

Critical analysis and extension of the Hirshfeld atoms in molecules

Patrick Bultinck^{a)}

Department of Inorganic and Physical Chemistry, Ghent University, Krijgslaan 281 (S3), 9000 Gent, Belgium

Christian Van Alsenoy

Department of Chemistry, Antwerp University, Universiteitsplein 1, 2610 Antwerpen, Belgium

Paul W. Ayers

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

Ramon Carbó-Dorca

Department of Inorganic and Physical Chemistry, Ghent University, Krijgslaan 281 (S3), 9000 Gent, Belgium and Institute of Computational Chemistry, University of Girona, Campus de Montilivi, 17005 Girona, Spain

(Received 31 October 2006; accepted 15 February 2007; published online 13 April 2007)

The computational approach to the Hirshfeld [Theor. Chim. Acta **44**, 129 (1977)] atom in a molecule is critically investigated, and several difficulties are highlighted. It is shown that these difficulties are mitigated by an alternative, iterative version, of the Hirshfeld partitioning procedure. The iterative scheme ensures that the Hirshfeld definition represents a mathematically proper information entropy, allows the Hirshfeld approach to be used for charged molecules, eliminates arbitrariness in the choice of the promolecule, and increases the magnitudes of the charges. The resulting “Hirshfeld-I charges” correlate well with electrostatic potential derived atomic charges.

© 2007 American Institute of Physics. [DOI: 10.1063/1.2715563]

INTRODUCTION

Although it lacks a true basis in quantum mechanics, few ideas have been as successful in chemistry as the view that molecules consist of atoms held together by chemical bonds. As a direct consequence, a lot of work has been devoted to the concept of the atoms in molecules (AIM). We stress that AIM are not restricted to Bader's,^{1,2} approach, which is henceforth referred to as the quantum chemical topology (QCT) approach. The common purpose of all AIM descriptions has always been to be able to understand issues such as molecular similarity, transferability, etc., between molecules. A logical step further is the realization of the functional group, considered a main determining concept for molecular properties. As the AIM is not an observable, there is no unique way to define it and many different approaches have been developed, sometimes giving quite divergent results. As stated by Parr *et al.* the AIM is a noumenon,³ meaning that it is a construction of the mind or the intellect and so, despite its utility, the AIM cannot be directly observed by experiment, nor can one measure enough properties of an AIM to define it unambiguously.

Within the large set of different definitions of the AIM, two main groups of methods can be discerned. The first one describes an atomic density based on the atom centered basis functions. The most famous among these Hilbert space approaches is, without doubt, the Mulliken technique⁴⁻⁷ where a projection operator is used to obtain atomic densities. This projection operator for an atom A is given by⁸

$$\pi_A = \sum_{\nu \in A} \sum_{\mu} S_{\nu\mu}^{(-1)} |\nu\rangle\langle\mu|, \quad (1)$$

where ν and μ are the atom centered basis functions used in the molecular calculation. \mathbf{S} is the overlap matrix of the basis functions with elements $S_{\nu\mu}$ and $S_{\nu\mu}^{(-1)}$ refers to an element of the inverse of the matrix \mathbf{S} . An atomic density is then obtained from the self-consistent-field (SCF) density via

$$\begin{aligned} \rho_A(\mathbf{r}) &= \sum_{\nu \in A} \sum_{\mu} S_{\nu\mu}^{(-1)} |\nu\rangle\langle\mu| \rho(\mathbf{r}) \\ &= \sum_{\nu \in A} \sum_{\mu} S_{\nu\mu}^{(-1)} |\nu\rangle\langle\mu| \left[\sum_{\sigma\lambda} D_{\sigma\lambda} |\sigma\rangle\langle\lambda| \right] \\ &= \sum_{\nu \in A} \sum_{\lambda} D_{\nu\lambda} |\nu\rangle\langle\lambda|, \end{aligned} \quad (2)$$

where now \mathbf{D} is the charge and bond order matrix. Despite some computational shortcomings,⁹ the Mulliken approach is still among the most popular approaches to the AIM. The Mulliken projection operator is just one example of a more general class of projection based AIM approaches, as was discussed by Clark and Davidson.¹⁰

A different approach identifies AIM by dissecting the Cartesian space in so-called atomic basins. Based on the topology of the electron distribution, Bader's QCT theory,^{1,2} yields the mutually exclusive atomic basins Ω_A , and using a weighting function $w_A(\mathbf{r})$ the AIM density becomes

$$\rho_A(\mathbf{r}) = w_A(\mathbf{r}) \rho(\mathbf{r}) = \delta(\mathbf{r} \in \Omega_A) \rho(\mathbf{r}), \quad (3)$$

where the notion $\delta(\mathbf{r} \in \Omega_A)$ is a so-called logical Dirac notation,¹¹⁻¹³ meaning that it results in 1 if the point \mathbf{r} lies in the atomic basin Ω_A of the atom A and 0 otherwise. Clearly,

^{a)}Fax: +(32) 926-449-83; electronic mail: patrick.bultinck@ugent.be

Bader's QCT approach is a binary approach; either all density in a point is attributed to the atom A or none of it is. An alternative method in Cartesian space, reminiscent of procedures used in x-ray crystallography, was suggested by Hirshfeld.¹⁴ The idea there is that AIM do not have sharp boundaries. Instead, the electron density in each point is distributed among all atoms. The weighting function $w_A(\mathbf{r})$ is now computed from a so-called promolecular density,

$$w_A(\mathbf{r}) = \frac{\rho_A^0(\mathbf{r})}{\rho_{\text{mol}}^0(\mathbf{r})} = \frac{\rho_A^0(\mathbf{r})}{\sum_A \rho_A^0(\mathbf{r})}. \quad (4)$$

$\rho_A^0(\mathbf{r})$ is the electron density computed for the isolated atom A , and the promolecular density $\rho_{\text{mol}}^0(\mathbf{r})$ is the density of the superposition of all $\rho_A^0(\mathbf{r})$ with all atoms A positioned as in the real molecule. The true molecular density is denoted $\rho_{\text{mol}}(\mathbf{r})$. The AIM densities are then obtained via

$$\rho_A(\mathbf{r}) = w_A(\mathbf{r})\rho(\mathbf{r}) = \frac{\rho_A^0(\mathbf{r})}{\rho_{\text{mol}}^0(\mathbf{r})}\rho_{\text{mol}}(\mathbf{r}). \quad (5)$$

This so-called Hirshfeld scheme is particularly popular within so-called conceptual density functional theory^{15,16} (DFT) and the weighting function, which allows identifying the AIM as that which is most similar to the isolated atom,¹⁷ has been shown to be directly derivable from information entropy.^{3,18–22} It is clear that the Hirshfeld approach results in overlapping atoms. As such, it is also reminiscent of other techniques such as the Voronoi approach of Baerends and co-workers,^{23,24} the fuzzy atoms of Mayer and Salvador,²⁵ or Becke's scheme.²⁶

Although AIM allow much more information to be obtained for every atom, the AIM density functions are most used as a source of atomic populations through straightforward integration. Despite the fact that the Hirshfeld AIM is quite popular, the present paper will show some shortcomings of the model. Subsequently, it will be shown that these shortcomings can be alleviated by a new scheme that we have implemented and tested for a wide range of molecules.

ANALYSIS OF THE HIRSHFELD AIM

In order to show the shortcomings of Hirshfeld's AIM, a diatomic molecule $A-B$ will be used as a typical example. The focus will lie mainly on atomic charges or populations. Four major shortcomings will be discussed. Not all of these shortcomings pertain to the deeper context of previous theoretical work on the Hirshfeld idea, but all are relevant to the way these charges are currently obtained computationally and used, for example, in reactivity studies. As will be shown, meaningful charges can be obtained from the present, newly derived algorithm.

(1) *Hirshfeld atomic charges tend to be virtually zero.* It can be argued that because there is no unique charge definition, the exact charges of AIM are unknown. However, nearly all other charge schemes give significantly larger atomic charges than the Hirshfeld scheme and there appears to be a consensus that Hirshfeld charges are too small. This observation was first dis-

cussed in detail by Davidson and Chakravorty, who also discussed the arbitrariness of the Hirshfeld method.²⁷ This discussion, as will be shown below, in fact, agrees with the presently developed new approach. The reason why Hirshfeld charges are so small can be related to the work by Ayers who has shown that the Hirshfeld weighting factor is such that it makes the AIM as similar as possible to the isolated atom.^{17,28} It is therefore not surprising that also the atomic properties of the AIM are as similar as possible to that of the isolated atom. One of these properties is the atomic population.

It will prove important to define atomic charges as

$$q_A = Z_A - N_A, \quad (6)$$

where N_A is the atomic electronic population obtained from integration over all space of the AIM density and Z_A is the atomic number. The reason to keep strictly this unique definition is that originally in Hirshfeld's paper: the atomic charge was considered equal to

$$q'_A = \int \rho_A^0(\mathbf{r})d\mathbf{r} - \int \rho_A(\mathbf{r})d\mathbf{r} = \int [\rho_A^0(\mathbf{r}) - \rho_A(\mathbf{r})]d\mathbf{r}, \quad (7)$$

where $\rho_A^0(\mathbf{r})$ is the isolated atom density of atom A . The essence of the new Hirshfeld procedure described in the present study is precisely that with definition (7), q'_A should be zero always and for all atoms. This naturally does not mean that q_A is zero in definition (6). The fact that the aforementioned atomic similarity¹⁷ results in Hirshfeld populations where $N_A \approx N_A^0$, giving atoms with nearly zero q_A , is due to the use of atomic densities from neutral atoms ($Z_A = N_A^0$) (Ref. 14) to construct the promolecular density. Both definitions (6) and (7) are naturally only equivalent when $Z_A = N_A^0$ and it is this condition which is lifted in the present study.

(2) *The Hirshfeld AIM populations depend on the choice of the promolecular density.* As has been mentioned above, the promolecular density is usually that obtained from the superposition of atomic densities computed for neutral atoms. However, this choice has no strict theoretical basis; this convention is adopted merely for computational convenience. The resulting atomic Hirshfeld charges are thus determined not only by the molecular electron density, but also on an expedient, but arbitrary, choice for the promolecular density. A clear illustration of the quite strong dependence of AIM charges on the promolecular density is a molecule with important ionic character such as LiF. If one uses as promolecule the combination of Li^0 and F^0 densities, spatially organized as in the molecule itself, the atomic charges are calculated to have an absolute value of 0.57. However, it is chemically reasonable to consider a promolecule consisting of Li^+ and F^- . One would hope that the same charges will be obtained, but this is not the case; now the charges have an absolute value of 0.98. Using the "chemically unreasonable" combination Li^- and F^+ , still other charges are obtained; this

time the charges have magnitude 0.30 although, interestingly, the Li AIM remains positive. This sensitivity to the choice of promolecular reference is clearly undesirable.

- (3) *Hirshfeld charges are only available for neutral molecules.* This point is a corollary of the previous discussion. When a charged molecule is considered, the ambiguity on how to choose the promolecular density is exacerbated. In practical applications, the neutral promolecular densities are usually retained, but then

$$\int \rho_{\text{mol}}(\mathbf{r})d\mathbf{r} \neq \int \rho_{\text{mol}}^0(\mathbf{r})d\mathbf{r}. \quad (8)$$

one is venturing on a quite dangerous path when connecting Hirshfeld charges to information entropy, as will be discussed below. As expressed by Davidson and Chakravorty, the choice of referencing all density differences to neutral atoms is an arbitrary choice.²⁷

- (4) *Care needs to be taken in connecting Hirshfeld charges to information entropy.* This point was the subject of the recent paper by Parr *et al.* with the thought-provoking title “What is an Atom in a Molecule?”³ The weighting function for the Hirshfeld AIM has been shown to be related to information entropy as the function minimizing the information loss during the formation of the molecule,^{3,18-21}

$$I = \int \rho_A(\mathbf{r}) \ln \left(\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} \right) d\mathbf{r} + \int \rho_B(\mathbf{r}) \ln \left(\frac{\rho_B(\mathbf{r})}{\rho_B^0(\mathbf{r})} \right) d\mathbf{r}. \quad (9)$$

As Parr *et al.* mentioned, there is a very basic requirement in this information theory connection. Information entropy²⁹ as expressed through this Kullback-Liebler measure³⁰ explicitly requires that for every atom

$$N_A = \int \rho_A(\mathbf{r})d\mathbf{r} = \int \rho_A^0(\mathbf{r})d\mathbf{r} = N_A^0. \quad (10)$$

That is, the AIM needs to have the same electronic population as the promolecular atom. Failure to meet this condition somewhat complicates the exact connection between the Hirshfeld approach and information theory. When choosing as promolecular density the superposition of neutral atoms, requirement (10) is not fulfilled, although the difference is not very big. Yet, as will be shown below, even this small difference has an important impact on the final results. A solution for this problem was presented by Parr *et al.* and will be discussed in some detail below.

THE ITERATIVE SOLUTION

The problem at hand is how to solve the above problems and still retain the elegance of Hirshfeld’s idea, as well as to implement it properly so the Hirshfeld approach can continue to be used in applications.

Returning to the fourth problem discussed above, it is worth noting that Parr *et al.*³ recently discussed an elegant solution to the problem of nonfulfillment of requirement (10). This solution relies on using the shape function $\sigma(\mathbf{r})$.³¹ It has been shown on several occasions that the shape func-

tion bears the same information as the density function and can, e.g., even be used in variational SCF procedure.³²⁻³⁶

The shape function $\sigma_A(\mathbf{r})$ for atom A is given as

$$\sigma_A(\mathbf{r}) = \frac{\rho_A(\mathbf{r})}{N_A}, \quad (11)$$

with an analogous expression for $\sigma_A^0(\mathbf{r})$.

The idea of Parr *et al.* is to rewrite Eq. (9) using these shape functions. This gives

$$I = N_A \int \sigma_A(\mathbf{r}) \ln \left(\frac{\sigma_A(\mathbf{r})}{\sigma_A^0(\mathbf{r})} \right) d\mathbf{r} + N_B \int \sigma_B(\mathbf{r}) \ln \left(\frac{\sigma_B(\mathbf{r})}{\sigma_B^0(\mathbf{r})} \right) d\mathbf{r} + N_A \ln \left(\frac{N_A}{N_A^0} \right) + N_B \ln \left(\frac{N_B}{N_B^0} \right). \quad (12)$$

The fact that by definition one has

$$\int \sigma_A(\mathbf{r})d\mathbf{r} = \int \sigma_A^0(\mathbf{r})d\mathbf{r} = 1 \quad (13)$$

means that the sum of the integrals in Eq. (12) is fully in accordance with information entropy. The extra term that appears is the so-called mixing entropy, which was recently examined in detail.²² This term reflects the charge transfer occurring during the formation of the molecule. It also reflects the nonadherence to requirement (10). Depending on the promolecule chosen, one obtains a different mixing entropy term. This results in different charges for the AIM depending on the promolecule chosen. Conceptually, one could regard the formation of a molecule out of the set of isolated atoms as a process where first all atoms retain their own population and then gradually exchange electrons. Naturally, what one is interested in for the computation of AIM properties is the situation when no net charge transfer between the AIM’s occurs anymore. That is, ideally the AIM are determined from the molecular density function and are independent of the choice for the electron density of the promolecule. This demands considering the situation when the mixing entropy is zero and the loss of information measure (12) is only related to the deformation of the atomic densities.

It is only after requirement (10) is fulfilled that the loss of information is given by the Parr-Nalewajski form [Eq. (9)] for the information entropy, although the definition of the AIM itself is not affected. Nonadherence to requirement (10) does, however, affect the value of the information. Including Eq. (10) in Eq. (12) gives

$$I = N_A \int \sigma_A(\mathbf{r}) \ln \left(1 + \frac{\sigma_A(\mathbf{r})}{\sigma_A^0(\mathbf{r})} \right) d\mathbf{r} + N_B \int \sigma_B(\mathbf{r}) \ln \left(1 + \frac{\sigma_B(\mathbf{r})}{\sigma_B^0(\mathbf{r})} \right) d\mathbf{r} = N_A \int \sigma_A(\mathbf{r}) \ln \left(\frac{N_A \sigma_A(\mathbf{r})}{N_A^0 \sigma_A^0(\mathbf{r})} \right) d\mathbf{r} + N_B \int \sigma_B(\mathbf{r}) \ln \left(\frac{N_B \sigma_B(\mathbf{r})}{N_B^0 \sigma_B^0(\mathbf{r})} \right) d\mathbf{r}$$

$$= \int \rho_A(\mathbf{r}) \ln \left(\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} \right) d\mathbf{r} + \int \rho_B(\mathbf{r}) \ln \left(\frac{\rho_B(\mathbf{r})}{\rho_B^0(\mathbf{r})} \right) d\mathbf{r}. \quad (14)$$

Once condition (10) is required, Eq. (14) no longer suffers the weakness of different normalizations of $\rho_A(\mathbf{r})$ and $\rho_A^0(\mathbf{r})$.

There is another interesting point that can be mentioned with respect to the recent study by Ayers.²² It can be shown that the individual atomic contributions to the information measure (12) are always positive when only deformation of atomic densities is present. In order to show this, it is worth noting that the contribution I_A from atom A to the information loss measure in Eq. (12),

$$I_A = N_A \int \sigma_A(\mathbf{r}) \ln \left(\frac{\sigma_A(\mathbf{r})}{\sigma_A^0(\mathbf{r})} \right) d\mathbf{r} + N_A \ln \left(\frac{N_A}{N_A^0} \right), \quad (15)$$

is not always positive. So, ironically, an atom can gain information upon the formation of the molecule. This follows from the fact that the mixing entropy contribution can be either positive or negative. However, after all net charge transfer has stopped, one has

$$\begin{aligned} I_A &= N_A \int \sigma_A(\mathbf{r}) \ln \left(\frac{\sigma_A(\mathbf{r})}{\sigma_A^0(\mathbf{r})} \right) d\mathbf{r} \\ &= \int \rho_A(\mathbf{r}) \ln \left(\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} \right) d\mathbf{r} \\ &\geq \int \rho_A(\mathbf{r}) d\mathbf{r} - \int \rho_A^0(\mathbf{r}) d\mathbf{r} \geq 0, \end{aligned} \quad (16)$$

which shows that after charge transfer has taken place, the deformation contributions are always positive meaning always a loss of information. Note also that the information/entropy of an atom is only an extensive function of the number of electrons when the net charge transfer is zero.

In order to obtain true AIM properties from a molecule with no net charge transfer between the atoms anymore, one needs to perform the Hirshfeld calculations in such a way that requirement (10) is satisfied not only for the entire promolecule, but also for every reference atom that contributes to the promolecule. To our knowledge, no algorithm for this has been developed and discussed in detail, although Davidson and Chakravorty suggested the idea of a self-consistent set of Hirshfeld charges to avoid the arbitrariness of the procedure.²⁷ The goal of the present paper is thus to find the proper way to implement a Hirshfeld scheme that takes away the arbitrary character of the method and at the same time resolves all the problems discussed above. This means that we need to find a way to have the molecular electron density determine its own promolecule, rather than imposing it from the start.

The algorithm described here is essentially an iterative refinement scheme for the Hirshfeld promolecular reference. Starting from some freely chosen promolecule with populations $\{N_A^0\}$, the Hirshfeld populations $\{N_A^1\}$ are computed. These populations will normally differ from the normalization of the isolated atom densities that constituted the promolecule $\{N_A^1 \neq N_A^0\}$. One then computes isolated atomic densities that normalize to $\{N_A^1\}$ and uses the promolecule

constructed from the latter densities in the next iteration. In general, the weighing function for atom A in iteration i is given by

$$w_A^i(\mathbf{r}) = \frac{\rho_A^{i-1}(\mathbf{r})}{\rho_{\text{mol}}^{i-1}(\mathbf{r})}. \quad (17)$$

Defining Δ_A^i as the absolute value of the difference between the AIM population for an atom A (N_A^i) after the i th iteration and the isolated atomic density normalization (N_A^{i-1}) that was used in the promolecule for this iteration,

$$\Delta_A^i = \text{abs}(N_A^i - N_A^{i-1}), \quad (18)$$

one tries to find the converged solution where for all atoms $\Delta_A^{\text{conv}} = 0$. In other words, one seeks the set of $\{N_A^{\text{conv}}\}$ for which the mixing entropy term has become zero. In the new scheme, the classically computed Hirshfeld charges are the result of the first iteration (provided one starts from $\{N_A^0 = Z_A\}$) and the subsequent iterations reduce the gap between the normalization of the reference atomic density and the AIM density.

Of course, the atomic populations are usually fractional numbers. This requires a method for computing the corresponding densities. Two ways have been derived, leading to the same expression. First, the atom condensed Fukui function is used. Using the notations $\text{lint}(x)$ to express the integer part of x , $\text{uint}(x)$ as the higher integer number or $\text{uint}(x) = \text{lint}(x) + 1$ and mentioning explicitly the population in the atomic densities, one has

$$\begin{aligned} \rho_A^{N_A}(\mathbf{r}) &= \rho_A^{\text{lint}(N_A)}(\mathbf{r}) + f_A^{\text{uint}(N_A)+}(\mathbf{r})[N_A - \text{lint}(N_A)] \\ &= \rho_A^{\text{lint}(N_A)}(\mathbf{r}) + [\rho_A^{\text{lint}(N_A)+1}(\mathbf{r}) - \rho_A^{\text{lint}(N_A)}(\mathbf{r})][N_A - \text{lint}(N_A)] \\ &= \rho_A^{\text{lint}(N_A)}(\mathbf{r})[\text{uint}(N_A) - N_A] + \rho_A^{\text{uint}(N_A)}(\mathbf{r})[N_A - \text{lint}(N_A)], \end{aligned} \quad (19)$$

where a finite difference approach for the Fukui function was used. Naturally all the densities $\rho_A^{\text{lint}(N_A)}$ and $\rho_A^{\text{uint}(N_A)}$ need to be obtained from atomic calculations. It is very gratifying to note that this approach gives exactly the same density as the second method, namely, the one obtained from ensemble theory expressions given by Perdew *et al.*³⁷ and discussed in detail by Yang *et al.*³⁸ and Ayers.³⁹

APPLICATION OF THE HIRSHFELD-I SCHEME

In order to test the new Hirshfeld approach, henceforth referred to as the Hirshfeld-I method⁴⁰ and to see whether it effectively solves the problems discussed above, atomic charges were computed using several methods. As a test set, 168 molecules have been chosen containing C, H, N, O, F, and Cl atoms. The test set is available as supplementary material⁴¹ and comprises the entire set previously used in electronegativity equalization studies by Bultinck *et al.*^{42,43} This set contains sufficiently diverse molecules with typical functional groups to allow the study of the performance of the new method. In total 2122 atoms are present in the test set, with 1082 H, 712 C, 125 N, 110 O, 65 F, and 28 Cl atoms.

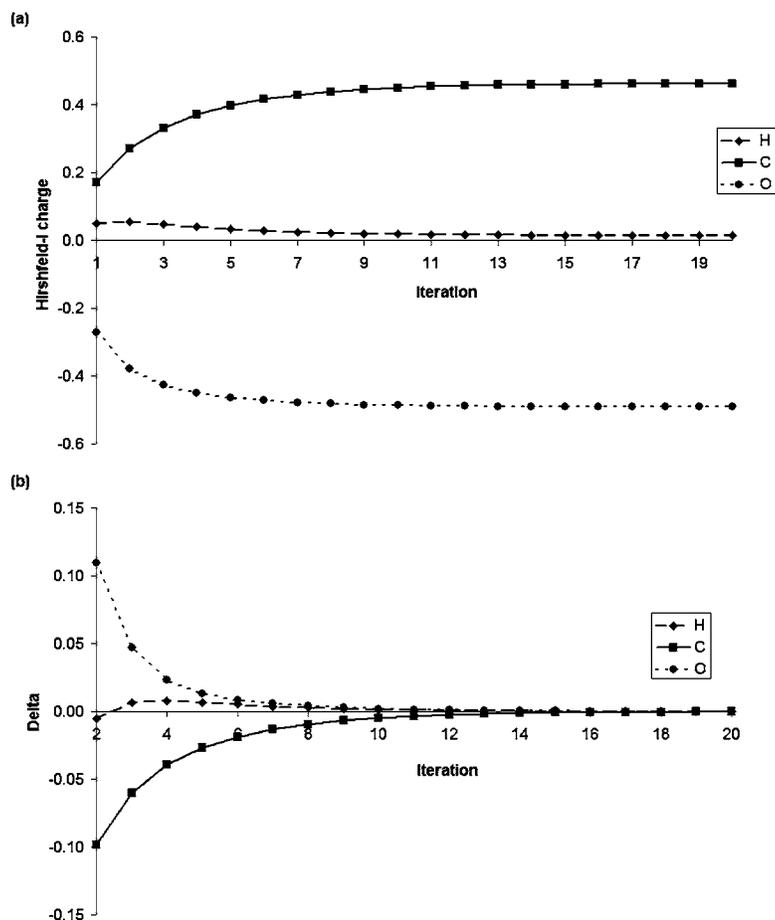


FIG. 1. Hirshfeld-I charge $\{q_A^i\}$ (a) and “delta” $\{\Delta_A^i\}$ (b) as a function of the number of iterations for the symmetry unique atoms in formaldehyde.

COMPUTATIONAL DETAILS

For all molecules, optimized B3LYP/6-31G* geometries were used. In order to implement the Hirshfeld-I scheme, formulas (5), (6), (18), and (19) were implemented in the STOCK program.⁴⁴ Atomic densities for fractional populations were computed as described above. The densities for integer population atomic densities from $\rho_A^{Z_A-2}$ to $\rho_A^{Z_A+2}$ for all elements encountered in the molecular set were obtained using ROHF/6-311++G** calculations with the ATOMSCF program by Roos *et al.*⁴⁵ Atoms are considered in the ground electronic state. Molecular electronic structure calculations were performed using the same basis set and using RHF calculations. In principle, it could be argued that one should take care with intramolecular basis set superposition error (BSSE) effects. Accounting for such effects could be done by, e.g., doing the atomic calculations in the full molecular basis set. As the basis set used in the present study is already quite large, it is assumed that BSSE will not have a dramatic effect on the resulting atomic populations. All molecular calculations were performed using GAUSSIAN 03.⁴⁶ For all molecules Mulliken populations were computed along with natural atomic populations.⁴⁷ Electrostatic potential (ESP) derived atomic populations were obtained using the CHELPG (Refs. 48 and 49) and Merz-Kollman-Singh⁵⁰ (MKS) methods.

Convergence of the Hirshfeld-I iterative scheme, implemented in STOCK, is considered obtained when $\Delta_A^{\text{conv}} < 0.0005$ for all atoms.

RESULTS AND DISCUSSION

Atomic charges

For most molecules, convergence in atomic charges was obtained within 25 iterations. In order to assess the effect of the iterative scheme, for the first iteration in the Hirshfeld-I scheme, the values for $\{N_A^0\}$ are chosen as $\{Z_A\}$. As a consequence, the solution of the first iteration gives the classical Hirshfeld populations. It is found that these original Hirshfeld populations $\{N_A^1\}$ differ only slightly from $\{Z_A\}$, in agreement with the similarity proof given by Ayers.¹⁷ Nevertheless, these small $\{\Delta_A^1\}$ can easily give rise to changes in atomic populations of several tenths, which is substantial and therefore gives quite drastic changes in atomic charges as well. To illustrate this, Fig. 1 shows how the atomic charges $\{q_A^i\}$ from Eq. (6) and $\{\Delta_A^i\}$ change upon continuing the iterative process for formaldehyde.

The figures clearly reveal how the Hirshfeld-I charges differ significantly from the original Hirshfeld charges (corresponding to the first iteration) and how much the charges can change despite a seemingly small $\{\Delta_A^1\}$ in the first iterations. Figure 1 also clearly shows that convergence obtains quite rapidly.

All Hirshfeld-I charges are contained within the interval $[-2, +2]$ for the present molecular set. Charges below -1 are quite rare. High positive charges do appear, especially for carbon atoms. Analysis of the data reveals that these cases almost exclusively concern carbon atoms with several elec-

TABLE I. Correlation coefficients (R^2 in %) between different charge sets for the neutral molecules.

	CHELPG	MKS	NPA	Mulliken	Hirshfeld	Hirshfeld-I
CHELPG	100	95	76	12	76	90
MKS	95	100	79	18	72	87
NPA	76	79	100	31	79	88
Mulliken	12	18	31	100	19	17
Hirshfeld	76	72	79	19	100	82
Hirshfeld-I	90	87	88	17	82	100

tronegative neighboring atoms. A typical case is the carbon sp^2 atom in the COO moiety. Such highly positive charges are not so uncommon, since they also appear in different other population analysis techniques, especially in electrostatic potential derived charges. The most negatively charged atoms are, somewhat surprisingly, nitrogen atoms, mostly connected to the COO moiety. The nitrogen atom there accumulates a lot of density and compensates in this way for the large loss in density of the above mentioned carbon atoms. Interestingly, this is very reminiscent also of the case of Bader's QCT charges for the same kind of molecules.⁵¹ There, nitrogen atoms were also found to occasionally bear very negative charges, even larger than oxygen atoms. The same is also true for electrostatic potential derived charges. Hydrogen atoms almost exclusively bear positive charges with only very few exceptions. Moreover, these negatively charged hydrogen atoms never bear a charge more negative than -0.04 . Charges on O, F, and Cl are in line with chemical intuition throughout.

Correlation analysis

Although atomic charges are not observables and different schemes may produce diverging results, a correlation analysis between different charge sets is interesting. Table I shows the correlation coefficients (R^2) between atomic charges obtained using the Mulliken approach, natural population analysis (NPA), electrostatic potential derived charges using the CHELPG algorithm, and the MKS approach on the one hand and the Hirshfeld and Hirshfeld-I charge sets on the other.

As was reported previously, the small Hirshfeld charges do not correlate very well with other charges.⁴³ Moreover, they are found to underestimate charges.²⁷ For a more in depth analysis of different charge schemes for the present molecular set, excluding the Hirshfeld-I scheme, the reader is

TABLE II. Regression equations between CHELPG (x) and Hirshfeld-I charges (y) and correlation coefficients (R^2 in %) for different elements and all elements together.

Element	Regression equation	R^2
H	$y=0.81x+0.04$	86
C	$y=0.99x-0.03$	84
N	$y=0.92x-0.02$	80
O	$y=0.97x-0.06$	79
F	$y=0.25x-0.21$	9
Cl	$y=0.85x+0.01$	80
All atoms	$y=0.98x$	90

referred to Ref. 43. Introduction of the iterative scheme gives Hirshfeld-I charges that do show a relevant linear correlation with the CHELPG charges and to slightly lesser extent with NPA and MKS. This is a quite unexpected feature since these ESP derived charges, in fact, do not result in a direct way from atomic densities, but are fitted to reproduce the ESP computed on a molecular surface. More importantly, as will be shown below, the slope of the regression line is quite close to 1. It is worth mentioning that the Hirshfeld-I charges correlate significantly better with ESP derived charges and NPA than the Hirshfeld charges; the correlation of Hirshfeld and Hirshfeld-I with Mulliken charges is comparable, but very poor. The manifest lack of correlation with Mulliken charges may be due to the basis set used, which included diffuse functions on both heavy atoms and hydrogen atoms. These are known to pose computational problems for the Mulliken analysis.⁹ Excluding the Mulliken charges, the Hirshfeld-I charges correlate less with the Hirshfeld charges than any of the other charge schemes. The NPA analysis also correlates favorably with the Hirshfeld-I charges. Again, the slope is relatively close to one (0.93) with zero intercept. In fact, except for the Mulliken scheme, all non-Hirshfeld related schemes correlate better with Hirshfeld-I charges than Hirshfeld charges.

In order to get a better insight into the correlation between CHELPG and Hirshfeld-I charges, Table II presents separate regressions between for the different elements present in the molecular set.

Since all molecules have zero charge, the regression line over all atoms automatically goes through the origin. Interestingly, the slope is nearly 1. Figure 2 shows the correlation

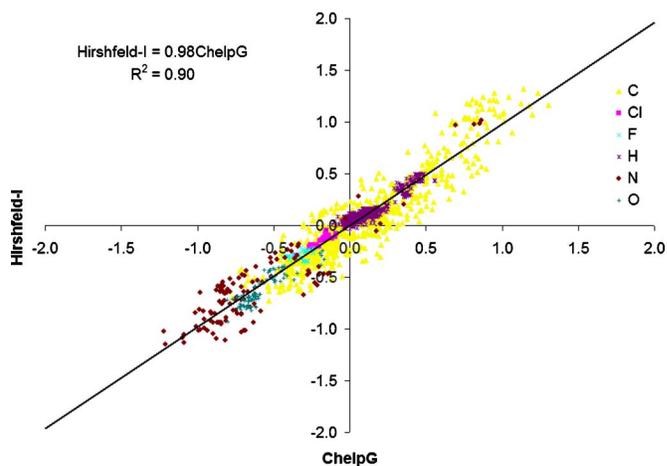


FIG. 2. Correlation between Hirshfeld-I and CHELPG charges.

obtained between the Hirshfeld-I and CHELPG charges for all atoms.

It could be argued that the atoms of the different elements form small islands that behave each as a dispersed point, thereby giving an artificially high correlation. The results for the separate elements show that there is not a very wide spread in correlation coefficients and that within each island the correlation remains noteworthy. This is also manifested in Fig. 2.

The set of fluorine charges is clearly an exception concerning the correlation. This highly electronegative and hard element, however, shows a very small spread in charges, so the regression line in that case bears little relevance. Furthermore, the charges differ on average only 0.03 a.u. with the Hirshfeld-I charges in most cases slightly less negative than the CHELPG based ones. It is worth noting that the Hirshfeld-I charges are significantly more negative than the Hirshfeld ones for fluorine.

Performance of the Hirshfeld-I scheme versus the original scheme

The aim of the present paper was to establish whether the Hirshfeld-I algorithm indeed solves the problems encountered with the original Hirshfeld scheme.

As has been shown above, the Hirshfeld-I scheme indeed produces charges that are substantially larger than the original Hirshfeld charges. The relevant regression equation between both procedures is

$$q_{\alpha}^{\text{Hirshfeld-I}} = 3.09q_{\alpha}^{\text{Hirshfeld}}. \quad (20)$$

In some cases, most notably when the COO group is present, charges can even grow unexpectedly large. In that sense they are reminiscent of previous observations with Bader's QCT charges.⁵¹ Note also that the correlation for the regression [Eq. (20)] (see Table I) is not particularly good (82%).

Secondly, within the iterative scheme the charges are independent of the initial choice for the promolecular density. To demonstrate this, we repeated the Hirshfeld-I algorithm for all 168 molecules, starting from three separate sets of initial atomic populations, and confirmed that the same results were always obtained. Specifically, after considering the classical Hirshfeld choice $\{N_A^0 = Z_A\}$ as described above, the iterative procedure was also initiated from Mulliken populations. Since these populations correlate quite poorly with the Hirshfeld-I charges obtained from $\{N_A^0 = Z_A\}$, this can be used as a test to see whether multiple solutions could exist for the Hirshfeld-I charges. The calculations revealed that for all molecules the same atomic charges are obtained irrespective of the starting promolecular density. As a second test, the iterative process was started from a set of populations where the Hirshfeld-I populations were reversed with respect to the $\{Z_A\}$. In other words, the iterative process started from populations given by $\{N_A^0 = 2Z_A - N_A^{\text{conv}}\}$. This is a very thorough test, since the internal polarization of the molecule is completely reversed. Again, the procedure converged to the same set of charges as the one produced from the other starting populations. This is very gratifying as this means that the molecular electron density determines the populations in the

TABLE III. Hirshfeld, Hirshfeld-I, and CHELPG charges for the symmetry unique atoms in $(\text{LiNH}_3)^+$.

	Hirshfeld	Hirshfeld-I	CHELPG
Li	0.751	1.009	0.939
N	-0.203	-1.278	-1.278
H	0.151	0.423	0.446

AIM and that this is a unique solution. Another example is LiF. As was shown above, the charges in LiF in the Hirshfeld scheme depend quite strongly on the promolecule chosen. In the Hirshfeld-I scheme, LiF always results with a charge of 0.93 in absolute value, irrespective of the promolecular density used. Davidson and Chakravorty pointed out the arbitrary nature of the Hirshfeld approach. One of the arbitrary choices was the fact that all density differences refer to neutral atoms. They tested the robustness of the method by computing a set of atomic charges for N_2 but changing the reference atomic densities. When using N^+N^- as a reference promolecule charges of 0.54 were found (0.53 in the present level of calculation), illustrating the arbitrariness of the choice of the promolecule. The more classical N^0N^0 promolecule calculations gave zero Hirshfeld charges. When starting N_2 charge calculations from N^+N^- as a reference promolecule the Hirshfeld-I algorithm quickly converges to zero charges, as chemically expected. The fact that there appears to be a unique minimum for the Hirshfeld-I charges is due to the fact that the information loss per atom for the Hirshfeld AIM is a strictly convex functional⁵² and minimization of a strictly convex functional with respect to a set of linear constraints always yields a unique minimum. These linear constraints mentioned of course reflect the requirement that the AIM density normalizes to the same as the isolated atom density.

The Hirshfeld-I charges for charged molecules do not violate the requirement that the molecular electron density normalization is equal to the normalization of the promolecular density. As an illustration, Table III reports charges obtained for $(\text{LiNH}_3)^+$. The molecule was optimized in the C_{3v} geometry on the RHF/6-311++G** level. Charges were obtained using $\{Z_A\}$ as starting points, so the first iteration gives the Hirshfeld charges. Convergence of the Hirshfeld-I charges was obtained after 20 iterations.

Again it is found that the charges are quite different between both schemes. It is clear how in the original Hirshfeld scheme, the charges are much smaller, in agreement with what is expected based on Ayers' maximal similarity argument.¹⁷ Also it is found again that the Hirshfeld-I charges are in better agreement with CHELPG charges. Furthermore, the same charges result for the Hirshfeld-I scheme when starting from different atomic populations. The quite negative charge on the nitrogen atom is certainly surprising in a cationic complex. This could possibly signal an important negative Fukui function on the nitrogen atom and may indicate a local reduction of an atom under a global oxidation process, the counterpart of the local oxidation under reduction predicted by Ayers.⁵³ The study of such effects via Hirshfeld-I will be reported elsewhere.

As an extra test of the performance of the Hirshfeld-I scheme for the calculation of atomic charges in charged systems, we also performed Hirshfeld-I calculations for the positive and negative ions derived from all molecules in the test set. For each of the 168 molecules the anion and cation were computed in the doublet spin state restraining the geometry to the same as the one used for the neutral molecule. As a first test to examine the applicability of the Hirshfeld-I scheme for such a large set of molecular ions, it was investigated whether still the same set of atomic charges is obtained when starting the iterations from different starting points. A first set of atomic starting populations was obtained using the regular Hirshfeld starting point. This means that for each atom N_A^0 is chosen in such a way that $\{N_A^0 = Z_A\}$. Note that, although this is a viable mathematical approach, the populations do not sum to the number of electrons in the molecule, which again hampers the link to information theory. On the other hand, already after the first iteration, the populations automatically sum the correct number of electrons. In a second set, populations were chosen as obtained from a Mulliken population analysis of the ionic density functions.

For the cations, as in the case of the neutral atoms, the two different starting points were found to result in the same set of atomic populations after convergence. This again illustrates that a unique set of atomic charges is obtained, although the number of iterations required from either starting point is higher than for the neutral molecules (roughly 40 iterations compared to 20).

For several molecular anions the SCF procedure was found to be quite problematic. Where it succeeded, in many cases the presence of diffuse basis functions caused the Mulliken analysis to yield completely awkward results, including extremely large charges. This is a well-known fact.⁹ It was found that the CHELPG charges also were out of line with chemical experience, showing, for example, carbon atoms with atomic charges of +4 and beyond. Therefore, all SCF calculations were repeated using the 6-311G** basis set which solved most of these problems. Only four molecules did not iterate to convergence in the SCF procedure and as a consequence were deleted from the test set leaving 164 molecules for further analysis. As for the Hirshfeld-I analysis, different starting points for the iterative procedure yielded exactly the same final set of atomic populations. This again shows the uniqueness of the iterative solution.

It is worth noting how the atomic charges from the Hirshfeld-I scheme for the cations do not correlate very well with those obtained from the classical Hirshfeld approach. The correlation coefficient amounts to only 70%. On the other hand, the correlation between the Hirshfeld-I charges for the cations correlates quite well with those obtained from the CHELPG algorithm. The correlation coefficient amounts to 88%. Moreover, the regression line has as equation $\text{HIRSHFELD-I} = 1.05 * \text{CHELPG} - 0.00$. The regression has a nearly unit slope and a very small intercept, reminiscent of the case of the neutral systems, as was shown in Table II. For the case of the anions, there is less good agreement. The correlation coefficient R^2 has dropped to 67% and the regression line is now $\text{HIRSHFELD-I} = 0.69 * \text{CHELPG} - 0.02$. The

range of charges in CHELPG is bigger than in the Hirshfeld-I case, with positive charges up to 1.69 and negative charges as negative as -1.43 whereas the Hirshfeld-I charges are contained in the interval $[-1.23, 1.28]$. It is not clear exactly why the correlation with the electrostatic derived charges is remarkably poorer for the anions. The question can, however, be raised to what extent the surface used to compute the electrostatic potential on for fitting the electrostatic charges should be dependent on the total charge of the system. As the anions are quite diffuse, it might be that using the same surface for the anions compared to the neutral and cationic systems is not fully appropriate. Again there is a poor correlation with the original Hirshfeld charges with $R^2 = 68\%$. Somewhat surprisingly, a good correlation is found with Mulliken atomic charges. The correlation coefficient amounts to 84%. However, as the correlation coefficient with the neutral molecules is so extremely poor, it is not allowed to draw a general conclusion from this finding.

The results of the above tests very clearly show how not only the Hirshfeld-I scheme can be applied to ionic systems as well, they are also applicable to non-closed-shell molecules.

Finally, from the standpoint of information entropy, the Hirshfeld-I atomic fragments are truly those that are most similar to the fractionally occupied isolated atomic densities. In the Hirshfeld-I treatment, there is only deformation of the atomic densities (the entropy of mixing term is zero) and the entropy of the atoms is always positive and is an extensive function of the number of electrons in the atom. This means that all of the mathematical prerequisites for Eq. (9) to be an information measure are strictly fulfilled. If one wishes to dissect a molecule in AIM, information entropy tells us that the Hirshfeld-I weighting functions are the ones that make the AIM as similar as possible to the isolated atoms provided that both have the same electronic population. In this sense the Hirshfeld-type AIM are the AIM that minimize the information lost relative to the separated-atom limit. It is also worth noting that the properties of Hirshfeld-I AIM exhibit maximum transferability from the fractionally charged atomic reference states (just as the properties of Hirshfeld AIM exhibit maximum transferability from the neutral atoms).

Just as in the original Hirshfeld scheme, in principle one should consider all possible atomic states in Eq. (19). This requires lengthy evaluations of the information entropy and the best set of states is then to be associated with the one that makes the information distance minimal. This is scarcely feasible, but it is worth noting that the effect of using different states in the original Hirshfeld scheme was shown to have a little effect on the final result.⁴⁴

CONCLUSIONS

A new Hirshfeld scheme, denoted Hirshfeld-I, has been proposed and implemented. This procedure uses an iterative approach to identify atoms in molecules. The Hirshfeld-I scheme allows calculations of atomic populations, adhering strictly to information theory. Furthermore, several other problems associated with the original Hirshfeld scheme are

solved. The Hirshfeld-I charges are much larger than the results from the original Hirshfeld scheme. In addition, the Hirshfeld-I charges correlate well with CHELPG electrostatic potential derived charges. The iterative scheme eliminates arbitrariness in the choice of the promolecule, so the atomic populations are determined solely by the molecular electronic density. Charged molecules are also tractable now. Open shell species can also be treated without problem. In agreement with Parr *et al.*,³ once the reference states for the atoms have been chosen, the present computational scheme provides a universal definition of AIM free of all atomic and molecular parameters other than density functions.

The present study has shown that in order to compute AIM properties, no net charge transfer should be taking place anymore in the molecule. It is only at that stage one can compute completely trustworthy, nonarbitrary, AIM properties.

ACKNOWLEDGMENTS

One of the authors (P.B.) wishes to thank Ghent University and the Fund for Scientific Research-Flanders (Belgium) for their grants to the Quantum Chemistry group at Ghent University. He also thanks Pratim Chattaraj (India) for interesting discussions on conceptual DFT. Another author (C.V.A.) gratefully acknowledges the University of Antwerp for access to the CALCUA computer cluster. One of the authors (R.C.D.) acknowledges the Ministerio de Ciencia y Tecnología for the grant: BQU2003-07420-C05-01; which has partially sponsored this work and also for a Salvador de Madariaga fellowship reference: PR2004-0547, which has permitted his stay at Ghent University during the period of 2004–2006. One of the authors (P.W.A.) thanks NSERC and the Canada Research Chairs for partial financial support.

¹R. F. W. Bader, *Atoms in Molecules: A Quantum Theory* (Clarendon, Oxford, 1990).

²R. F. W. Bader, *Chem. Rev.* (Washington, D.C.) **91**, 893 (1991).

³R. G. Parr, P. W. Ayers, and R. F. Nalewajski, *J. Phys. Chem. A* **109**, 3957 (2005).

⁴R. S. Mulliken, *J. Chem. Phys.* **23**, 1833 (1955).

⁵R. S. Mulliken, *J. Chem. Phys.* **23**, 1841 (1955).

⁶R. S. Mulliken, *J. Chem. Phys.* **23**, 2338 (1955).

⁷R. S. Mulliken, *J. Chem. Phys.* **23**, 2343 (1955).

⁸R. Carbó-Dorca and P. Bultinck, *J. Math. Chem.* **36**, 201 (2004).

⁹F. Jensen, *Introduction to Computational Chemistry* (Wiley, Chichester, 1999).

¹⁰A. E. Clark and E. R. Davidson, *Int. J. Quantum Chem.* **93**, 384 (2003).

¹¹R. Carbó-Dorca, *J. Mol. Struct.: THEOCHEM* **537**, 41 (2001).

¹²R. Carbó-Dorca, *J. Math. Chem.* **32**, 201 (2002).

¹³R. Carbó-Dorca, *Int. J. Quantum Chem.* **91**, 607 (2003).

¹⁴F. L. Hirshfeld, *Theor. Chim. Acta* **44**, 129 (1977).

¹⁵R. G. Parr and W. Yang, *Density Functional Theory of Atoms and Molecules* (Oxford University Press, New York, 1989).

¹⁶P. Geerlings, F. De Proft, and W. Langenaeker, *Chem. Rev.* (Washington, D.C.) **103**, 1793 (2003).

¹⁷P. W. Ayers, *J. Chem. Phys.* **113**, 10886 (2000).

¹⁸R. F. Nalewajski and R. G. Parr, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 8879 (2000).

¹⁹R. F. Nalewajski, E. Switka, and A. Michalak, *Int. J. Quantum Chem.* **87**, 198 (2002).

²⁰R. F. Nalewajski, *Chem. Phys. Lett.* **372**, 28 (2003).

²¹R. F. Nalewajski and E. Broniatowska, *Int. J. Quantum Chem.* **101**, 349 (2005).

²²P. W. Ayers, *Theor. Chem. Acc.* **115**, 301 (2006).

²³F. M. Bickelhaupt, N. J. R. van Eikema Hommes, C. Fonseca Guerra, and E. J. Baerends, *Organometallics* **15**, 2923 (1996).

²⁴C. Fonseca Guerra, J.-W. Handgraaf, E. J. Baerends, and F. M. Bickelhaupt, *J. Comput. Chem.* **25**, 189 (2003).

²⁵I. Mayer and P. Salvador, *Chem. Phys. Lett.* **383**, 368 (2004).

²⁶A. D. Becke, *J. Chem. Phys.* **88**, 2547 (1988).

²⁷E. R. Davidson and S. Chakravorty, *Theor. Chim. Acta* **83**, 319 (1992).

²⁸P. W. Ayers, R. C. Morrison, and R. K. Roy, *J. Chem. Phys.* **116**, 8731 (2002).

²⁹J. R. Pierce, *An Introduction to Information Theory: Symbols, Signals and Noise* (Dover, New York, 1980).

³⁰S. Kullback, *Information Theory and Statistics* (Wiley, New York, 1959).

³¹R. G. Parr and L. J. Bartolotti, *J. Phys. Chem.* **87**, 2810 (1983).

³²R. Carbó-Dorca, *J. Math. Chem.* **23**, 365 (1998).

³³P. W. Ayers, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 1959 (2000).

³⁴P. Bultinck and R. Carbó-Dorca, *J. Math. Chem.* **36**, 201 (2004).

³⁵F. De Proft, P. W. Ayers, K. D. Sen, and P. Geerlings, *J. Chem. Phys.* **120**, 9969 (2004).

³⁶P. W. Ayers, *Phys. Rev. A* **71**, 062506 (2005).

³⁷J. P. Perdew, R. G. Parr, M. Levy, and J. L. Balduz, Jr., *Phys. Rev. Lett.* **49**, 1691 (1982).

³⁸W. Yang, Y. Zhang, and P. W. Ayers, *Phys. Rev. Lett.* **84**, 84 (2000).

³⁹P. W. Ayers, *J. Math. Chem.* (in press).

⁴⁰In order to honor the work of Hirshfeld, the new scheme retains his name and is simply extended with the “-I” extension to emphasize the iterative character of the new algorithm.

⁴¹See EPAPS Document No. E-JCPSA6-126-309713 for the molecular test set. This document can be reached via a direct link in the online article’s HTML reference section or via the EPAPS homepage (<http://www.aip.org/epaps/epaps.html>).

⁴²P. Bultinck, W. Langenaeker, P. Lahorte, F. De Proft, P. Geerlings, M. Waroquier, and J. P. Tollenaere, *J. Phys. Chem. A* **106**, 7887 (2002).

⁴³P. Bultinck, W. Langenaeker, P. Lahorte, F. De Proft, P. Geerlings, C. Van Alsenoy, and J. P. Tollenaere, *J. Phys. Chem. A* **106**, 7895 (2002).

⁴⁴B. Rousseau, A. Peeters, and C. Van Alsenoy, *Chem. Phys. Lett.* **324**, 189 (2000).

⁴⁵B. Roos, C. Salez, A. Veillard, and E. Clementi, ATOMSCF, IBM Technical Report RJ 518, IBM Research Laboratory, San Jose, CA (1968).

⁴⁶M. J. Frisch, G. W. Trucks, H. B. Schlegel *et al.*, GAUSSIAN03, Revision B.05, Gaussian, Inc., Wallingford, CT, 2004.

⁴⁷A. E. Reed, L. A. Curtiss, and F. Weinhold, *Chem. Rev.* (Washington, D.C.) **88**, 899 (1988).

⁴⁸M. M. Francl, C. Carey, L. E. Chilian, and D. M. Gange, *J. Comput. Chem.* **17**, 367 (1996).

⁴⁹C. M. Breneman and K. B. Wiberg, *J. Comput. Chem.* **11**, 361 (1990).

⁵⁰U. C. Singh and P. A. Kollman, *J. Comput. Chem.* **5**, 129 (1984).

⁵¹P. Bultinck, R. Vanholme, P. Popelier, F. De Proft, and P. Geerlings, *J. Phys. Chem. A* **108**, 10359 (2004).

⁵²P. W. Ayers and P. Bultinck (unpublished).

⁵³P. W. Ayers, *Phys. Chem. Chem. Phys.* **8**, 3387 (2006).