular Modeling, Ghent University, Proeftuinstraat 86, B-9000 Gent, Belgium*
[2]*Department of Chemistry, McMaster University, Hamilton, Ontario L8S 4M1, Canada*
[3]*Department of Chemistry, East Carolina University, Greenville, North Carolina 27858, USA*
[4]*Department of Inorganic and Physical Chemistry, Ghent University, Krijgslaan 281(S3), 9000 Gent, Belgium*

(Received 10 December 2008; accepted 3 April 2009; published online 18 May 2009)

A simple explanation is given for the exactness of the extended Koopmans' theorem, (EKT) for computing the removal energy of any many-electron system to the lowest-energy ground state ion of a given symmetry. In particular, by removing the electron from a "removal orbital" of appropriate symmetry that is concentrated in the asymptotic region, one obtains the exact ionization potential and the exact Dyson orbital for the corresponding state of the ion. It is argued that the EKT is not restricted to many-electron systems but holds for any finite many-body system, provided that the interaction vanishes for increasing interparticle distance. A necessary and sufficient condition for the validity of the EKT for any state (not just the lowest-energy states of a given symmetry) in terms of the third-order reduced density matrix is stated and derived. © *2009 American Institute of Physics*. [DOI: 10.1063/1.3130044]

I. EXTENDED KOOPMANS' THEOREM

As an electron in an N-electron system moves further and further away, the wavefunction of the system begins to resemble an $N-1$ electron system, polarized by the distant electron. In the asymptotic limit,

$$\lim_{|\mathbf{r}_N|\to\infty} \Psi^{(N)}(\mathbf{r}_1,s_1;\mathbf{r}_2,s_2;\ldots;\mathbf{r}_N,s_N) \propto \Psi_s^{(N-1)}(\mathbf{r}_1,s_1;\mathbf{r}_2,s_2;\ldots;\mathbf{r}_{N-1},s_{N-1}), \tag{1}$$

where $\Psi_s^{(N-1)}$ is the lowest-energy state of the $N-1$ electron system that is consistent with the spin-and spatial-symmetry states of the original N-electron state ($\Psi^{(N)}$) and the electron that is removed. Thus any eigenfunction of an N-electron system contains within it *exact* information about the low-energy eigenfunctions of the $N-1$ electron system. The extended Koopmans' theorem (EKT) is a practical numerical procedure for extracting this exact information.[1–3] The EKT often provides reasonably accurate approximations for some of the higher-energy states of the $N-1$ electron system also.[4–11]

The EKT is derived from the orbital representation of the electron removal process in Eq. (1),[1–3]

$$\Psi_k^{(N-1)}(\mathbf{z}_1;\ldots;\mathbf{z}_{N-1}) = \int \phi_k(\mathbf{z}_N)\Psi^{(N)}(\mathbf{z}_1;\ldots;\mathbf{z}_N)d\mathbf{z}_N, \tag{2}$$

where $\mathbf{z}=(\mathbf{r},s)$ denotes both the spatial and spin coordinates of the electrons. The EKT removal orbital, ϕ_k, is determined by minimizing the energy of the $N-1$ electron state or, equivalently, optimizing the ionization energy

$$\mathcal{I}_k = E_k^{(N-1)} - E_0^{(N)} = \frac{\langle \int \phi_k(\mathbf{z}_N)\Psi^{(N)}(\mathbf{z}_1,\ldots,\mathbf{z}_N)d\mathbf{z}_N | \int \phi_k(\mathbf{z}_N)(\hat{H}^{(N-1)} - \hat{H}^{(N)})\Psi^{(N)}(\mathbf{z}_1,\ldots,\mathbf{z}_N)d\mathbf{z}_N\rangle_{1,\ldots,N-1}}{\langle \int \phi_k(\mathbf{z}_N)\Psi^{(N)}(\mathbf{z}_1,\ldots,\mathbf{z}_N)d\mathbf{z}_N | \int \phi_k(\mathbf{z}_N)\Psi^{(N)}(\mathbf{z}_1,\ldots,\mathbf{z}_N)d\mathbf{z}_N\rangle_{1,\ldots,N-1}}. \tag{3}$$

This gives the EKT equation,

$$\int \phi_k(\mathbf{z}_N)\left(-\frac{\nabla_N^2}{2} + v_{\text{ext}}(\mathbf{z}_N) + \sum_{i=1}^{N-1}\frac{1}{|\mathbf{r}_i-\mathbf{r}_N|}\right) \times \Psi^{(N)}(\mathbf{z}_1,\ldots,\mathbf{z}_N)d\mathbf{z}_N = -\varepsilon_k \int \phi_k(\mathbf{z}_N)\Psi^{(N)}(\mathbf{z}_1,\ldots,\mathbf{z}_N)d\mathbf{z}_N. \tag{4}$$

The orthonormality of the $N-1$ electron wavefunctions defined in Eq. (2) implies that the EKT removal orbitals are orthonormal in the metric of the one-electron reduced density matrix,

$$\int\int \phi_k(\mathbf{z})\gamma(\mathbf{z},\mathbf{z}')\phi_l(\mathbf{z}')d\mathbf{z}d\mathbf{z}' = \delta_{kl}. \tag{5}$$

The EKT equation is more commonly written in terms of density matrices, either in the spatial representation,[3,12]

a)Electronic mail: ayers@mcmaster.ca.

$$\int\int \phi_k(\mathbf{z}_1')\left(-\frac{\nabla_1^2}{2}+v_{\text{ext}}(\mathbf{z}_1)+\frac{(N-1)}{|\mathbf{r}_1-\mathbf{r}_2|}\right)\times\Gamma_2^{(N)}(\mathbf{z}_1,\mathbf{z}_2;\mathbf{z}_1',\mathbf{z}_2)d\mathbf{z}_2d\mathbf{z}_1' = -\varepsilon_k\int\int \phi_k(\mathbf{z}_1')\Gamma_2^{(N)}(\mathbf{z}_1,\mathbf{z}_2;\mathbf{z}_1',\mathbf{z}_2)d\mathbf{z}_2d\mathbf{z}_1', \tag{6}$$

$$\int \phi_k(\mathbf{z}_1')\left(-\frac{\nabla_1^2}{2}+v_{\text{ext}}(\mathbf{z}_1)\right)\Gamma_1^{(N)}(\mathbf{z}_1;\mathbf{z}_1')d\mathbf{z}_1' + 2\int\int \phi_k(\mathbf{z}_1')\left(\frac{1}{|\mathbf{r}_1-\mathbf{r}_2|}\right)\Gamma_2^{(N)}(\mathbf{z}_1,\mathbf{z}_2;\mathbf{z}_1',\mathbf{z}_2)d\mathbf{z}_2d\mathbf{z}_1' = -\varepsilon_k\int \phi_k(\mathbf{z}_1')\Gamma_1^{(N)}(\mathbf{z}_1;\mathbf{z}_1')d\mathbf{z}_1', \tag{7}$$

or the orbital representation,[1,2,9]

$$\sum_\lambda \langle\Psi^{(N)}|a_\lambda^+[a_\kappa,\hat{H}]|\Psi^{(N)}\rangle c_{k\lambda} = -\varepsilon_k\sum_\lambda \langle\Psi^{(N)}|a_\lambda^+a_\kappa|\Psi^{(N)}\rangle c_{k\lambda}, \tag{8}$$

$$\sum_\lambda\left(\sum_\mu h_{\kappa\mu}\gamma_{\lambda\mu}+\sum_{\mu\nu\sigma}V_{\kappa\mu\nu\sigma}\Gamma_{\lambda\mu\nu\sigma}\right)c_{k\lambda} = \varepsilon_k\sum_\lambda \gamma_{\kappa\lambda}c_{k\lambda}, \tag{9}$$

where $|\phi_k\rangle=\Sigma_\lambda \mathbf{c}_{k\lambda}a_\lambda^+|0\rangle$ is the expression for the EKT removal orbital, and the reduced density matrices are defined as

$$\gamma_{\lambda\mu} = \langle\Psi^{(N)}|a_\mu^+a_\lambda|\Psi^{(N)}\rangle, \tag{10}$$

$$\Gamma_{\lambda\mu\nu\sigma} = \langle\Psi^{(N)}|a_\nu^+a_\sigma^+a_\mu a_\lambda|\Psi^{(N)}\rangle. \tag{11}$$

Finally, the EKT can be expressed in terms of quantities related to electron propagator theory as

$$\int M_1^{(-)}(\mathbf{z},\mathbf{z}')\phi_k(\mathbf{z}')d\mathbf{z}' = \varepsilon_k\int M_0^{(-)}(\mathbf{z},\mathbf{z}')\phi_k(\mathbf{z}')d\mathbf{z}', \tag{12}$$

where

$$M_n^{(-)}(\mathbf{z},\mathbf{z}') = \frac{1}{2\pi i}\int E^n G(\mathbf{z},\mathbf{z}',E)e^{i\eta E}dE = \sum_k (-\mathcal{I}_k)^n g_k(\mathbf{z})g_k(\mathbf{z}') \tag{13}$$

are the energy moments of the electron propagator that are associated with electron removal. The kth Dyson orbital for electron removal is denoted $g_k=\langle\Psi_k^{(N-1)}|\Psi^{(N)}\rangle_{1,\ldots,N-1}$ and the associated exact removal energy is denoted $\mathcal{I}_k=E_k^{(N-1)}-E_0^{(N)}$. The zeroth energy moment is just the one-electron reduced density matrix, $M_0^{(-)}=\gamma$.

In all of these expressions, the quantity

$$-\varepsilon_k \approx \mathcal{I}_k \tag{14}$$

gives an approximation to the electron removal energy. If one uses the Hartree–Fock wavefunction/density matrices/electron propagator in these quantities, then ε_k is identical to the Hartree–Fock orbital energy and Eq. (14) is just the usual statement of Koopmans' theorem. So Eqs. (4)–(12) represent an extension of Koopmans' theorem to correlated electronic structure theory calculations.

The EKT is an important tool in electronic structure theory because it provides a simple and computationally efficient way to extract information about the successive ionization potentials (IPs) of a system from correlated electronic structure methods. Together, the (generally approximate) electron removal energies, ε_k, and their associated approximate Dyson orbitals,

$$g_k^{(\text{EKT})}(\mathbf{z}) = \int \gamma(\mathbf{z},\mathbf{z}')\phi_k(\mathbf{z}')d\mathbf{z}' \tag{15}$$

provide "correlated" analogs that have similar conceptual utility to the single-particle orbitals from independent particle models such as Hartree–Fock and Kohn–Sham density functional theory (DFT).[13,14] The analogous orbitals and energies for electron attachment can be defined [just interchange the creation and annihilation operators in Eq. (8)],[2] but, like the corresponding quantities in Hartree–Fock theory, they are usually less accurate than the corresponding ionization-related quantities.

The EKT has been used to extract IPs from many different types of correlated calculations. The most straightforward implementation uses the result of CASSCF/MCSCF/full-configuration interaction (CI) calculations (using GAMESS-US);[9] it is also possible to use the EKT to obtain IPs from Møller–Plesset perturbation theory and QCISD calculations [using GAUSSIAN IOp(6/51)].[15] The EKT has been used to test approximate self-energies in electron propagator approaches.[16] IPs in density-matrix functional theory can be computed with EKT because the two-electron matrix elements needed to evaluate Eq. (9) are approximated in terms of the one-particle reduced density matrix in such approaches.[17–20] It is clear that the EKT could provide similar information for computational approaches using higher-order reduced density matrices also.[21–29] EKT has also been explored as a method for computing the change in electron density from electron removal/attachment (the so-called Fukui function[30–32]) and other reactivity indicators in DFT-based chemical reactivity theory.[33] There are "generalized" EKT that use higher-order density matrices to obtain higher-order IPs, accurate electron attachment energies, and excitation energies.[33]

II. THE EXACTNESS OF THE EKT

The EKT is *exact* for the lowest-energy electron removal energy, ε_0. This remarkable statement has been the subject of much doubt and debate.[7,8,11,12,34–38] For example, it seems remarkable that the lowest ionization energy of any system could be determined *exactly* from using Eq. (12) since $M_1^{(-)}$ is the operator representing the *average* electron removal energy. Indeed, if one compares the EKT result to results from perturbation theory,[34,35] one observes that second-order terms are missing; it is just that these terms seem to be zero for the lowest-energy electron removal process.[36]

The usual argument for the exactness of the EKT uses the fact that the first-order density matrix can be expressed using approximate Dyson orbitals from the EKT,

$$\gamma(\mathbf{z},\mathbf{z}') = \sum_k (g_k^{(\mathrm{EKT})}(\mathbf{z}'))^* g_k^{(\mathrm{EKT})}(\mathbf{z}). \quad (16)$$

Moreover, by analogy to the argument for the asymptotic decay of the Hartree–Fock orbitals,[39] Morrell, *et al.* were able to show that the $r\to\infty$ asymptotic decay of the EKT Dyson orbitals is governed by the smallest EKT IP,

$$g_k^{(\mathrm{EKT})}(\mathbf{z}) \sim e^{-r\sqrt{-2\varepsilon_0}}. \quad (17)$$

Since the $r\to\infty$ asymptotic decay of the density matrix is[40–43]

$$\gamma(\mathbf{z},\mathbf{z}') \sim e^{-r\sqrt{2\mathcal{I}_0}}, \quad (18)$$

it follows that $\varepsilon_o=-\mathcal{I}_0$. The EKT is thus exact for the first IP. It should be noted that the derivation of Eq. (18) uses the Coulomb form of interaction between the particles. Whether the EKT works for non-Coulombic interactions is an open problem; we will show that the EKT is very general and gives a non-Coulombic example where it is exact.

While the asymptotic argument for the exactness of the EKT is convincing by the standards of mathematical rigor that are typical in theoretical chemical physics, it may be criticized. In particular, it is certainly possible to expand one function (say, e^{-2r}) in terms of an infinite number of functions which have a different asymptotic decay [e.g., the orthonormal set $(4\pi)^{-1/2}\, L_n^2(r)e^{-r/2}$, where $L_n^2(r)$ are associated Laguerre polynomials]. The fact that the asymptotic decay of the density matrix is given by Eq. (18) thus strongly suggests—but does not rigorously prove—that the functions used in the expansion, $g_k^{(\mathrm{EKT})}$, also decay in the same manner.

Our assertion is that the EKT holds precisely because of Eq. (1). That is, if the EKT removal orbital, $\phi_k(\mathbf{z})$, is concentrated sufficiently far away from the molecule, then the $N-1$ electron wavefunction that results from Eq. (2) will be accurate. Moreover, the further ϕ_k moves from the molecule, the more accurate $\Psi_k^{(N-1)}$ becomes. This also implies that the optimal EKT removal orbital should be interpreted as a generalized function (a limit of a sequence of functions): there is no well-defined physical orbital that solves Eq. (4) in the basis set limit.

To make this argument mathematically precise, consider a trial EKT removal orbital with the specific form $\tilde{\phi}(\mathbf{z}_N) = (\eta/\pi)^{3/2} \exp(-\eta|\mathbf{r}_N-\mathbf{r}_0|^2)\sigma(N)$, where $\sigma(N)$ is the spin of the electron being removed. Since we are restricting the form of the trial orbital and the EKT can be derived from the variational principle for the $N-1$ electron wavefunction [cf. Eq. (3)], employing $\tilde{\phi}(\mathbf{z}_N)$ gives an upper bound to the true ionization energy. If we take the limit as $\eta\to\infty$ and $|\mathbf{r}_0|\to\infty$, the trial orbital becomes a delta function at infinite distance from the system. In this limit the $N-1$ electron wavefunction used in the EKT [cf. Eq. (2)] is the exact $N-1$ electron wavefunction obtained by moving an $\sigma(N)$-spin electron to infinity [cf. Eq (1)]. The EKT is exact for the lowest IP of each spin.

Taking the $\eta\to\infty$ limit and completely localizing the electron removal orbital makes the connection to Eq. (1) clear, but this is not required for the validity of the EKT. Consider the Dyson orbital expansion of the N-electron wavefunction,

$$\Psi_0^{(N)}(\mathbf{z}_1,\ldots,\mathbf{z}_N) = \sum_{k=0}^{\infty} g_k(\mathbf{z}_N)\Psi_k^{(N-1)}(\mathbf{z}_1,\mathbf{z}_2,\ldots,\mathbf{z}_{N-1}). \quad (19)$$

For sufficiently large $|\mathbf{r}_0|$ and η, $\tilde{\phi}(\mathbf{z}_N)$ will be localized far from the system in question and will have appreciable overlap only with the asymptotically dominant Dyson orbital of the same spin. The EKT gives the exact lowest ionization energy if and only if the Dyson orbital corresponding to that ionization is asymptotically dominant. [If a different Dyson orbital is asymptotically dominant (e.g., for reasons of symmetry), then the EKT will give the exact IP to an excited state of the $N-1$ electron system.] For Coulomb systems, Katriel and Davidson showed that the Dyson orbital for ionization to the $N-1$ electron ground state decays faster than the other Dyson orbitals by at least a factor of $|\mathbf{r}_N|^{-2}$. Since the Dyson orbitals for the lowest α-spin IPs are asymptotically dominant, the EKT gives the exact removal energy and Dyson orbital for the lowest-energy ionizations of α-spin and β-spin electrons in Coulomb systems.

Our new derivation of the EKT indicates that the EKT removal orbital is concentrated far away from the system. To confirm this hypothesis numerically, we performed a series of full-CI calculations on the beryllium atom using the six s-type functions from the cc-pV5Z basis and an additional spherically-symmetric radial Gaussian function, $\chi(\mathbf{r}) \sim e^{-5(r-r_o)^2}$. This extra function is a spherical shell that peaks r_0 atomic units from the beryllium nucleus. The value of r_0 is successively taken to be 1.0, 7.0, 9.0, or 11.0 a.u., yielding radial wavefunctions concentrated at increasing distances from the nucleus. Note that the Coulomb integrals involving the added spherical shell orbital were calculated using numerical quadrature. The EKT removal orbital that corresponds to the lowest IP is expected to be concentrated as far from the atom as possible. As Fig. 1 indeed shows, for the more distant spherical shells, the EKT removal orbital is nearly proportional to the added radial shell function. We infer that as the basis set approaches completeness, the EKT removal orbital becomes concentrated infinitely far away from the atom or molecule. The cc-pV5Z basis is already sufficiently diffuse to give good results for the IPs and the computed lowest IP is almost the same as the full-CI value ($\mathrm{IP}_{\mathrm{EKT},r_0=1}$=0.299 00 a.u.; $\mathrm{IP}_{\mathrm{full\text{-}CI}}$=0.298 98 a.u.) even for the nearest spherical shell (r_0=1 a.u.). For spherical shells concentrated in the outer regions of the atom, $r_0 \geqslant 7$ a.u., the lowest EKT IP agrees with the full-CI value to the limits of our numerical precision.

Further confirmation of our interpretation was obtained by revisiting the accurate calculations on the beryllium atom from Ref. 10. We started with the 6-311G basis for Be (located at the origin) and then extended this basis set by adding 12 ghost atoms at positions $\mathbf{r}_0$, distributed symmetrically along the z-axis at $z_0=\pm 1.0, \pm 2.0, \ldots, \pm 6.0$ Å. A single primitive Gaussian wavefunction $e^{-|\mathbf{r}-\mathbf{r}_0|^2}$ is centered on each

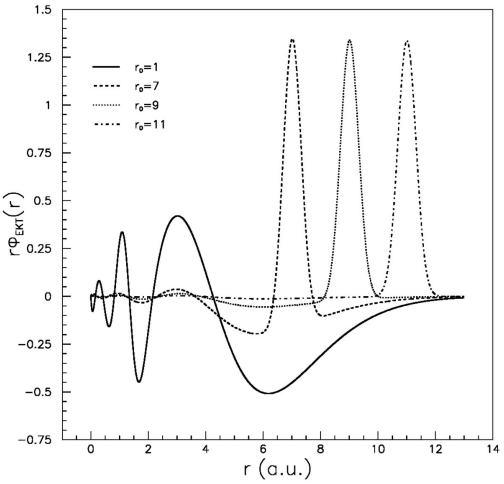


FIG. 1. The radial wavefunction of the first EKT removal orbital for the Be atom, obtained with a basis consisting of the 6s-orbitals of the cc-pV5Z basis set and an extra orbital as explained in the text. A.u. are used. The extra orbital is located at $r_0=1,7,9,11$ a.u.

ghost atom. The EKT orbital was then computed from an MCSCF calculation using the first 18 of the 25 orbitals computed from this basis. The generalized eigenvalue problem in Eq. (9) is ill-conditioned when the first-order density matrix is nearly singular, so we eliminated a natural orbital with occupation $\sim 10^{-12}$ from our the calculations of the EKT orbital. The resulting EKT orbital is plotted in Fig. 2; notice that the orbital density is concentrated on the distant ghost atoms. The orbital density is located mainly on the ghost atoms at 4.0 and 5.0 Å instead of the most distant ghost atom (at 6.0 Å) because the natural orbital we eliminated from the calculation because of its tiny occupation number is located mostly on the furthest ghost atoms.

The asymptotic arguments employed by Morrell *et al.*[41] rely on the Coulombic r_{ij}^{-1} interactions between electrons.[3] Based on the observed differences between the expressions for electron removal energies from the EKT and from order-by-order expansion, it was speculated that perhaps the EKT is only exact for the lowest electron removal energy in Coulombic systems. However, the argument formulated in the

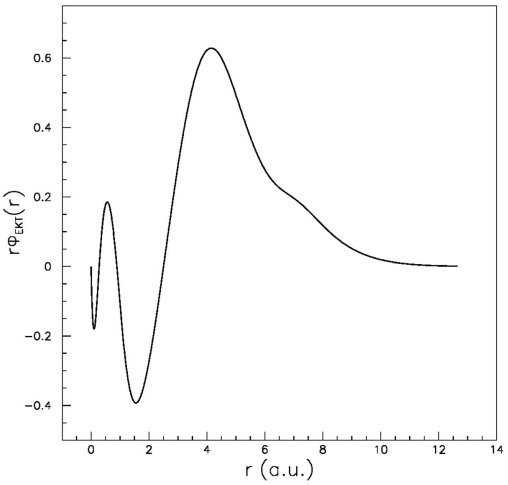


FIG. 2. The wavefunction of the first EKT removal orbital on the z-axis for the Be atom in the presence of 12 ghost atoms located on the z-axis $\pm 1, \pm 2, \ldots, \pm 6$ Å from the Be nucleus. The Be atom is described with the 6-311G basis; each ghost atom is described by a single primitive s-type Gaussian. A.u. are used.

previous paragraphs relies only on the validity of Eq. (1), which is true for any system, including systems with short-range interparticle potential. The only proviso is that the interaction vanishes with increasing interparticle distance. To ensure that the EKT is indeed true for any system, we considered a self-bound system of bosons interacting via delta-function attractive potentials in one dimension with the Hamiltonian

$$\hat{H} = \frac{-\hbar^2}{2m}\sum_{i=1}^{N}\frac{\partial^2}{\partial x_i^2} - g\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\delta(x_i - x_j). \tag{20}$$

This model is exactly solvable for arbitrary particle number.[44,45] For all $N \geqslant 2$, this system has only one bound state with energy $E_0^{(N)} = -(12)^{-1}\lambda g N(N^2-1)$, where $\lambda = mg/(2\hbar^2)$.

As in the previous section, we focused our study on the four-particle system, $N=4$, with IP $\mathcal{I}_0 = 3\lambda g$. In this case, the one-particle reduced density matrix ($\gamma = M_0^{(-)}$) and the removal energy matrix ($M_1^{(-)}$) have the expressions

$$\begin{aligned} M_0^{(-)}(x,x') = \frac{3\lambda}{5}(&30e^{-3s} - 15e^{-6s+3v} - 9e^{-6s-5v} + 3e^{-9s+6v} \\ &+ 2e^{-9s-4v} + e^{-9s-10v} + \Theta(v-s)(-3e^{-9s+6v} \\ &+ 3e^{6s-9v} + 15e^{-6s+3v} - 15e^{3s-6v} - 30e^{-3s} \\ &+ 30e^{-3v}) + \Theta(v-3x)(10e^{-7v} - 15e^{-3s-6v} \\ &+ 9e^{-6s-5v} - 3e^{6s-9v} - 2e^{-9s-4v} + e^{9s-10v})), \end{aligned} \tag{21}$$

$$\begin{aligned} M_1^{(-)}(x,x') = \frac{\lambda^2 g}{5}(&270e^{-3s} - 135e^{-6s+3v} - 9e^{-6s-5v} \\ &+ 27e^{-9s+6v} + 18e^{-9s-4v} + e^{-9s-10v} + \Theta(v-s) \\ &\times(-27e^{-9s+6v} + 27e^{6s-9v} + 135e^{-6s+3v} \\ &- 135e^{3s-6v} - 270e^{-3s} + 270e^{-3v}) \\ &+ \Theta(v-3x)(130e^{-7v} - 135e^{-3s-6v} + 9e^{-6s-5v} \\ &- 27e^{6s-9v} - 22e^{-9s-4v} + e^{9s-10v})), \end{aligned} \tag{22}$$

where $v = \lambda|x-x'|$, $s = \lambda|x+x'|$, and $\Theta(x)$ is the Heaviside step function. The exact Dyson orbital is

$$g_0(x) = \sqrt{2\lambda}\left(3e^{-3\lambda|x|} - \tfrac{3}{2}e^{-9\lambda|x|} + \tfrac{3}{10}e^{-15\lambda|x|}\right). \tag{23}$$

In the $x \to +\infty$ asymptotic limits, these quantities have the asymptotic forms

$$M_0^{(-)}(x,x') \sim g_0(x')3\sqrt{2\lambda}e^{-3\lambda x}, \tag{24}$$

$$M_1^{(-)}(x,x') \sim \mathcal{I}_0(g_0(x')3\sqrt{2\lambda}e^{-3\lambda x}), \tag{25}$$

$$g_0(x) \sim 3\sqrt{2\lambda}e^{-3\lambda x}. \tag{26}$$

For this self-bound system, the characteristic exponential rate of asymptotic decay, $\alpha = 3\lambda$, is related to the IP as $\alpha = \sqrt{2\mu\mathcal{I}_0}$, where $\mu = mN/(N-1)$ is the reduced mass for the motion of one particle (far away from the others) relative to the center of mass of the other $N-1$ particles.

We substituted Eqs. (21) and (22) into Eq. (12) and solved for the EKT removal orbital and its eigenvalue on a grid of evenly spaced points. The equation to be solved is very ill-conditioned, and it is necessary to eliminate all of the natural orbitals with small eigenvalues. Once this is done, the equation can be solved, and we observe that the EKT removal orbital, $\phi_0(x)$ is concentrated entirely at the edge of the x-interval covered by the grid. In accord with our expectations, the EKT removal energy, ε_0, is correct, and the EKT Dyson orbital, $g_0^{(\mathrm{EKT})}(x)$, matches the analytic result from Eq. (26).

III. IONIZATION TO OTHER SYMMETRY STATES

It has been repeatedly noted that for atoms the EKT seems to be accurate not only for the lowest IP but also for the lowest IP to certain other symmetry states.[5,7,10,11] We now explain this finding. By removing an electron with a specified spin and angular momentum (i.e., a specific m_s and m_ℓ), one can often access excited states of the ion. For example, consider the N-electron wavefunction with $M_S^{(N)} = S^{(N)}$. Removing a majority-spin electron will produce a state of the ionized system with $M_S^{(N-1)} = S^{(N)} - \frac{1}{2}$; removing a minority-spin electron will produce a state of the ion with $M_S^{(N-1)} = S^{(N)} + \frac{1}{2}$. If the ground state of the ionized system has smaller spin multiplicity than the ground state of the original molecule, then one should obtain *two* IPs exactly: one for electron removal from the spin-majority channel and one for electron removal from the spin-minority channel. In cases where the spin multiplicity of the ionized system is very different from that of the original molecule, the EKT may fail to give the lowest IP and instead give exact results only for higher-energy excited states of the cation. These results are similar to the situation in spin DFT, where the eigenvalues of the highest-occupied orbitals of both α-spin and β-spin equal -1 times the associated electron removal energies.[46–49]

As an example of this procedure, we studied the electron removal from the Be^+ ^2S ground state using the basis set of six s-type functions described in connection with Fig. 1 and an extra spherical shell at $r_0 = 11$ a.u. The EKT ionization energies are reported and compared to the energy difference computed from full CI in Table I. We consider the $M_S^{(3)} = 1/2$ state of the Be^+ doublet, so α-spin electrons are in the majority. The EKT results for removing an α-spin electron from the Be^+ are obtained by restricting the indices κ and λ in Eq. (9) to α-spin orbitals. This accesses states with $M_S^{(2)} = 0$, so that the EKT provides approximate IPs to both the singlet and triplet states in Be^{+2}.

If one removes an α-spin electron, then it is clear from Table I that only the lowest singlet state is described adequately, whereas the triplet IP deviates from the exact result by 0.059 a.u. This can be understood from previous consid-

TABLE I. IPs in a.u. for the ^2S ground state of Be^+. Removing a majority-spin α-electron produces a state with $M_s^{(2)}=0$, which can give either a singlet state (first column) or triplet state (second column) of Be^{+2}. Removing a β-spin electron produces only the triplet state. Exact results from the energy difference of full-CI calculations are reported in the rows labeled FCI. Results for removing α-spin and β-spin electrons, using the EKT, are labeled EKT_α and EKT_β, respectively. The absolute deviation from the exact IPs is in parentheses.

		Ionization potential	
Basis size	Calculation	$E_0^{(2)}(^1S)-E_0^{(3)}(^2S)$	$E_0^{(2)}(^3S)-E_0^{(3)}(^2S)$
7	FCI	0.665 87	4.997 61
	EKT_α	0.665 87 (2.3×10^{-10})	5.056 77 (5.9×10^{-2})
	EKT_β	⋯	4.999 38 (1.8×10^{-3})
16	FCI	0.665 92	4.997 15
	EKT_α	0.665 92 (2.2×10^{-7})	5.039 63 (4.2×10^{-2})
	EKT_β	⋯	4.997 14 (8.2×10^{-6})

erations: the Dyson orbital for the transition to the singlet ground state is asymptotically dominant, and exact EKT results are expected only for the singlet not for the triplet.

Alternatively, one can remove a β-spin electron from Be^+; in this case, one accesses states with $M_S^{(2)}=1$ and the EKT provides approximate IPs for the triplet state only. The Dyson orbital for the transition to the lowest-energy triplet state should now be dominant as there is no longer competition from the transition to the singlet state. One expects the EKT to be exact for the lowest triplet state and, indeed, the EKT result for the β-spin IP is much more accurate (by about a factor of 30). However, the computed IP_β is still much less accurate than the prediction for the singlet. We ascribe this to finite basis set effects: the Dyson orbital for the transition to the singlet has primarily $2s$ character, which the valence pentuple-zeta basis (cc-pV5Z) describes very well. The Dyson orbital for the transition to the triplet, however, has primarily $1s$ character. The core basis functions in the cc-pV5Z basis have a relatively little flexibility and are ill-suited for describing the core relaxation effects that accompany ionization to the triplet state. To enhance the description of the Be core, we extended the basis set by decontracting the cc-pV5Z $1s$-type orbitals. Specifically, we kept the two linear combinations of the five most rapidly decaying Gaussians (which describe the electron-nuclear cusp that is present in every system), while adding as Gaussian primitives the remaining exponentials. We supplemented this set by adding four more Gaussian primitives, with exponents (.5, 1.5, 3.0, and 5.0) chosen to fill the gaps in the cc-pV5Z sampling of Gaussian exponential parameters. The results of the EKT calculation in this basis set of 15 s-type Gaussians, plus a spherical shell at 11.0 a.u., are reported in Fig. 1. Notice that the EKT IP to the triplet $M_S^{(2)}=1$ state is two orders of magnitude more accurate now, with accuracy comparable to the EKT IP to the singlet state. This confirms our mathematical argument that, in the basis set limit, the EKT gives exact results for both α-spin and β-spin ionizations.

The case of angular momentum symmetry is similar. By choosing the EKT removal orbital with different values of m_ℓ, one can obtain exact IPs to excited states of the ion that have higher-angular momentum than the ground state.

IV. NECESSARY AND SUFFICIENT CONDITIONS FOR EXACTNESS

When is the EKT exact not only for the removal energy to the lowest state of a given symmetry, but also to certain other, higher-order, excited states? There does not seem to be any easy criterion for this, though it is known that the EKT is accurate (though perhaps not exact) even for ionization to highly excited states of the cation [e.g., the $1s^22s^2\rightarrow1s^12s^2$ ionization of Be (Ref. 10)]. Pernal and Cioslowski gave a criterion for all of the EKT removal energies to be exact in terms eigenvalues, $\{\nu_k\}$, and eigenvectors, $\{\Phi_k\}$, of the $N-1$ particle reduced density matrix,

$$\Gamma_{N-1}^{(N)}=\sum_k \nu_k\Phi_k(z_1'\ldots z_{N-1}')\Phi_k(\mathbf{z}_1\ldots\mathbf{z}_N). \tag{27}$$

Specifically, all of the EKT removal energies are exact if for every j such that $\nu_j>0$ and k such that $\nu_k=0$, the energy matrix element between Φ_j and Φ_k vanishes: $\langle\Phi_j|\hat{H}^{(N-1)}|\Phi_k\rangle=0$.[12] This criterion is unlikely to be satisfied and is difficult to evaluate because it requires the $N-1$ electron reduced density matrix. It is worth recalling that whenever $\nu_j>0$, it is equal to one of the natural orbital occupation numbers.[50]

It seems more likely that *some* of the EKT removal energies are exact, while others are approximate. In that case, the Pernal–Cioslowski condition will be violated. A simple test that is *necessary and sufficient* for the exactness of any specific EKT removal energy and the associated Dyson orbital is obtained by noting that the EKT approximation to the $N-1$ electron wavefunction, Eq. (2), is exact if and only if

$$\langle\Psi_k^{(N-1)}|(\hat{H}^{(N-1)}-E_k^{(N-1)})^2|\Psi_k^{(N-1)}\rangle=0. \tag{28}$$

Equivalently,

$$\langle(\hat{H}^{(N-1)}-(E^{(N)}-\varepsilon_k))\Psi_k^{(N-1)}| \times(\hat{H}^{(N-1)}-(E^{(N)}-\varepsilon_k))\Psi_k^{(N-1)}\rangle=0. \tag{29}$$

As shown in the Appendix, Eq. (29) can be simplified using Eq. (2), giving a necessary and sufficient condition for the exactness of the EKT in terms of the two-and three-particle reduced density matrices,

$$0 = \int \cdots \int d\mathbf{z}_1 d\mathbf{z}_1' d\mathbf{z}_2 d\mathbf{z}_3 \Bigg\{ \phi_k(\mathbf{z}_1')\phi_k(\mathbf{z}_1) \Bigg[\left(-\frac{1}{2}\nabla_{1'}^2 + v_{\text{ext}}(\mathbf{r}_1') + (N-1)|\mathbf{r}_2 - \mathbf{r}_1'|^{-1} - \varepsilon_k \right) \times \left(-\frac{1}{2}\nabla_1^2 + v_{\text{ext}}(\mathbf{r}_1) + (N-1)|\mathbf{r}_2 - \mathbf{r}_1|^{-1} - \varepsilon_k \right) + (N-2) \left(-\frac{1}{2}\nabla_{1'}^2 + v_{\text{ext}}(\mathbf{r}_1') + (N-1)|\mathbf{r}_2 - \mathbf{r}_1'|^{-1} - \varepsilon_k \right) \times \left(-\frac{1}{2}\nabla_1^2 + v_{\text{ext}}(\mathbf{r}_1) + (N-1)|\mathbf{r}_3 - \mathbf{r}_1|^{-1} - \varepsilon_k \right) \Bigg] \times \Gamma_3(\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3; \mathbf{z}_1', \mathbf{z}_2, \mathbf{z}_3) \Bigg\}. \tag{30}$$

This necessary and sufficient condition can be stated, equivalently, in the orbital representation as

$$0 = \sum_{\lambda\lambda'} c_{k\lambda} c_{k\lambda'} \sum_{\substack{\mu\nu\sigma \\ \mu'\nu'\sigma'}} \Gamma_{\mu\nu\sigma\mu'\nu'\sigma'} \times \left(\sum_{\kappa} (w_{\lambda\kappa\nu'\sigma'} w_{\nu\sigma\lambda'\kappa} \delta_{\mu\mu'} - w_{\lambda\mu'\nu'\sigma'} w_{\nu\sigma\lambda'\mu}) \right), \tag{31}$$

where

$$w_{\mu\nu\mu'\nu'} = \frac{\delta_{\nu\nu'}}{N-1}(h_{\mu\mu'} - \varepsilon_k \delta_{\mu\mu'}) + V_{\mu\nu\mu'\nu'} \tag{32}$$

This provides a simple criterion for whether the higher EKT removal energies are exact and, if they are not exact, a measure of their accuracy. Because it requires the three-particle reduced density matrix, it is much harder to evaluate than the EKT (although it is much, much easier than the Pernal–Cioslowski condition based on the $N-1$ electron density matrix).

V. SUMMARY

The EKT is a valuable tool for computing ionization energies. If the exact N-electron wavefunction is used to formulate the EKT equations [cf. Eqs. (4)–(12)], then the first EKT removal energy, ε_0, and the lowest EKT Dyson orbital, $g_0^{(\text{EKT})}(\mathbf{z})$, are exact. This result can be simply explained by noting that the wavefunction that remains after an electron is moved infinitely far from an N-electron system is precisely the lowest-energy wavefunction of the $N-1$ electron ion that can be attained given the spin and symmetry constraints inherited from the original N-electron system. The EKT works because in the infinite basis set (alternative, the infinite-grid) limit, the EKT removal orbital, $\phi_0(\mathbf{z})$, is located infinitely far from the molecule in question. Because Eq. (1) is completely general, the EKT is also completely general: the EKT is not limited to molecular systems, fermions, or particles interacting with long-range forces. The same argument used in this paper could be used to establish the exactness of the generalized EKT for multiple ionization.[33]

The EKT is also exact for some other electron removal energies, $\varepsilon_k < \varepsilon_0$. In particular, the preceding asymptotic argument can be used to show that the removal energy to the lowest-energy state of a specified symmetry can sometimes be attained by selecting the appropriate spin and spatial symmetry for the EKT removal orbital. Higher-energy states of a given symmetry *might* also be exact, but we do not know of any asymptotic criterion. However, Eq. (30) provides a necessary and sufficient condition for the exactness of any EKT removal energy in terms of the three-particle reduced density matrix.

We conclude with some practical advice for EKT computations. Although one might naïvely expect that diffuse functions are not very important for describing the wavefunction of atomic and molecular cations, it is very important to use diffuse basis sets when computing IPs and Dyson orbitals with EKT. It is also helpful to add a cage of ghost atoms far from the system (but not too far, if the natural orbital occupations corresponding to these basis functions are too small then the EKT equations become numerically unstable), as various linear combinations of the basis functions on these ghost atoms can be used to construct EKT removal orbitals of various symmetries. To obtain accurate results for higher IPs, it is important to choose a basis that can describe changes in the core orbitals.

After this article was accepted for publication, we became aware of an article in press, by Matthias Ernzerhof, revisiting the EKT.[51] The conclusion in Ernzerhof's paper is similar to ours: the EKT is arbitrarily accurate for the ground state and the EKT removal orbitals are "generalized functions" that are concentrated asymptotically.

ACKNOWLEDGMENTS

P.W.A. acknowledges support from the Canada Research Chairs, the Alfred P. Sloan foundation, and NSERC. D.V.N. acknowledges support from the research council (BOF) of Ghent University. P.B. acknowledges support from FWO-Flanders.

APPENDIX: DERIVATION OF EQUATION (30)

This appendix derives Eq. (30) from Eq. (29). Substituting Eq. (2) into Eq. (29) gives

$$0 = \Bigg\langle (\hat{H}^{(N-1)} - (E^{(N)} - \varepsilon_k)) \int \phi_k(\mathbf{z}_N) \Psi^{(N)}(\mathbf{z}_1, \ldots, \mathbf{z}_N) d\mathbf{z}_N \times \Big| (\hat{H}^{(N-1)} - (E^{(N)} - \varepsilon_k)) \times \int \phi_k(\mathbf{z}_N) \Psi^{(N)}(\mathbf{z}_1, \ldots, \mathbf{z}_N) d\mathbf{z}_N \Bigg\rangle. \tag{A1}$$

Because the $N-1$ electron Hamiltonian is a linear operator and does not depend on the coordinates of the Nth electron,

the operator can be brought inside the integral

$$0=\left\langle \int \phi_k(\mathbf{z}_N)(\hat{H}^{(N-1)}-(E^{(N)}-\varepsilon_k))\Psi^{(N)}(\mathbf{z}_1,\ldots,\mathbf{z}_N)d\mathbf{z}_N \middle| \int \phi_k(\mathbf{z}_N)(\hat{H}^{(N-1)}-(E^{(N)}-\varepsilon_k))\Psi^{(N)}(\mathbf{z}_1,\ldots \mathbf{z}_N)d\mathbf{z}_N \right\rangle. \tag{A2}$$

Adding and subtracting $\hat{H}^{(N)}$ and using the N-electron Schrödinger equation gives

$$\begin{aligned}0=&\left\langle \int \phi_k(\mathbf{z}_N)(\hat{H}^{(N)}+(\hat{H}^{(N-1)}-\hat{H}^{(N)})-(E^{(N)}-\varepsilon_k))\Psi^{(N)}(\mathbf{z}_1,\ldots,\mathbf{z}_N)d\mathbf{z}_N \right.\\ &\left.\times \left| \int \phi_k(\mathbf{z}_N)(\hat{H}^{(N)}+(\hat{H}^{(N-1)}-\hat{H}^{(N)})-(E^{(N)}-\varepsilon_k))\Psi^{(N)}(\mathbf{z}_1,\ldots,\mathbf{z}_N)d\mathbf{z}_N \right\rangle \right.\\ =&\left\langle \int \phi_k(\mathbf{z}_N)(\hat{H}^{(N-1)}-\hat{H}^{(N)}+\varepsilon_k)\Psi^{(N)}(\mathbf{z}_1,\ldots,\mathbf{z}_N)d\mathbf{z}_N \middle| \int \phi_k(\mathbf{z}_N)(\hat{H}^{(N-1)}-\hat{H}^{(N)}+\varepsilon_k)\Psi^{(N)}(\mathbf{z}_1,\ldots,\mathbf{z}_N)d\mathbf{z}_N \right\rangle.\end{aligned} \tag{A3}$$

Using the form of the Hamiltonian, we have

$$\begin{aligned}0=&\left\langle \int \phi_k(\mathbf{z}_N)\left(-\frac{\nabla_N^2}{2}+v_{\text{ext}}(\mathbf{z}_N)+\sum_{i=1}^{N-1}\frac{1}{|\mathbf{r}_i-\mathbf{r}_N|}-\varepsilon_k\right)\Psi^{(N)}(\mathbf{z}_1,\ldots,\mathbf{z}_N)d\mathbf{z}_N \right.\\ &\left.\times \left| \int \phi_k(\mathbf{z}_N)\left(-\frac{\nabla_N^2}{2}+v_{\text{ext}}(\mathbf{z}_N)+\sum_{j=1}^{N-1}\frac{1}{|\mathbf{r}-\mathbf{r}_N|}-\varepsilon_k\right)\Psi^{(N)}(\mathbf{z}_1,\ldots,\mathbf{z}_N)d\mathbf{z}_N \right\rangle \right.\end{aligned}. \tag{A4}$$

To write this equation in terms of the reduced density matrices, pull the summations out of the integrals so that the identical-particle nature of electrons can be exploited.

$$\begin{aligned}0=&\left\langle \sum_{i=1}^{N-1}\int \phi_k(\mathbf{z}_N)\left(-\frac{\nabla_N^2}{2(N-1)}+\frac{v_{\text{ext}}(\mathbf{z}_N)}{N-1}+\frac{1}{|\mathbf{r}_i-\mathbf{r}_N|}-\frac{\varepsilon_k}{N-1}\right)\Psi^{(N)}(\mathbf{z}_1,\ldots,\mathbf{z}_N)d\mathbf{z}_N \right.\\ &\left.\times \left| \sum_{j=1}^{N-1}\int \phi_k(\mathbf{z}_N)\left(-\frac{\nabla_N^2}{2(N-1)}+\frac{v_{\text{ext}}(\mathbf{z}_N)}{N-1}+\frac{1}{|\mathbf{r}_j-\mathbf{r}_N|}-\frac{\varepsilon_k}{N-1}\right)\Psi^{(N)}(\mathbf{z}_1,\ldots,\mathbf{z}_N)d\mathbf{z}_N \right\rangle \right.\\ =&\sum_{i=1}^{N-1}\left\langle \int \phi_k(\mathbf{z}_N)\left(-\frac{\nabla_N^2}{2(N-1)}+\frac{v_{\text{ext}}(\mathbf{z}_N)}{N-1}+\frac{1}{|\mathbf{r}_i-\mathbf{r}_N|}-\frac{\varepsilon_k}{N-1}\right)\Psi^{(N)}(\mathbf{z}_1,\ldots,\mathbf{z}_N)d\mathbf{z}_N \right.\\ &\left.\times \left| \int \phi_k(\mathbf{z}_N)\left(-\frac{\nabla_N^2}{2(N-1)}+\frac{v_{\text{ext}}(\mathbf{z}_N)}{N-1}+\frac{1}{|\mathbf{r}_i-\mathbf{r}_N|}-\frac{\varepsilon_k}{N-1}\right)\Psi^{(N)}(\mathbf{z}_1,\ldots,\mathbf{z}_N)d\mathbf{z}_N \right\rangle \right.\\ &+\sum_{i=1}^{N-1}\sum_{\substack{j=1\\ j\neq i}}^{N-1}\left\langle \int \phi_k(\mathbf{z}_N)\left(-\frac{\nabla_N^2}{2(N-1)}+\frac{v_{\text{ext}}(\mathbf{z}_N)}{N-1}+\frac{1}{|\mathbf{r}_i-\mathbf{r}_N|}-\frac{\varepsilon_k}{N-1}\right)\Psi^{(N)}(\mathbf{z}_1,\ldots,\mathbf{z}_N)d\mathbf{z}_N \right.\\ &\left.\times \left| \int \phi_k(\mathbf{z}_N)\left(-\frac{\nabla_N^2}{2(N-1)}+\frac{v_{\text{ext}}(\mathbf{z}_N)}{N-1}+\frac{1}{|\mathbf{r}_j-\mathbf{r}_N|}-\frac{\varepsilon_k}{N-1}\right)\Psi^{(N)}(\mathbf{z}_1,\ldots,\mathbf{z}_N)d\mathbf{z}_N \right\rangle \right.\end{aligned}. \tag{A5}$$

Using the fact electrons are identical particles, one has

$$\begin{aligned}0=&\frac{1}{N-1}\left\langle \int \phi_k(\mathbf{z}_N)\left(\frac{\nabla_N^2}{2}+v(\mathbf{z}_N)+\frac{N-1}{|\mathbf{r}_{N-1}-\mathbf{r}_N|}-\varepsilon_k\right)\Psi^{(N)}(\mathbf{z}_1,\ldots,\mathbf{z}_N)d\mathbf{z}_N \right.\\ &\left.\times \left| \int \phi_k(\mathbf{z}_N)\left(-\frac{\nabla_N^2}{2}+v(\mathbf{z}_N)+\frac{N-1}{|\mathbf{r}_{N-1}-\mathbf{r}_N|}-\varepsilon_k\right)\Psi^{(N)}(\mathbf{z}_1,\ldots,\mathbf{z}_N)d\mathbf{z}_N \right\rangle \right.\\ &+\frac{N-2}{N-1}\left\langle \int \phi_k(\mathbf{z}_N)\left(-\frac{\nabla_N^2}{2}+v(\mathbf{z}_N)+\frac{N-1}{|\mathbf{r}_{N-1}-\mathbf{r}_N|}-\varepsilon_k\right)\Psi^{(N)}(\mathbf{z}_1,\ldots,\mathbf{z}_N)d\mathbf{z}_N \right.\\ &\left.\times \left| \int \phi_k(\mathbf{z}_N)\left(-\frac{\nabla_N^2}{2}+v(\mathbf{z}_N)+\frac{N-1}{|\mathbf{r}_{N-2}-\mathbf{r}_N|}-\varepsilon_k\right)\Psi^{(N)}(\mathbf{z}_1,\ldots,\mathbf{z}_N)d\mathbf{z}_N \right\rangle \right.\end{aligned}. \tag{A6}$$

This can be rewritten in terms of reduced density matrices

$$0=\frac{2}{N(N-1)^2}\int\int\int\int d\mathbf{z}_{N-1}d\mathbf{z}_N d\mathbf{z}'_{N-1}d\mathbf{z}'_N\Bigg\{\delta(\mathbf{z}_{N-1}-\mathbf{z}'_{N-1})\phi_k(\mathbf{z}'_N)\phi_k(\mathbf{z}_N)\Bigg(-\frac{\nabla^2_{N'}}{2}+v(\mathbf{z}'_N)+\frac{N-1}{|\mathbf{r}'_{N-1}-\mathbf{r}'_N|}-\varepsilon_k\Bigg)$$
$$\times\Bigg(-\frac{\nabla^2_N}{2}+v(\mathbf{z}_N)+\frac{N-1}{|\mathbf{r}_{N-1}-\mathbf{r}_N|}-\varepsilon_k\Bigg)\Gamma^{(N)}_2(\mathbf{z}'_{N-1},\mathbf{z}'_N;\mathbf{z}_{N-1},\mathbf{z}_N)\Bigg\}+\frac{6}{N(N-1)^2}\int\int\cdots\int\int d\mathbf{z}_{N-2}\cdots d\mathbf{z}_N d\mathbf{z}'_{N-2}\cdots d\mathbf{z}'_N$$
$$\times\Bigg\{\delta(\mathbf{z}_{N-1}-\mathbf{z}'_{N-1})\delta(\mathbf{z}_{N-2}-\mathbf{z}'_{N-2})\phi_k(\mathbf{z}'_N)\phi_k(\mathbf{z}_N)\Bigg(-\frac{\nabla^2_{N'}}{2}+v(\mathbf{z}'_N)+\frac{N-1}{|\mathbf{r}'_{N-1}-\mathbf{r}'_N|}-\varepsilon_k\Bigg)$$
$$\times\Bigg(-\frac{\nabla^2_N}{2}+v(\mathbf{z}_N)+\frac{N-1}{|\mathbf{r}_{N-2}-\mathbf{r}_N|}-\varepsilon_k\Bigg)\Gamma^{(N)}_3(\mathbf{z}'_{N-2},\mathbf{z}'_{N-1},\mathbf{z}'_N;\mathbf{z}_{N-2},\mathbf{z}_{N-1},\mathbf{z}_N)\Bigg\}. \tag{A7}$$

To derive this equation the normalization of the k-particle density matrices to $N(N-1)\cdots(N-k)/k!$ must be taken into account. Multiplying both sides of the equation by $N(N-1)^2/2$ and explicitly performing the integrations with respect to the delta function gives

$$0=\int\cdots\int d\mathbf{z}_1 d\mathbf{z}'_1 d\mathbf{z}_2\Bigg\{\phi_k(\mathbf{z}'_1)\phi_k(\mathbf{z}_1)\Bigg(-\frac{1}{2}\nabla^2_{1'}+v_{\text{ext}}(\mathbf{r}'_1)+(N-1)|\mathbf{r}_2-\mathbf{r}'_1|^{-1}-\varepsilon_k\Bigg)$$
$$\times\Bigg(-\frac{1}{2}\nabla^2_1+v_{\text{ext}}(\mathbf{r}_1)+(N-1)|\mathbf{r}_2-\mathbf{r}_1|^{-1}-\varepsilon_k\Bigg)\Gamma_2(\mathbf{z}_1,\mathbf{z}_2;\mathbf{z}'_1,\mathbf{z}_2)\Bigg\}$$
$$+3\int\cdots\int d\mathbf{z}_1 d\mathbf{z}'_1 d\mathbf{z}_2 d\mathbf{z}_3\Bigg\{\phi_k(\mathbf{z}'_1)\phi_k(\mathbf{z}_1)\Bigg(-\frac{1}{2}\nabla^2_{1'}+v_{\text{ext}}(\mathbf{r}'_1)+(N-1)|\mathbf{r}_2-\mathbf{r}'_1|^{-1}-\varepsilon_k\Bigg)$$
$$\times\Bigg(-\frac{1}{2}\nabla^2_1+v_{\text{ext}}(\mathbf{r}_1)+(N-1)|\mathbf{r}_3-\mathbf{r}_1|^{-1}-\varepsilon_k\Bigg)\Gamma_3(\mathbf{z}_1,\mathbf{z}_2,\mathbf{z}_3;\mathbf{z}'_1,\mathbf{z}_2,\mathbf{z}_3)\Bigg\}. \tag{A8}$$

Using the relationship between the third- and second- order density matrices gives

$$0=\frac{3}{(N-2)}\int\cdots\int d\mathbf{z}_1 d\mathbf{z}'_1 d\mathbf{z}_2 d\mathbf{z}_3\Bigg\{\phi_k(\mathbf{z}'_1)\phi_k(\mathbf{z}_1)\Bigg(-\frac{1}{2}\nabla^2_{1'}+v_{\text{ext}}(\mathbf{r}'_1)+(N-1)|\mathbf{r}_2-\mathbf{r}'_1|^{-1}-\varepsilon_k\Bigg)$$
$$\times\Bigg(-\frac{1}{2}\nabla^2_1+v_{\text{ext}}(\mathbf{r}_1)+(N-1)|\mathbf{r}_2-\mathbf{r}_1|^{-1}-\varepsilon_k\Bigg)\Gamma_3(\mathbf{z}_1,\mathbf{z}_2,\mathbf{z}_3;\mathbf{z}'_1,\mathbf{z}_2,\mathbf{z}_3)\Bigg\}$$
$$+3\int\cdots\int d\mathbf{z}_1 d\mathbf{z}'_1 d\mathbf{z}_2 d\mathbf{z}_3\Bigg\{\phi_k(\mathbf{z}'_1)\phi_k(\mathbf{z}_1)\Bigg(-\frac{1}{2}\nabla^2_{1'}+v_{\text{ext}}(\mathbf{r}'_1)+(N-1)|\mathbf{r}_2-\mathbf{r}'_1|^{-1}-\varepsilon_k\Bigg)$$
$$\times\Bigg(-\frac{1}{2}\nabla^2_1+v_{\text{ext}}(\mathbf{r}_1)+(N-1)|\mathbf{r}_3-\mathbf{r}_1|^{-1}-\varepsilon_k\Bigg)\Gamma_3(\mathbf{z}_1,\mathbf{z}_2,\mathbf{z}_3;\mathbf{z}'_1,\mathbf{z}_2,\mathbf{z}_3)\Bigg\}. \tag{A9}$$

Multiplying by $(N-2)/3$ and combining terms gives Eq. (30).

[1] O. W. Day, D. W. Smith, and R. C. Morrison, J. Chem. Phys. **62**, 115 (1975).
[2] D. W. Smith and O. W. Day, J. Chem. Phys. **62**, 113 (1975).
[3] M. M. Morrell, R. G. Parr, and M. Levy, J. Chem. Phys. **62**, 549 (1975).
[4] J. C. Ellenbogen, O. W. Day, D. W. Smith, and R. C. Morrison, J. Chem. Phys. **66**, 4795 (1977).
[5] R. C. Morrison, C. M. Dixon, and J. R. Mizell, Int. J. Quantum Chem. **S28**, 309 (1994).
[6] R. C. Morrison, Int. J. Quantum Chem. **49**, 649 (1994).
[7] R. C. Morrison, J. Chem. Phys. **99**, 6221 (1993).
[8] R. C. Morrison, J. Chem. Phys. **96**, 3718 (1992).
[9] R. C. Morrison and G. H. Liu, J. Comput. Chem. **13**, 1004 (1992).
[10] R. C. Morrison and P. W. Ayers, J. Chem. Phys. **103**, 6556 (1995).
[11] D. Sundholm and J. Olsen, J. Chem. Phys. **99**, 6222 (1993).
[12] K. Pernal and J. Cioslowski, J. Chem. Phys. **114**, 4359 (2001).
[13] R. C. Morrison, W. Tong, and O. W. Day, Int. J. Quantum Chem. **60**, 421 (1996).
[14] W. Tong, R. C. Morrison, and O. W. Day, Int. J. Quantum Chem. **60**, 411 (1996).
[15] J. Cioslowski, P. Piskorz, and G. H. Liu, J. Chem. Phys. **107**, 6804 (1997).
[16] S. Verdonck, D. Van Neck, P. W. Ayers, and M. Waroquier, Phys. Rev. A **74**, 062503 (2006).
[17] K. Pernal and J. Cioslowski, Chem. Phys. Lett. **412**, 71 (2005).
[18] P. Leiva and M. Piris, J. Chem. Phys. **123**, 214102 (2005).
[19] P. Leiva and M. Piris, Int. J. Quantum Chem. **107**, 1 (2007).
[20] P. Leiva and M. Piris, J. Mol. Struct.: THEOCHEM **770**, 45 (2006).
[21] C. Garrod, M. V. Mihailovic, and M. Rosina, J. Math. Phys. **16**, 868 (1975).
[22] L. Cohen and C. Frishberg, Phys. Rev. A **13**, 927 (1976).
[23] H. Nakatsuji, Phys. Rev. A **14**, 41 (1976).
[24] C. Valdemoro, L. M. Tel, and E. Perez-Romero, Adv. Quantum Chem. **28**, 33 (1997).
[25] M. Nakata, H. Nakatsuji, M. Ehara, M. Fukuda, K. Nakata, and K. Fujisawa, J. Chem. Phys. **114**, 8282 (2001).
[26] M. Ehara, M. Nakata, H. Kou, K. Yasuda, and H. Nakatsuji, Chem. Phys. Lett. **305**, 483 (1999).
[27] D. A. Mazziotti, Acc. Chem. Res. **39**, 207 (2006).
[28] D. A. Mazziotti, Phys. Rev. A **65**, 062511 (2002).
[29] D. A. Mazziotti, Phys. Rev. A **57**, 4219 (1998).

[30] R. G. Parr and W. T. Yang, J. Am. Chem. Soc. **106**, 4049 (1984).
[31] W. T. Yang, R. G. Parr, and R. Pucci, J. Chem. Phys. **81**, 2862 (1984).
[32] P. W. Ayers and M. Levy, Theor. Chem. Acc. **103**, 353 (2000).
[33] P. W. Ayers and J. Melin, Theor. Chem. Acc. **117**, 371 (2007).
[34] B. T. Pickup and J. G. Snijders, Chem. Phys. Lett. **153**, 69 (1988).
[35] B. T. Pickup, Chem. Phys. Lett. **33**, 422 (1975).
[36] J. Olsen and D. Sundholm, Chem. Phys. Lett. **288**, 282 (1998).
[37] R. C. Morrison, Z. X. Zhou, and R. G. Parr, Theor. Chim. Acta **86**, 3 (1993).
[38] D. Sundholm and J. Olsen, J. Chem. Phys. **98**, 3999 (1993).
[39] N. C. Handy, M. T. Marron, and H. J. Silverstone, Phys. Rev. **180**, 45 (1969).
[40] J. Katriel and E. R. Davidson, Proc. Natl. Acad. Sci. U.S.A. **77**, 4403 (1980).
[41] R. Ahlrichs, M. Hoffmann-Ostenhof, T. Hoffmann-Ostenhof, and J. D. Morgan III, Phys. Rev. A **23**, 2106 (1981).
[42] M. Levy, J. P. Perdew, and V. Sahni, Phys. Rev. A **30**, 2745 (1984).
[43] M. Levy and R. G. Parr, J. Chem. Phys. **64**, 2707 (1976).
[44] D. Van Neck, A. E. L. Dieperink, and M. Waroquier, Phys. Rev. C **53**, 2231 (1996).
[45] D. Van Neck, S. Rombouts, and S. Verdonck, Phys. Rev. C **72**, 054318 (2005).
[46] O. V. Gritsenko and E. J. Baerends, J. Chem. Phys. **120**, 8364 (2004).
[47] O. V. Gritsenko, B. Braida, and E. J. Baerends, J. Chem. Phys. **119**, 1937 (2003).
[48] O. V. Gritsenko and E. J. Baerends, J. Chem. Phys. **117**, 9154 (2002).
[49] P. W. Ayers and W. T. Yang, J. Chem. Phys. **124**, 224108 (2006).
[50] B. C. Karlson and J. M. Keller, Phys. Rev. **121**, 659 (1961).
[51] M. Ernzerhof, J. Chem. Theory Comput. **5**, 793 (2009).

Bibliography

[1] S. G. Arturo and D. E. Knox. *J. Mol. Struc.: THEOCHEM*, 770:31, 2006.

[2] P. L. A. Popelier, M. Devereux, and M. Rafat. *Acta Cryst.*, 60:427, 2004.

[3] P. Geerlings, F. De Proft, and W. Langenaeker. *Chem. Rev.*, 103:1793, 2003.

[4] W. H. Dickhoff and D. Van Neck. *Many-body theory exposed! Propagator description of quantum mechanics in many-body systems*. World Scientific Publishing: Singapore, 2005.

[5] F. Jensen. *Introduction to computational chemistry*. John Wiley and Sons, 1999.

[6] R. J. Meier. *Vibrational Spectroscopy*, 1:26, 2007.

[7] A. K. Rappe and C. J. Casewit. *Molecular Mechanics Across Chemistry*. University Science Books, 1997.

[8] R. F. W. Bader. *Atoms in Molecules: A Quantum Theory*. Oxford University Press: Oxford, 1994.

[9] P. A. M. Dirac. *The principles of quantum mechanics*. Oxford University Press, 1988.

[10] A. J. Coleman. *Int. J. Quant. Chem.*, 85:196, 2001.

[11] J. A. Pople, R. Seeger, and R. Krishnan. *12*, Int. J. Quantum Chem.:149, 1977.

[12] K. Raghavachari, H. B. Schlegel, and J. A. Pople. *J. Chem. Phys.*, 72:4654, 1980.

[13] K. Raghavachari and J. A. Pople. *Int. J. Quantum Chem.*, 20:1067, 1981.

[14] Attila Szabo and Neil S. Ostlund. *Modern quantum chemistry: introduction to advanced electronic structure theory*. Courier Dover Publications, 1996.

[15] R. J. Bartlett and G. D. Purvis. *Int. J. Quantum Chem.*, 14:561, 1978.

[16] J. A. Pople, R. Krishnan, H. B. Schlegel, and J. S. Binkley. *Int. J. Quantum Chem.*, 14:545, 1978.

[17] J. Cizek. *Advances in Chem. Phys*, 14:35, 1969.

[18] G. D. Purvis and R. J. Bartlett. *J. Chem. Phys.*, 76:1910, 1982.

[19] G. E. Scuseria, C. L. Janssen, and H. F. Schaefer. *J. Chem. Phys.*, 89:7382, 1988.

[20] G. E. Scuseria and H. F. Schaefer. *J. Chem. Phys.*, 90:3700, 1989.

[21] C. Moller and M. S. Plesset. *Phys. Rev.*, 46:618, 1934.

[22] M. J. Frisch M. Head-Gordon, J. A. Pople. *Chem. Phys. Lett.*, 153(6):503, 1988.

[23] J. A. Pople R. Krishnan. *Int. J. Quant. Chem.*, 14(1):91, 1978.

[24] C. C. J. Roothaan. *Rev. Mod. Phys.*, 23:69, 1951.

[25] J. A. Pople and R. K. Nesbet. *J. Chem. Phys.*, 22:571, 1954.

[26] R. McWeeny and G. Dierksen. *J. Chem. Phys.*, 49:4852, 1968.

[27] P. Lowdin. *Phys. Rev.*, 97:1509, 1955.

[28] A. J. Coleman. *Rev. Mod. Phys.*, (35):668, 1963.

[29] C. Garrod and J. K. Percus. *J. Math. Phys.*, 5:1756, 1964.

[30] B. Verstichel, H.van Aggelen, D. Van Neck, P. W. Ayers, and P. Bultinck. *Phys. Rev. A*, 80.

[31] W. Koch and M. C. Holthausen. *A Chemist's Guide to Density Functional Theory*. Wiley-VCH: Weinheim, 2000.

[32] P. Hohenberg and W. Kohn. *Phys. Rev. B*, 136:864, 1964.

[33] L. Li and R. G. Parr. *J. of Chem. Phys.*, 84(3):1704, 1986.

[34] I. Mayer and A. Hamza. *Int. J. Quantum Chem.*, 103:798, 2005.

[35] R. S. Mulliken. *J. Chem. Phys.*, 23:1833, 1955.

[36] J. F. Rico, R. Lopez, and G. Ramirez. *J. Chem. Phys.*, 110:4213, 1999.

[37] J. F. Rico, R. Lopez, and G. Ramirez. *J. Chem. Theory. Comput.*, 1:1083, 2005.

[38] R. F. W. Bader. A quantum theory of molecular structure and its applications. *Chem. Rev.*, 91(5):893, 1991.

[39] P. Popelier. *Atoms In Molecules: An Introduction*. Prentice-Hall: Harlow, Great Britain, 2000.

[40] F. L. Hirshfeld. *Theor. Chem. Acc.*, 44(2):129, 1977.

[41] K. Kullback and R. A. Leibler. *Ann. Math. Stat.*, 22:79, 1951.

[42] R. F. Nalewajski and R. G. Parr. *Proc. Natl. Acad. Sci. USA*, 97:8879, 2000.

[43] R. F. Nalewajski and R. G. Parr. *J. Phys. Chem. A*, 105:7391, 2001.

[44] P. Bultinck, C. Van Alsenoy, P. W. Ayers, and R. Carbo-Dorca. *J. Chem. Phys.*, 126(14):144111, 2007.

[45] P. Bultinck. *Faraday Discussions*, 135:244, 2007.

[46] P. Bultinck, P. W. Ayers, S. Fias, K. Tiels, and C. Van Alsenoy. *Chem. Phys. Lett.*, 444(1-3):205, 2007.

[47] P. Bultinck, D. L. Cooper, and D. Van Neck. *Phys. Chem. Chem. Phys.*, 11(18):3424, 2009.

[48] J. P. Perdew, R. G. Parr, M. Levi, and J. L. Balduz Jr. *Phys. Rev. Lett.*, 49:1691, 1982.

[49] W. Yang, Y. Zhang, and P. W. Ayers. *Phys. Rev. Lett.*, 84:5172, 2000.

[50] D. Ghillemijn, P. Bultinck, D. Van Neck, and P. W. Ayers. *J. Comput. Chem.*, 32(8):1561, 2011.

[51] T. C. Lillestolen and R. J. Wheatley. *Chem. Commun.*, 45:5909, 2008.

[52] R.F.W. Bader. *J. Chem. Phys.*, 73:2871, 1980.

[53] P. W. Ayers, R. G. Parr, and A. Nagy. *Int. J. Quantum Chem.*, 90:309, 2002.

[54] J. S. M. Anderson, P. W. Ayers, and J. I. Rodriguez Hernandez. *J. Phys. Chem. A*, in press, 2010.

[55] L.A. Montero, L.A. Diaz, and R. Bader. *Introduction to Advanced Topics of Computational Chemistry*. Editorial de la Universidad de La Habana: Havana, 2003.

[56] P. Bultinck, S. Fias, C. Van Alsenoy, P. W. Ayers, and R. Carbo-Dorca. *J. Chem. Phys.*, 127:034102, 2007.

[57] I. Mayer and A. Hamza. *Theor. Chem. Acc.*, 105:360, 2001.

[58] P. Salvador, M. Duran, and I. Mayer. *J. Chem. Phys.*, 115:1153, 2001.

[59] A. M. Pendas, M. A. Blanco, and E. Francisco. *J. Chem. Phys.*, 120:4581, 2004.

[60] A. M. Pendas, E. Francisco, and M. A. Blanco. *J. Comput. Chem.*, 26:344, 2005.

[61] C. J. Cramer. *Essentials of Computational Chemistry: Theories and Methods*. Wiley, 2002.

[62] J. Meister and W. H. E. Schwarz. *J. Phys. Chem.*, 98:8245, 1994.

[63] B. H. Bezler, K. M. Merz, and P. A. Kollman. *J. Comput. Chem.*, 11:431, 1990.

[64] D. E. Williams. *Rev. Comput. Chem.*, 2:219, 1991.

[65] W. J. Mortier, K. Van Genechten, and J. Gasteiger. *J. Am. Chem. Soc.*, 107:829, 1985.

[66] W. J. Mortier, S. K. Ghosh, and S. Shankar. *J. Am. Chem. Soc.*, 108:4315, 1986.

[67] W. J. Mortier. *Structure and Bonding*, 66:125, 1987.

[68] K. A. Van Genechten, W. J. Mortier, and P. Geerlings. *Chem. Commun.*, page 1278, 1986.

[69] K. A. Van Genechten, W. J. Mortier, and P. Geerlings. *J. Chem. Phys.*, 86:5063, 1987.

[70] L. Uytterhoeven, W. J. Mortier, and P. Geerlings. *J. Phys. Chem. Solids*, 50:479, 1989.

[71] T. Verstraelen, V. Van Speybroeck, and M. Waroquier. *J. Chem. Phys.*, 131:044127, 2009.

[72] David L. Mobley, Elise Dumont, John D. Chodera, and Ken A. Dill. *J. Phys. Chem. B*, 111(2241), 2007.

[73] A. R. Edmonds. *Angular Momentum in Quantum Mechanics*. Princeton University Press, 1996.

[74] S. Van Damme, P. Bultinck, and S. Fias. *J. Chem. Theor. Comput.*, 5(2):334, 2009.

[75] M. M. Francl and L. E. Chirlian. *Rev. Comput. Chem.*, 14:1, 2000.

[76] J. Maranon, H. Grinberg, and N. S. Nudelman. *Int. J. Quant. Chem.*, 22(1):69, 1982.

[77] K. B. Wiberg. *Tetrahedron*, 24:1083, 1968.

[78] R. F. W. Bader and M. E. Stephens. *J. Am. Chem. Soc.*, 97:7391, 1975.

[79] M. Giambiagi, M. S. Giambiagi, D. R. Grempel, and C. D. J. Heymann. *J. Chim. Phys. Phys.-Chim. Biol.*, 72:15, 1975.

[80] I. Mayer. *Chem. Phys. Lett.*, 97:270, 1983.

[81] R. L. Fulton. *J. Phys. Chem.*, 97:7516, 1993.

[82] J. G. Angyan, M. Loos, and I. Mayer. *J. Phys. Chem.*, 98:5244, 1994.

[83] I. Mayer and P. Salvador. *Chem. Phys. Lett.*, 383:368, 2004.

[84] X. Fradera, M. A. Austen, and R. F. W. Bader. *J. Phys. Chem. A*, 103:304, 1999.

[85] R. Ponec and D. L. Cooper. *J. Mol. Struct. (Theochem)*, 727:133, 2005.

[86] I. Mayer. *J. Comput. Chem.*, 28:204, 2007.

[87] R. G. Parr and W. Yang. *Density-Functional Theory of Atoms and Molecules*. Oxford Science Publications, 1988.

[88] R. P. Iczkowski and J. L. Margrave. *Am. Chem. Soc.*, 83:3547, 1961.

[89] E. P. Gyftopoulos and G. N. Hatsopoulos. *Proc. Natl. Acad. Sci. U.S.A.*, 60:786, 1965.

[90] R. G. Parr and R. G. Pearson. *J. Am. Chem. Soc.*, 105:7512, 1983.

[91] P. Politzer. *J. Chem. Phys.*, 86:1072, 1987.

[92] A. T. Maynard, M. Huang, W. G. Rice, and D.G. Covell. *Proc. Natl. Acad. Sci. U.S.A.*, 95:11578, 1998.

[93] R. G. Parr, L. Von Szentpaly, and S. B. Liu. *J. Am. Chem. Soc.*, 121:1922, 1999.

[94] P. W. Ayers, J. S. M. Anderson, and L. J. Bartolotti. *Int. J. Quant. Chem.*, 101:520, 2005.

[95] P. K. Chattaraj and B. Maiti. *J. Phys. Chem. A*, 105:169, 2001.

[96] R. S. Mulliken. *Chem. Phys.*, 2:782, 1934.

[97] K. Fukui, Y. Yonezawa, and H. Shingu. *J. Chem. Phys.*, 20:722, 1952.

[98] R. G. Parr and W. Yang. *J. Am. Chem. Soc.*, 106:4049, 1984.

[99] P. W. Ayers and M. Levy. *Theor. Chem. Acc.*, 103:353, 2000.

[100] W. Yang and R. G. Parr. *Proc. Natl. Acad. Sci. U.S.A.*, 82:6723, 1985.

[101] W. Yang and W. J. Mortier. *J. Am. Chem. Soc.*, 108:5708, 1986.

[102] P. W. Ayers, R. C. Morrison, and R. K. Roy. *J. Chem. Phys.*, 116:8731, 2002.

[103] R. F. W. Bader and P. M. Beddall. *J. Chem. Phys.*, 56:3320, 1972.

[104] R. F. W. Bader, P. M. Beddall, and J. Jr. Peslak. *J. Chem. Phys.*, 58:557, 1973.

[105] S. Srebrenik and R. F. W. Bader. *J. Chem. Phys.*, 63:3945, 1975.

[106] S. Srebrenik, R. F. W. Bader, and T. T. Nguyen-Dang. *J. Chem. Phys.*, 68:3667, 1978.

[107] J. Cioslowski and J. Karwowski. *Fundamentals of Molecular Similarity*. Kluwer Academic: Dordrecht, 2001.

[108] R. F. Nalewajski and J. Mrozek. *Int. J. Quant. Chem.*, 51(4):187, 1994.

[109] A. Michalak, R. L. DeKock, and T. Ziegler. *J. Phys. Chem. A*, 112(31):7256, 2008.

[110] A. Michalak, M. Mitoraj, and T. Ziegler. *J. Phys. Chem. A*, 112(9):1933, 2008.

[111] I. Mayer and P. Salvador. *J. Chem. Phys.*, 130(23):234106, 2009.

[112] Diego R. Alcoba, Luis Lain, Alicia Torre, and Roberto C. Bochicchio. *Journal of Chemical Physics*, 123(14):144113, 2005.

[113] C.F. Matta and R.F.W. Bader. *J. Phys. Chem. A*, 110:6365, 2006.

[114] Y. Zhang and A. Wasserman. *J. Chem. Theory Comput.*, 6(11):3312, 2010.

[115] A. M. Pendas, M. A. Blanco, and E. Francisco. *J. Comput. Chem.*, 28:161, 2007.

[116] D. Vanfleteren, D. Van Neck, P. Bultinck, P. W. Ayers, and M. Waroquier. *J. Chem. Phys.*, 132:164111, 2010.

[117] I. Mayer. *Int. J. Quant. Chem.*, 26:151, 1984.

[118] R. Ponec and D. L. Cooper. *Faraday Discuss*, 135:31, 2007.

[119] M. A. Blanco, A. M. Pendas, and E. Francisco. *J. Chem. Theory Comput.*, 1:1096, 2005.

[120] A. M. Pendas, E. Francisco, and M. A. Blanco. *J. Phys. Chem. A*, 110:12864, 2006.

[121] A. M. Pendas, E. Francisco, and M. A. Blanco. *J. of Chem. Phys.*, 125:184112, 2006.

[122] D. Vanfleteren, D. Ghillemijn, D. Van Neck, P. Bultinck, M. Waroquier, and P. W. Ayers. *J. Comput. Chem.*, submitted, 2011.

[123] P. W. Ayers. *J. Math. Chem.*, 43:285, 2008.

[124] I. Mayer. *Simple Theorems, Proofs, And Derivations In Quantum Chemistry*. Kluwer Academic Pub, 2003.

[125] P. R. Watson, M. A. Van Hove, and K. Hermann. *NIST Computational Chemistry Comparison and Benchmark DataBase - Release 15a*. National Institute of Standards and Technology, Gaithersburg, MD, 2004.

[126] P. Bultinck, D. L. Cooper, and R. Ponec. *J. Phys. Chem. A*, 114:8754, 2010.

[127] F. R. Burden and R. M. Wilson. *Adv. Phys.*, 21:825, 1972.

[128] R. Ahlrichs and P. R. Taylor. *J. Chim. Phys.*, 78:315, 1981.

[129] R.A. Poirier, R.E. Kari, and I.G. Csizmadia. *Handbook of Gaussian Basis Sets*. Elsevier: Amsterdam, 1985.

[130] D. Vanfleteren, D. Van Neck, P. Bultinck, P. W. Ayers, and M. Waroquier. *J. Chem. Phys.*, 133:231103, 2010.

[131] D. Vanfleteren, D. Van Neck, P. Bultinck, P. W. Ayers, and M. Waroquier. *J. Chem. Phys.*, submitted, 2011.

[132] A. D. McNaught and A. Wilkinson. *IUPAC. Compendium of Chemical Terminology, 2nd ed. (the "Gold Book")*. Blackwell Scientific Publications: Oxford, 1997.

[133] C. Hansch, A. Leo, and R. W. Taft. *Chem. Rev.*, 91(2):165, 1991.

[134] J. E. Huheey. *J. Org. Chem.*, 36:204, 1971.

[135] L. Komorowski, J. Lipinski, and M. J. Pyka. *J. Phys. Chem.*, 97:3166, 1993.

[136] E. P. Serjeant and B. Dempsey. *Ionization Constants of Organic Acids in Solution, IUPAC Chemical Data Series No. 23*. Pergamon Press: Oxford, 1979.

[137] L. J. Pauling. *Am. Chem. Soc.*, 54:3570, 1932.

[138] E. A. Keiter J. E. Huheey and R. L. Keiter. *Inorganic Chemistry: Principles of Structure and Reactivity (4th Edition)*. Prentice Hall, 1997.

[139] R. T. Sanderson. *Science*, 114:670, 1951.

[140] R. T. Sanderson. *Science*, 116:41, 1952.

[141] R. T. Sanderson. *J. Chem. Educ.*, 29:539, 1952.

[142] R. T. Sanderson. *Science*, 121:207, 1955.

[143] R. T. Sanderson. *Chemical Bonds and Bond Energy*. Academic Press: New York, 1976.

[144] R. T. Sanderson. *Polar Covalence*. Academic Press: New York, 1983.

[145] R.G. Parr, R. A. Donnelly, M. Levy, and W. E. Palke. *J. Chem. Phys.*, 68:3801, 1978.

[146] W. Kohn, A. D. Becke, and R. G. Parr. *J. Phys. Chem.*, 100:12974, 1996.

[147] T. A. Koopmans. *Physica*, 1:104, 1933.

[148] T. T. Nguyen-Dang, R. F. W. Bader, and H. Essen. *Int. J. Quantum Chem.*, 22:1049, 1982.

[149] L. C. Allen. *J. Am. Chem. Soc.*, 111:9005, 1989.

[150] L. C. Allen. *Acc. Chem. Res.*, 23:175, 1990.

[151] L. C. Allen and E. T. Knight. *J. Mol. Struct. (THEOCHEM)*, 261:313, 1992.

[152] L. C. Allen. *J. Am. Chem. Soc.*, 114:1510, 1992.

[153] L. C. Allen. *Int. J. Quantum Chem.*, 49:253, 1994.

[154] R. A. Donnelly and R. G. Parr. *J. Chem. Phys.*, 69:4431, 1978.

[155] P. Politzer and H. Weinstein. *J. Chem. Phys.*, 71:4218, 1979.

[156] R. G. Parr and L. J. Bartolotti. *J. Am. Chem. Soc.*, 104:3801, 1982.

[157] P. Jaramillo, P. Perez, R. Contreras, W. Tiznado, and P. Fuentealba. *J. Phys. Chem. A*, 110:8181, 2006.