

# The Atomic Partial Charges Arboretum: Trying to See the Forest for the Trees

Minsik Cho,<sup>[a, b]</sup> Nitai Sylvetsky,<sup>[a]</sup> Sarah Eshafi,<sup>[a, c]</sup> Golokesh Santra,<sup>[a]</sup> Irena Efremenko,<sup>[a]</sup> and Jan M. L. Martin<sup>\*[a]</sup>

Atomic partial charges are among the most commonly used interpretive tools in quantum chemistry. Dozens of different ‘population analyses’ are in use, which are best seen as proxies (indirect gauges) rather than measurements of a ‘general ionicity’. For the GMTKN55 benchmark of nearly 2,500 main-group molecules, which span a broad swathe of chemical space, some two dozen different charge distributions were evaluated at the PBE0 level near the 1-particle basis set limit. The correlation matrix between the different charge distributions exhibits a block structure; blocking is, broadly speaking, by charge distribution class. A principal component analysis on the entire dataset suggests that nearly all variation can be accounted for by just two ‘principal components of ionicity’: one has all the distributions going in sync, while the second

corresponds mainly to Bader QTAIM vs. all others. A weaker third component corresponds to electrostatic charge models in opposition to the orbital-based ones. The single charge distributions that have the greatest statistical similarity to the first principal component are iterated Hirshfeld (Hirshfeld-I) and a minimal-basis projected modification of Bickelhaupt charges. If three individual variables, rather than three principal components, are to be identified that contain most of the information in the whole dataset, one representative for each of the three classes of Corminboeuf et al. is needed: one based on partitioning of the density (such as QTAIM), a second based on orbital partitioning (such as NPA), and a third based on the molecular electrostatic potential (such as HLY or CHELPG).

## 1. Introduction

Atomic partial charges  $q_A$  are a central concept in general chemistry. Unfortunately, they do not correspond to a single, well-defined, quantum mechanical observable. The very notion of atomic partial charges, however, does imply that “atoms in molecules” (in the broad sense of the word, not the narrow one of Bader/QTAIM<sup>[1]</sup> charges) are meaningful concepts.

Parr, Ayers, and Nalewajski<sup>[2]</sup> consider partial charges to be *noumena*: The term *noumenon*, originally coined by Plato from the Greek word *noös* (knowledge, cognition), is defined by the Oxford English Dictionary as “an object knowable by the mind or intellect, not by the senses; specifically (in Kantian philosophy) an object of purely intellectual intuition”. Frenking and

Krapp<sup>[3]</sup> speak of “unicorns”: “my[th]ical animal[s] whose appearance is known to everybody although nobody has ever seen one”.

In reaction, Matta and Bader<sup>[4,5]</sup> strongly took exception. They point out that Bader’s QTAIM charges, in particular, can be seen as expectation values of well-defined operators, namely: Heaviside “step” functions that are unity within one set of zero-flux surfaces and zero outside. The fact that QTAIM charges often are at odds with “chemical intuition” (or at least chemical received wisdom) leaves the reader even more perplexed, and more in need of a guide.<sup>[6]</sup>

Perhaps a third term is more apropos here, namely the statistical one “proxy variable”. A proxy variable is a (fairly easily) measurable quantity that acts as an indirect “proxy” for a deeper concept that eludes direct measurement. Examples from economics are proxies for the standard of living such as GDP per capita or PPP (purchasing power parity). Other examples, from psychology, are IQ scores when used as proxies for general intelligence, or (relatedly) SAT scores as predictors for scholastic success. (As a hoary quip goes, the only thing standardized cognitive tests directly measure is performance on those tests.) In that sense, partial charges are “proxy variables” for a deeper chemical concept that Meister and Schwarz<sup>[7]</sup> have termed “[molecular] ionicity”.


Corminboeuf and coworkers<sup>[8]</sup> as well as Cramer and Truhlar<sup>[9]</sup> have attempted taxonomies of breeds within this particular species of “chemical unicorn”. Corminboeuf et al. distinguish three broad classes of computed partial charges:


(1) fitting to an observable quantity such as the molecular electrostatic potential on the van der Waals surface, such as done in CHELP<sup>[10]</sup> (charges from electrostatic potentials),

[a] M. Cho, N. Sylvetsky, S. Eshafi, G. Santra, Dr. I. Efremenko, Prof. Dr. J. M. L. Martin  
Department of Organic Chemistry  
Weizmann Institute of Science  
76100 Rehovot (Israel)  
E-mail: gdrshom@weizmann.ac.il

[b] M. Cho  
Present Address: Department of Chemistry  
Brown University  
Providence, Rhode Island 02912 (USA)

[c] S. Eshafi  
Present Address: Integrated Science Program  
McMaster University  
Hamilton, Ontario L8S 4 M1 (Canada)

 Supporting information for this article is available on the WWW under <https://doi.org/10.1002/cphc.202000040>

 © 2020 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

- CHELPG<sup>[11]</sup> (ditto using a grid), Merz-Kollman,<sup>[12]</sup> and HLY<sup>[13]</sup> (Hu–Lu–Yang) charges;
- (2) partitioning in terms of atomic orbitals, such as done in Mulliken charges<sup>[14]</sup> and their variants, Natural Population Analysis (NPA),<sup>[15]</sup> and Intrinsic Bond Orbitals,<sup>[16]</sup> (IBO, also known<sup>[17]</sup> as Quasi-Atomic Natural Bond Orbitals<sup>[18]</sup>);
  - (3) atoms-in-molecules partitioning of the electron density, either into disjoint QTAIM<sup>[1]</sup> or Voronoi<sup>[19]</sup> cells, or into overlapping “stockholder domains” (such as the original Hirshfeld population analysis<sup>[20,21]</sup> and its manifold offspring, e.g.,<sup>[20–26]</sup>)<sup>1</sup>.

In a separate category they place “experimental” charges, which are measured (or rather, for which trends are inferred) from NMR or EXAFS chemical shifts, dipole moments, and the like.

Cramer and Truhlar<sup>[9]</sup> proposed a slightly different taxonomy which, with variations, we will adhere to in the present work.

- Class I charges are derived from experimentally measurable properties, e.g., from observed deformation densities (as in the work of the late Philip Coppens<sup>[27,28]</sup>), from the electronegativity equalization principle, or from, e.g., dipole moments of diatomic and small (usually highly symmetric) polyatomic molecules.
- Class II charges are extracted from the molecular orbitals (Mulliken, NPA,...) or the electron density (Bader QTAIM, Hirshfeld,...).
- Class III charges are extracted from the wave function or electron density by fitting a physical observable (e.g., the electrostatic potential) derived from it.
- Class IV charges, such as the CM5 model introduced by Truhlar and coworkers,<sup>[29]</sup> are based on semiempirical adjustment of a well-defined Class II or Class III model to better reproduce one or more physical observables (e.g. the dipole moment).

(We note in passing that the legal fiction known as “formal charges” might be termed Class 0.)

A number of review articles have been devoted to partial charges, such as Wiberg and Rablen,<sup>[30,31]</sup> Bachrach,<sup>[32]</sup> Cramer and Truhlar,<sup>[9]</sup> and most recently Ayers and coworkers<sup>[26]</sup> (which focuses principally on Hirshfeld-type approaches from an information theoretical point of view).

Normally, “observability” in the quantum mechanical sense of the word implies the existence of an operator for which one can obtain an expectation value. Cioslowski and Surjan<sup>[33]</sup> propose a weaker, more general definition where a uniquely defined computational protocol that extracts uniquely defined numerical values from the wave function, which have well-defined infinite basis set limits, is said to satisfy a “generalized observability criterion”. (It should be pointed out that the hoary Mulliken population analysis,<sup>[14]</sup> for instance, does not satisfy said criterion on account of its pathological basis set dependence – nor do modified Mulliken analyses like Löwdin,<sup>[34]</sup> Ros-Schuit,<sup>[35]</sup> or Bickelhaupt.<sup>[36]</sup> However, minimal basis set projections<sup>[37]</sup> – denoted here as MBS-Mulliken,<sup>[14,37]</sup> MBS-Bick-

elhaupt (Ref. [36] and vide infra), etc. – do satisfy the Cioslowski-Surjan criterion.<sup>[36]</sup>)

So, there are now several dozen charge distributions “on the market”. But how distinct are they really? How different a picture do they really paint? Answers to this question might be found in feature reduction techniques – most notably from the best-known such technique, PCA (principal component analysis).

Investigating this question requires a dataset of molecules large and chemically diverse enough that any conclusions reached cannot (easily) be dismissed as sample selection artefacts. The almost 2,500 unique species in the GMTKN55 benchmark<sup>[38]</sup> (general main-group thermochemistry, kinetics, and noncovalent interactions, 55 problem subsets) for density functional methods are one useful starting point. (See Refs.<sup>[38–42]</sup> for example applications to DFT.)

We shall show that there is indeed a “method in the madness” and that almost all of the data variation can be represented by two to three principal components.

## Computational Methods

### Population Analyses Considered

We considered the following population analyses as representatives of the different classes in the Cramer and Truhlar<sup>[9]</sup> taxonomy (where we shall introduce some subclasses below):

(1) for Class I charges, the EEQ (electronegativity equalization principle) charges as implemented in the DFTD4 program of the Grimme group (Section II.A of Ref. [43]);

(2) Class II is subdivided here into:

- Class IIa: from partitioning orbitals. (NB: this effectively corresponds to Corminboeuf’s Class 2.) We can subdivide further into:

- IIa1: Mulliken<sup>[14]</sup> and variants (Löwdin,<sup>[34]</sup> Bickelhaupt,<sup>[36]</sup>...) applied in the full basis set. We considered these but discarded them on account of their pathological basis set dependence.

- IIa1 M: projection of the MOs onto a minimal basis set<sup>[37]</sup> followed by the above techniques. This eliminates the basis set hypersensitivity and leads to values that satisfy the extended Cioslowski-Surjan observability criterion.<sup>[33]</sup> We here consider MBS-Mulliken<sup>[37]</sup> and (by straightforward extension) MBS-Bickelhaupt. The latter differs from MBS-Mulliken in that off-diagonal overlap populations are taken into account by diagonal-overlap weighted average instead of simple average:

$$Q'_{ii,Bickelhaupt} = q_{ii} + \sum_{i \neq j} \left( \frac{q_{ij}}{q_{ii} + q_{jj}} \right) (q_{ij} + q_{ji})$$

$$Q'_{ii,Mulliken} = q_{ii} + \frac{1}{2} \sum_{i \neq j} (q_{ij} + q_{ji})$$

- IIa2: techniques based on some form of natural or intrinsic atomic orbital: here we consider Weinhold’s NPA (natural population analysis<sup>[15]</sup>) and the IAO (intrinsic atomic orbitals) of Knizia,<sup>[16]</sup> which are functionally equivalent<sup>[17]</sup> to Ruedenberg’s QAOs (quasi-atomic orbitals<sup>[18,44]</sup>).

- Class IIb: from partitioning the electron density into atomic domains & integrating over those. (This effectively corresponds

<sup>1</sup>F. L. Hirshfeld himself used the term “stockholder population analysis” for what is today commonly known by his last name.

to Class 3 in Corminboeuf's taxonomy.) We can further distinguish between fuzzy and discrete domains.

- llb1: fuzzy domains: we primarily focus on variants of the "stockholder" population analysis of Hirshfeld,<sup>[20,21]</sup> specifically:
- his original method (in which the deformation density is partitioned based on the promolecular density);
- the iterated Hirshfeld<sup>[22]</sup> (Hirshfeld-I) method, where the proatom densities that define the promolecular density are iteratively interpolated between ionization states;
- the Iterative Stockholder Approach<sup>[23]</sup> (ISA) in which proatoms are avoided through angular integration around an atom;
- MBIS (minimal basis iterative stockholder<sup>[25]</sup>), which mitigates certain issues with proatoms that are unbound in vacuo (e.g., N<sup>-</sup>, N<sup>2-</sup>, N<sup>3-</sup>);
- Finally, the DDEC6 (density derived electrostatic and chemical approach, version 6) of Manz<sup>[24](a-f)</sup> which is designed for computational resilience in problematic systems and contains many adjustments to ensure proper behavior in solids<sup>2</sup>. This approach has been further developed for additional properties (e.g., bond orders,<sup>[24](c)</sup> dispersion coefficients<sup>[24](e)</sup>) that are beyond the scope of the present paper.

Stockholder-type charges have been rationalized based on information theory<sup>[2,45]</sup> and were the subject of a recent review.<sup>[26]</sup> (We considered Salvador's topological fuzzy Voronoi charges<sup>[46]</sup> for a representative sample consisting of the W4-17 benchmark,<sup>[47]</sup> and found said charges to be very similar in practice to Bader QTAIM charges,  $R^2 = 0.97$ .)

- llb2: disjoint/discrete domains: the most important representative of this are QTAIM (quantum theory of atoms in molecules<sup>[1]</sup>), a.k.a. "Bader charges", in which zero-flux surfaces are used to partition the electron density into atomic cells within which numerical integration takes place.

Maslen and Spackman<sup>[48,49]</sup> proposed a modification inspired by Hirshfeld charges: while zero-flux surfaces are determined as in QTAIM, the integration is performed instead on the deformation density (i.e.  $\rho_{\text{molecule}} - \rho_{\text{promolecule}}$ ). This mitigates the tendency of QTAIM to amplify charges: zero-flux surfaces tend to run far from high electronegative atoms like O and F, at the expense of less electronegative atoms.

The Voronoi Deformation Densities (VDD) of Fonseca-Guerra et al.,<sup>[19]</sup> go one step further, in that they partition the deformation density through simple Voronoi tessellation: any point in space closer to a given atom than to the others is assigned to that atom.

(3) Class III can be further subdivided into:

- Class IIIa: based on the electrostatic potential. (This corresponds to Class 1 in Corminboeuf's taxonomy.) The original CHELP of Chirlian and Franci<sup>[10]</sup> was modified by Breneman and Wiberg (CHELPG)<sup>[11]</sup> to improve rotational invariance. Merz-Kollman (a.k.a. Merz-Singh-Kollman<sup>[12,50]</sup>), Hu-Lu-Yang (HLY<sup>[13]</sup>), and RESP (restrained electrostatic potential<sup>[51]</sup>) all have the same physical basis but merely represent different integration grid specifics for the ESP. As HLY appears to be the most stable numerically and is available for all elements represented in our dataset, we have focused primarily on HLY.

- Class IIIb: based on other electrostatic properties. Here we consider Cioslowski's APT (atomic polar tensor<sup>[52]</sup>) charges, based on the trace of the dipole moment derivatives.

(4) under Class IV we consider CM5 (charge model 5),<sup>[29]</sup> an empirical adjustment of Hirshfeld to better reproduce molecular dipole moments, and ADCH:<sup>[53]</sup> (atomic dipole corrected Hirshfeld), an adjustment to ensure reproduction of molecular dipole moment obtained at the same calculation level.

In addition, we consider ACP, atomic charge partitioning<sup>[54]</sup> and especially its iterative variant i-ACP,<sup>[55]</sup> which arguably can be pigeonholed both under Class IV (on account of their adjustment for better electrical properties) and under llb1 (as they bear an obvious kinship to MBIS in representing proatoms by Slater-type functions).

## Other Computational Details

All electronic structure calculations were carried out using the PBE0 functional<sup>[56,57]</sup> with the def2-TZVPP basis set<sup>[58]</sup> using the Gaussian,<sup>[59]</sup> MOLPRO,<sup>[60]</sup> and Q-CHEM<sup>[61]</sup> electronic structure program systems. SG-3<sup>[62]</sup> or equivalent integration grids (e.g. Grid = UltraFine in Gaussian) were employed.

The reference geometries were obtained from the online supporting information to the GMTKN55 paper and used without further optimization. Species with trivial charge distributions (e.g. atoms, homonuclear diatomics, tetrahedral P<sub>4</sub>,...) were removed from the dataset, as were the handful of duplicate species, which left 2,125 unique molecules. Details of the electronic structure and postprocessing software used to generate each specific set of partial charges are given in Table 1. The various aspects of this process were automated using a collection of scripts developed in-house. For easy data manipulation, Python scripts using numpy and pandas libraries were written. Duplicates were identified from the input geometries (by comparison of the rotational constants). Holes in the data were generally from technical limitations dealing with molecules that contain heavy elements for which some parameter was lacking in the available implementation.

Principal component analysis was initially performed upon 25 different population analysis methods. Our initial analysis identified some near-redundancies (see below), leading to some winnowing.

Upon completing the analysis, its statistical stability was tested by adding elements that were randomly chosen from the dataset to provide noise. The structure of principal components along with the explained variance values was compared between the untouched and adulterated data to assess how robust the implied structure markers within the data are.

## 2. Results and Discussion

The full correlation matrix of all variables, as well as the full partial charges dataset, are given in Microsoft Excel format in the Supporting Information.

One thing becomes immediately clear if the matrix is blocked by charge distribution type: blocks with very high

<sup>2</sup> Among many other things, proatom densities are evaluated in the presence of compensating background charges.

**Table 1.** Details of software used for various population analyses.

Software	Atomic Partial Charges
Multiwfn <sup>[63]</sup>	Atomic Dipole Corrected Hirshfeld (ADCH), <sup>[53]</sup> Voronoi Deformation Density (VDD), <sup>[19]</sup> Becke, <sup>[64]</sup> Ros-Schuit, <sup>[35]</sup> CHelpG (CHarges from Electrostatic Potentials on a Grid), <sup>[11]</sup> Merz-Kollman, <sup>[12]</sup> Stout-Politzer, <sup>[65]</sup> Restricted Electrostatic Potential (RESP), <sup>[51]</sup> Bickelhaupt, <sup>[36]</sup>
Gaussian 16 <sup>[59]</sup>	Bader Quantum Theory of Atoms in Molecules (QTAIM) <sup>[11]</sup> Mulliken, <sup>[14]</sup> Hirshfeld, <sup>[20,21]</sup> Iterative Hirshfeld, <sup>[22]</sup> HLY (Hu–Lu–Yang), <sup>[13]</sup> Minimal Basis Set Mulliken, <sup>[37]</sup> Natural Population Analysis (NPA), <sup>[15]</sup> Charge Model 5 (CM5), <sup>[29]</sup> Atomic Polar Tensor Charges (APT) <sup>[52]</sup>
MOLPRO 2019 <sup>[60]</sup>	Intrinsic Bond Orbitals (IBO) <sup>[16]</sup>
Horton 3 <sup>[66]</sup>	Iterative Stockholder Analysis (ISA), <sup>[23]</sup> Minimal Basis Set Iterative Stockholder Analysis (MBIS) <sup>[25]</sup>
ACP <sup>[67]</sup> and I-ACP <sup>[67]</sup>	Adjusted Charge Partitioning (ACP), <sup>[54]</sup> Iterative Adjusted Charge Partitioning (I-ACP) <sup>[55]</sup>
DFTD4 <sup>[43]</sup>	Electronegativity Equilibration Charges (EEQ) <sup>[68]</sup>
DDEC6 <sup>[24]</sup>	DDEC6 (density derived electrostatic and chemical approach, version 6) <sup>[24]</sup>
IBOView <sup>[16]</sup>	Fuzzy Voronoi <sup>[46]</sup> (graphical user interface only)

correlation coefficients within them emerge. Keeping all of these around in the principal component analysis leads to a large number of near-zero eigenvalues.

By way of illustration, we show here the block between the four electrostatic charge variants:

	CHelpG	MK	RESP	HLY
CHelpG	1.000	0.983	0.982	0.945
MK	0.983	1.000	0.995	0.981
RESP	0.982	0.995	1.000	0.975
HLY	0.945	0.981	0.975	1.000

Its eigenvalues and eigenvectors (i.e. the principal components of the correlation matrix) are:

3.9307	0.0554	0.0101	0.0038
0.4974	0.6680	-0.5201	0.1892
0.5036	-0.0025	0.1703	-0.8470
0.5027	0.0725	0.7364	0.4468
0.4962	-0.7406	-0.3976	0.2172

Note that the eigenvalues of a correlation matrix add up to its dimension, *in casu*, four. The largest eigenvector (effectively the arithmetic mean of all four variables) has eigenvalue 3.93, the next largest (corresponding to CHELPG vs. HLY) having eigenvalue 0.055. One naive variable selection technique, discussed by Jolliffe<sup>[69]</sup> in Section 9.3 of his textbook on Principal Component Analysis, is to eliminate the variables that have the largest coefficient in the eigenvectors with the lowest eigenvalues. This ‘backward elimination’ (BE) strategy leads to Merz-Kollman and RESP being eliminated, leaving just CHELPG and HLY. The eigenvectors for the remaining 2×2 correlation matrix are just (CHELPG±HLY)/√2. Whether one chooses to retain CHELPG or HLY for further analysis is basically arbitrary: CHELPG is available in a larger number of codes, while HLY is implemented (in Gaussian) for all elements in the Periodic Table. In the present paper, we have elected to retain HLY.

Between the four iterative Class IIb1 methods, we have the following correlation matrix:

	ISA	Hirshfeld-I	MBIS	DDEC6
ISA	1.0000	0.9595	0.9698	0.9720
Hirshfeld-I	0.9595	1.0000	0.9862	0.9875
MBIS	0.9698	0.9862	1.0000	0.9924
DDEC6	0.9720	0.9875	0.9924	1.0000

Its largest two eigenvalues are 3.934 and 0.045, the former corresponding to all charges moving in tandem and the latter to ISA pulling in the opposite direction from the three others. Within the lower 3-variable block, the top eigenvalue is 2.977, showing that these three charge types on the whole contain very similar information.

The squared correlation matrix for 18 variables has been given in Table 2, reordered to maximize blocking. Some comments concerning it are in order here:

- Note the high correlation ( $R^2=0.96$ ) between VDD and ordinary/original Hirshfeld. Both are based on partitioning the same promolecular density: the key difference is of course that Hirshfeld uses fuzzy boundaries derived from  $\rho_{\text{proatom}}/\rho_{\text{promolecule}}$  while VDD employs discrete Voronoi tessellation. Apparently, the impact of this difference on the statistical similarity of the variables is smaller than one might naively expect.
- In contrast, squared correlations between ordinary Hirshfeld and the various iterative Class IIb1 methods are surprisingly small, around 0.80 except for DDEC6 ( $R^2=0.85$ ). Indeed, ISA has a greater correlation with HLY electrostatic charges ( $R^2=0.87$ ), as does MBIS ( $R^2=0.84$ ); DDEC6’s  $R^2$  with HLY is similar to its correlation with ordinary Hirshfeld.
- The iterative Class IIb1 methods emerge as a block, out of which ISA “only” has  $R^2=0.92$ – $0.94$  with the others, but the remaining three have no  $R^2$  below 0.97 between them. ACP and i-ACP can be seen as peripheral members of the block: in light of the conceptual similarity between MBIS and i-ACP, the relatively low  $R^2[\text{MBIS}, \text{i-ACP}]=0.86$  is somewhat surprising. Rather more surprisingly, this latter sub-block of DDEC6, MBIS, and Hirshfeld-I has  $R^2=0.94$ – $0.95$  with MBSBickelhaupt, that is, Bickelhaupt’s modified Mulliken after minimal basis set projection. (While  $R^2[\text{DDEC6}, \text{MBIS}]=0.99$  and  $R^2[\text{DDEC6}, \text{Hirshfeld-I}]=0.98$  suggest great similarity, these coefficients of determination would likely have



**Table 2.** The squared correlation matrix for a subset of all variables. For clarity, the variables have been ordered to maximize blocking. The color gradient runs from white for  $R^2 \leq 0.80$  to deep blue for  $R^2 \geq 0.98$ .

	VDD	Hirshfeld	HLY	ISA	DDEC6	MBIS	Hirshfeld-I	MBS-Bickelhaupt	NPA	IBO	MBS-Mulliken	i-ACP	ACP	CM5	EEQ	QTAIM	APT	ADCH
VDD	1.00	0.96	0.64	0.78	0.81	0.77	0.78	0.78	0.74	0.78	0.70	0.78	0.78	0.76	0.73	0.64	0.68	0.68
Hirshfeld	0.96	1.00	0.67	0.79	0.85	0.80	0.80	0.79	0.78	0.82	0.76	0.75	0.80	0.78	0.73	0.61	0.62	0.72
HLY	0.64	0.67	1.00	0.87	0.84	0.84	0.79	0.76	0.73	0.73	0.72	0.71	0.77	0.73	0.70	0.48	0.51	0.68
ISA	0.78	0.79	0.87	1.00	0.94	0.94	0.92	0.89	0.81	0.80	0.77	0.89	0.85	0.78	0.77	0.69	0.74	0.69
DDEC6	0.81	0.85	0.84	0.94	1.00	0.99	0.98	0.94	0.92	0.91	0.90	0.85	0.89	0.84	0.81	0.66	0.67	0.75
MBIS	0.77	0.80	0.84	0.94	0.99	1.00	0.97	0.95	0.93	0.89	0.90	0.86	0.90	0.84	0.82	0.66	0.66	0.74
Hirshfeld-I	0.78	0.80	0.79	0.92	0.98	0.97	1.00	0.94	0.92	0.89	0.88	0.87	0.86	0.80	0.78	0.73	0.73	0.68
MBSBickelhaupt	0.78	0.79	0.76	0.89	0.94	0.95	0.94	1.00	0.92	0.87	0.90	0.89	0.92	0.87	0.86	0.75	0.71	0.71
NPA	0.74	0.78	0.73	0.81	0.92	0.93	0.92	0.92	1.00	0.96	0.98	0.73	0.85	0.84	0.81	0.60	0.55	0.73
IBO	0.78	0.82	0.73	0.80	0.91	0.89	0.89	0.87	0.96	1.00	0.96	0.71	0.81	0.82	0.80	0.56	0.55	0.75
MBSMulliken	0.70	0.76	0.72	0.77	0.90	0.90	0.88	0.90	0.96	0.96	1.00	0.68	0.84	0.85	0.83	0.52	0.49	0.76
i-ACP	0.78	0.75	0.71	0.89	0.85	0.86	0.87	0.89	0.73	0.71	0.68	1.00	0.88	0.77	0.77	0.84	0.83	0.59
ACP	0.78	0.80	0.77	0.85	0.89	0.90	0.86	0.92	0.85	0.81	0.84	0.88	1.00	0.91	0.89	0.63	0.60	0.75
CM5	0.76	0.78	0.73	0.78	0.84	0.84	0.80	0.87	0.84	0.82	0.85	0.77	0.91	1.00	0.97	0.51	0.48	0.81
EEQ	0.73	0.73	0.70	0.77	0.81	0.82	0.78	0.86	0.81	0.80	0.83	0.77	0.89	0.97	1.00	0.54	0.50	0.76
QTAIM	0.64	0.61	0.48	0.69	0.66	0.66	0.74	0.75	0.60	0.56	0.52	0.84	0.63	0.51	1.00	0.86	0.86	0.36
APT	0.68	0.62	0.51	0.74	0.67	0.66	0.73	0.71	0.55	0.55	0.49	0.83	0.60	0.48	0.50	1.00	1.00	0.36
ADCH	0.68	0.72	0.68	0.69	0.75	0.74	0.68	0.71	0.73	0.75	0.76	0.59	0.75	0.81	0.76	0.36	0.36	1.00

- been lower for a more solid state/materials science oriented sample, where DDEC6 should be more resilient.)
- (d) MBSMulliken, NPA, and IBO form a block with all  $R^2 \geq 0.96$ , and indeed  $R^2$  between NPA and MBSMulliken reaches 0.98. Of these three, NPA has the largest cross-correlations with iterative Class IIb1 methods (other than ISA), and IBO the smallest. MBSBickelhaupt correlates slightly less well with NPA and IBO, but better with the iterative Class IIb1 methods.
- (e) CM5 and EEQ exhibit a surprisingly large  $R^2 = 0.97$  between them; this might be ascribed to the design of CM5 as an empirical correction to Hirshfeld for better electrostatics. However, the correlation between either method and HLY, for instance, is fairly low ( $R^2 = 0.70$  for EEQ, 0.73 for CM5) while rather better correlations are found with some of the iterative Hirshfeld variants, and particularly with ACP. Somewhat surprisingly, since ADCH like CM5 was designed as a correction to Hirshfeld for better properties, ADCH correlates less well with CM5 than one might expect ( $R^2 = 0.81$ ).
- (f) QTAIM has relatively high  $R^2 = \{0.86, 0.84\}$  with  $\{APT, i-ACP\}$ , but much smaller ones with the ‘‘Hirshfeld’’ and ‘‘NPA’’ blocks, or with the electrostatic blocks.
- (g) i-ACP acts as a kind of ‘bridge’ to the various iterative Hirshfeld flavors, with  $R^2 = 0.85-0.89$ . The older ACP, on the other hand, correlates less well with the Hirshfeld family but rather better,  $R^2 = 0.89$  and 0.91, with EEQ and CM5, respectively.

Principal component analysis (PCA) on the correlation matrix reveals that the first two principal components, with eigenvalues 15.93 and 0.82, respectively, account for most of the variation in the dataset. Let us now switch to PCA on the covariance matrix. (Table 3)

Normalizing the eigenvalues by the number of variables, we find that the first three eigenvectors (i.e., the first three principal components) contain as much variation as 20 out of the total 21 primitive variables.

The first eigenvector,  $PC_1$ , has all charges pulling in the same direction with different strengths:  $PC_1$  can thus be said to correspond to Schwarz and Meister’s<sup>[7]</sup> ‘principal component of ionicity’. The different coefficients correspond roughly to how pronounced charge differences are: e.g., for the original Hirshfeld we see 0.08, for iterative Hirshfeld 0.26, for DDEC6 0.21, and for QTAIM 0.38.

$PC_2$  is another matter: it pits QTAIM, APT, and i-ACP with same signs against every other distribution with opposite sign.  $PC_2$  can hence be seen as a secondary ‘‘principal component of ionicity’’. Why APT, despite its different physical principles, marches in comparative lockstep with QTAIM is somewhat opaque to the authors.

$PC_3$ , which may still be of marginal statistical significance, pits the electrostatic charges and ISA against the orbital-based charges.

In order to verify how well determined  $PC_2$  and  $PC_3$  are, about 10% random numbers were added to the sample and the analysis repeated. The eigenvalue for  $PC_2$  clearly stays in well-defined territory, but  $PC_3$  appears to be in danger of

**Table 3.** First six principal components found after PCA on the covariance matrix.

Eigenvalues	2.145	0.157	0.072	0.022	0.017	0.010
VDD	-0.08	0.00	-0.01	0.15	0.22	-0.26
Hirshfeld	-0.08	0.02	-0.02	0.12	0.18	-0.27
CHELPG	-0.21	0.01	0.37	0.04	-0.03	-0.03
MK	-0.21	0.11	0.39	-0.08	-0.11	-0.08
RESP	-0.21	0.10	0.37	-0.05	-0.09	-0.11
HLY	-0.22	0.20	0.38	-0.16	-0.28	-0.14
ISA	-0.23	0.03	0.21	-0.07	0.14	0.12
DDEC6	-0.21	0.09	-0.01	-0.09	0.10	0.10
MBIS	-0.26	0.11	-0.02	-0.13	0.02	0.32
Hirshfeld-I	-0.26	0.03	-0.08	-0.23	0.06	0.22
MBSBickelhaupt	-0.26	0.02	-0.14	0.12	-0.08	0.25
NPA	-0.28	0.20	-0.35	-0.22	-0.05	0.01
IBO	-0.21	0.17	-0.23	-0.16	0.16	-0.35
MBSMulliken	-0.28	0.29	-0.37	-0.15	-0.07	-0.08
i-ACP	-0.20	-0.13	0.07	0.28	0.09	0.32
ACP	-0.14	0.06	-0.02	0.31	0.01	0.21
CM5	-0.14	0.13	-0.05	0.43	0.01	0.10
EEQ	-0.14	0.11	-0.06	0.46	-0.02	0.09
QTAIM	-0.38	-0.72	-0.18	0.05	-0.47	-0.24
APT	-0.24	-0.40	0.09	-0.14	0.71	-0.02
ADCH	-0.13	0.20	-0.02	0.37	0.12	-0.49

‘drowning in the noise’. We therefore caution against over-interpreting PC<sub>3</sub>.

What happens if we retain just one representative of each “physical” group? This leads to the following:

Eigenvalues	0.909	0.096	0.023	0.014	0.009	0.003
HLY	0.309	-0.448	<b>0.705</b>	-0.403	-0.120	-0.175
Hirshfeld-I	0.385	-0.253	-0.072	0.184	-0.289	<b>0.816</b>
EEQ	0.208	-0.247	-0.110	-0.102	<b>0.920</b>	0.164
APT	0.386	0.309	0.429	<b>0.724</b>	0.155	-0.156
QTAIM	0.630	<b>0.584</b>	-0.194	-0.469	-0.052	-0.046
NPA	<b>0.410</b>	-0.487	-0.515	0.224	-0.171	-0.501

Principal component analysis is one way to address the feature selection problem. PCs still require calculating or measuring all the variables, however: hence, in data science and machine learning, some attention has been devoted to the variable selection problem, i.e.: to extracting the subset of original variables that contains most of the information in the dataset. A simplistic way is ‘hunting’ for the variables with the largest coefficients in the PCs, but Jolliffe’s textbook on PCA<sup>[69]</sup> cautions against this. Ref. [70] discusses more sophisticated approaches and their implementation in the subselect module of the R statistical software system.<sup>[71]</sup>

“Subselect” works through simulated annealing to optimize a choice of objective functions.<sup>[70]</sup> We ended up using the GCD (generalized coefficient of determination) criterion, that is, the overlap between the subspace spanned by variable subset size  $n$  and that spanned by first  $n$  principal components. (The analysis was run using R version 3.6.1 on the senior author’s Macbook Pro.)

However, if we apply *subselect* to the covariance matrix of the various population types, we find that the objective function goes through a maximum for 3 variables, and that these are MBSMulliken, QTAIM, and HLY. (Solutions using another electrostatic potential charge instead of HLY, or NPA instead of MBSMulliken, are essentially of the same quality.) For a single variable, we get Hirshfeld-I as the optimum; for two

variables, MBSMulliken and QTAIM. Expanding to four and five variables successively adds EEQ and APT, respectively.

If one forces inclusion of Hirshfeld-I, then successive additions lead to (again, using the GCD criterion):

2 variables: Hirshfeld-I + QTAIM

3 variables: Hirshfeld-I + QTAIM + ISA

4 variables: Hirshfeld-I + QTAIM + ISA + EEQ

Beyond which the “mold is broken”:

5 variables: Hirshfeld-I + QTAIM + EEQ + HLY + APT

However, multiple regression in terms of (MBSMulliken, HLY, QTAIM) reveals that Hirshfeld-I can be reproduced at  $R^2=0.961$  by  $0.252 \text{ HLY} + 0.432 \text{ MBSMulliken} + 0.205 \text{ QTAIM}$ . In other words, if those other three variables are present then Hirshfeld-I is effectively redundant.

Intriguingly, said three variables correspond to one representative each of the three Corninboeuf classes.

Let us consider the eigenvalues and eigenvectors of the covariance matrix for this 3-variable subset:

	HLY	QTAIM	NPA
HLY	0.120	0.152	0.127
QTAIM	0.152	0.398	0.208
NPA	0.127	0.208	0.183
Eigenvalue:	0.60	0.08	0.02

HLY	-0.38	0.56	-0.74
QTAIM	-0.78	-0.62	-0.07
NPA	-0.50	0.55	0.67

The structure here is very clear: (1) Principal ionicity; (2) QTAIM in opposition to the others; (3) Electrostatic versus orbital.

Another angle on the issue is afforded by considering the squared Pearson correlations  $R^2$  for least-squares fits of the individual charges in terms of the first few principal components. The result is shown in Table 4.

We note that the first principal component has the strongest correlation with Hirshfeld-I ( $R^2=0.9752$ ) and with MBSBickelhaupt ( $R^2=0.9765$ ), while MBIS and DDEC6 (both iterative Hirshfeld variants) come quite close at  $R^2=0.963$  and  $0.964$ , respectively. The effect of adding the second principal component is essentially negligible for Hirshfeld-I and MBSBickelhaupt (both of which have small loadings in PC2) but QTAIM is now the winner at  $R^2=0.9842$ , albeit closely followed by the iterative Class IIb1 methods other than ISA. If we consider the largest increases in  $R^2$  from one to two PCs, then QTAIM followed by APT seems to be most associated with PC2. The corresponding increase from adding PC3 is by far the largest for HLY, which jumps from 0.84 to 0.95. To reach 0.99 territory, HLY requires adding in PC5 and PC6: its loading in PC4 is too small to be useful. CM5 and EEQ would benefit from a 4<sup>th</sup> PC.

From the converse point of view, we may consider how well the three first PCs would be expressed as linear combinations of the three variables (MBSMulliken or NPA, QTAIM, HLY). The  $R^2$  for these fits are 0.990 for PC1, 0.977 for PC2, and 0.843 for PC3. Increasing the latter number further (as well as good fits for PC4 and PC5) can be achieved by adding EEQ and APT.

### 3. Conclusions

From a principal component analysis of about two dozen different charge distributions for a sample of over 2,000 main-group molecules, we can establish the following.

1. There are two very well-defined “principal components of ionicity” PCq1 and PCq2, and a more weakly defined third component PCq3.
2. The trio of individual charge distributions that best describes the space covered by PCq1, PCq2, and PCq3 is QTAIM with MBSMulliken (or NPA) and HLY (or another electrostatic potential charge) — in other words, one representative each of the three classes of Corminboeuf et al. If we omit PCq3, then HLY (or other electrostatic charge) can be omitted.
3. For the single charge distribution that most closely resembles PCq1, Hirshfeld-I and MBS-Bickelhaupt are effectively tied, both with  $R^2 > 0.975$ . If QTAIM, MBSMulliken, and HLY are included, however, Hirshfeld-I can be represented as their linear combination with  $R^2 > 0.96$ : Hirshfeld-I  $\approx 0.252$  HLY + 0.432 MBSMulliken + 0.205 QTAIM.
4. The “first principal component of ionicity” PCq1 corresponds to the one posited by Schwarz and Meister: all charges move in concert at different amplitudes.
5. The “second principal component of ionicity” PCq2 corresponds to the difference between QTAIM-type charges and other charge types, except that (intriguingly, in view of the different physics) APT somewhat resembles QTAIM.
6. The more weakly defined PCq3 corresponds to the opposition between electrostatic and orbital-based charges, leaving QTAIM nearly unaffected. Further components are statistically too weakly defined.
7. Most types of partial charges can be fitted quite well as linear combinations of PCq1, PCq2, and PCq3, though some require a fourth or fifth principal component for a good fit. Intriguingly, original Hirshfeld is the most resistant to expansion in principal components, requiring as many as ten of them.
8. Behavior of IBO (intrinsic bond orbital) charges is statistically very similar to NPA (natural population analysis) as well as MBSMulliken, that is, Mulliken population analysis after minimal basis set projection.

**Table 4.**  $R^2$  for least-squares fits of the individual charges as linear combinations of the first few principal components.

	C(PC <sub>1</sub> )	C(PC <sub>2</sub> )	C(PC <sub>3</sub> )	$R^2$ (PC <sub>1</sub> )	$R^2$ (PC <sub>1,2</sub> )	$R^2$ (PC <sub>1,2,3</sub> )
Hirshfeld	0.085	0.027	0.017	0.8092	0.8158	0.8164 <sup>[e]</sup>
CM5	0.151	0.145	0.019	0.8348	0.8981	0.8983 <sup>[c]</sup>
Hirshfeld-I	0.278	0.052	0.007	<b>0.9752</b>	0.9780	0.9780
NPA	0.304	0.253	-0.342	0.9160	0.9680	0.9921
HLY	0.229	0.193	0.601	0.7959	0.8419	0.9552 <sup>[a]</sup>
ISA	0.252	0.032	0.404	0.9362	0.9338	0.9835
MBIS	0.277	0.137	0.111	0.9629	0.9821	0.9853
DDEC6	0.227	0.108	0.104	0.9638	0.9818	0.9860
MBSMulliken	0.301	0.340	-0.365	0.8766	0.9677	<b>0.9943</b>
EEQ	0.155	0.122	-0.003	0.8297	0.8718	0.8718 <sup>[d]</sup>
QTAIM	0.419	-0.699	-0.274	0.8022	<b>0.9842</b>	0.9913
IBO	0.226	0.201	-0.216	0.8913	0.9492	0.9661
ACP	0.155	0.075	0.069	0.9007	0.9180	0.9216
i-ACP	0.213	-0.121	0.185	0.9207	0.9450	0.9594
APT	0.258	-0.401	0.187	0.7774	0.9303	0.9387 <sup>[b]</sup>
MBSBickelhaupt	0.281	0.047	-0.075	<b>0.9765</b>	0.9787	0.9801

[a]  $R^2=0.9972$  with 6 components: C(PC4) = -0.080, C(PC5) = -0.470, C(PC6) = +0.470. [b]  $R^2=0.9938$  with 5 components: C(PC4) = -0.328, C(PC5) = 0.666. [c]  $R^2=0.9801$  with 4 components: C(PC4) = 0.455. [d]  $R^2=0.9681$  with 4 components: C(PC4) = 0.509. [e]  $R^2=0.9324$  with 7 components,  $R^2=0.9927$  with 10 components.

Somewhat surprisingly, discrete vs. fuzzy partitioning of the deformation density (i.e., Voronoi Deformation Densities vs. original Hirshfeld charges) causes a much smaller difference than iterating the pro-atomic charges that make up the promolecular density.

We also define a minimal basis set-projected variant of Bickelhaupt charges, which should be easy to implement in any electronic structure system, or postprocessing code for same, for which one has source code available.

In all, no single charge distribution tells the whole story, but a trio of well-defined ones – i.e., one representative each of the three major classes of Corminboeuf and coworkers – should be able to cover most aspects of it.

## Acknowledgments

M.C. and S.E. acknowledge fellowships from the Kupcynet-Getz International Summer School at the Weizmann Institute of Science, and N.S. a doctoral fellowship from the Feinberg Graduate School. This work was supported in part by the Israel Science Foundation (grant 1358/15) as well as by two internal Weizmann Institute funding sources: the Helen and Martin Kimmel Center for Molecular Design and a research grant from the estate of Emile Mimran. J.M.L.M. thanks Prof. Paul W. Ayers and Dr. Farnaz Heidar-Zadeh for helpful discussions and Dr. Christoph Bannwarth for feedback on a preprint. Finally, the authors would like to pay tribute to the legacy of the Weizmann Institute crystallographer and father of what he termed “stockholder population analysis”, Prof. Frederick L. Hirshfeld (1927–1991).<sup>[72]</sup>

## Conflict of Interest

The authors declare no conflict of interest.

**Keywords:** atoms in molecules · charge distributions · molecular modelling · population analysis · principal components of ionicity

- [1] R. F. W. Bader, *Atoms In Molecules: A Quantum Theory*, Clarendon Press, Oxford, 1990.
- [2] R. G. Parr, P. W. Ayers, R. F. Nalewajski, *J. Phys. Chem. A* **2005**, *109*, 3957–3959.
- [3] G. Frenking, A. Krapp, *J. Comput. Chem.* **2007**, *28*, 15–24.
- [4] R. F. W. Bader, C. F. Matta, *J. Phys. Chem. A* **2004**, *108*, 8385–8394.
- [5] F. Matta, R. F. W. Bader, *J. Phys. Chem. A* **2006**, *110*, 6365–6371.
- [6] M. Maimonides, *The Guide for the Perplexed*, Routledge, 1904.
- [7] J. Meister, W. H. E. Schwarz, *J. Phys. Chem.* **1994**, *98*, 8245–8252.
- [8] F. Gonther, S. N. Steinmann, M. D. Wodrich, C. Corminboeuf, *Chem. Soc. Rev.* **2012**, *41*, 4671–4687.
- [9] J. W. Storer, D. J. Giesen, C. J. Cramer, D. G. Truhlar, *J. Comput.-Aided Mol. Des.* **1995**, *9*, 87–110.
- [10] L. E. Chirlian, M. M. Francl, *J. Comput. Chem.* **1987**, *8*, 894–905.
- [11] C. M. Breneman, K. B. Wiberg, *J. Comput. Chem.* **1990**, *11*, 361–373.
- [12] B. H. Besler, K. M. Merz, P. A. Kollman, *J. Comput. Chem.* **1990**, *11*, 431–439.
- [13] H. Hu, Z. Lu, W. Yang, *J. Chem. Theory Comput.* **2007**, *3*, 1004–1013.
- [14] R. S. Mulliken, *J. Chem. Phys.* **1955**, *23*, 1833–1840.
- [15] F. Weinhold, C. R. Landis, *Valency and Bonding*, Cambridge University Press, Cambridge, 2005.
- [16] G. Knizia, *J. Chem. Theory Comput.* **2013**, *9*, 4834–4843.
- [17] T. Janowski, *J. Chem. Theory Comput.* **2014**, *10*, 3085–3091.
- [18] W. C. Lu, C. Z. Wang, M. W. Schmidt, L. Bytautas, K. M. Ho, K. Ruedenberg, *J. Chem. Phys.* **2004**, *120*, 2629–2637.
- [19] C. Fonseca Guerra, J.-W. Handgraaf, E. J. Baerends, F. M. Bickelhaupt, *J. Comput. Chem.* **2004**, *25*, 189–210.
- [20] F. L. Hirshfeld, *Theor. Chim. Acta* **1977**, *44*, 129–138.
- [21] F. L. Hirshfeld, *Isr. J. Chem.* **1977**, *16*, 198–201.
- [22] P. Bultinck, C. Van Alsenoy, P. W. Ayers, R. Carbó-Dorca, *J. Chem. Phys.* **2007**, *126*, 144111.
- [23] T. C. Lillestolen, R. J. Wheatley, *J. Chem. Phys.* **2009**, *131*, 144101.
- [24] a) T. A. Manz, N. G. Limas, *RSC Adv.* **2016**, *6*, 47771–47801; b) N. G. Limas, T. A. Manz, *RSC Adv.* **2016**, *6*, 45727–45747; c) N. G. Limas, T. A. Manz, *RSC Adv.* **2017**, *7*, 45552–45581; d) N. G. Limas, T. A. Manz, *RSC Adv.* **2018**, *8*, 2678–2707; e) T. A. Manz, T. Chen, D. J. Cole, N. G. Limas, B. Fiszbein, *RSC Adv.* **2019**, *9*, 19297–19324; f) T. A. Manz, T. Chen, *RSC Adv.* **2019**, *9*, 33310–33336.
- [25] T. Verstraelen, S. Vandenbrande, F. Heidar-Zadeh, L. Vanduyfhuys, V. Van Speybroeck, M. Waroquier, P. W. Ayers, *J. Chem. Theory Comput.* **2016**, *12*, 3894–3912.
- [26] F. Heidar-Zadeh, P. W. Ayers, T. Verstraelen, I. Vinogradov, E. Vöhringer-Martinez, P. Bultinck, *J. Phys. Chem. A* **2018**, *122*, 4219–4245.
- [27] P. Coppens, *Annu. Rev. Phys. Chem.* **1992**, *43*, 663–692.
- [28] P. Coppens, *Phys. Scr.* **2013**, *87*, 048104.
- [29] A. V. Marenich, S. V. Jerome, C. J. Cramer, D. G. Truhlar, *J. Chem. Theory Comput.* **2012**, *8*, 527–541.
- [30] K. B. Wiberg, P. R. Rablen, *J. Comput. Chem.* **1993**, *14*, 1504–1518.
- [31] K. B. Wiberg, P. R. Rablen, *J. Org. Chem.* **2018**, *83*, 15463–15469.
- [32] S. M. Bachrach, in *Rev. Comput. Chem. Vol. 5* (Eds.: D. J. Boyd, K. B. Lipkowitz), Wiley, 1994, pp. 171–227.
- [33] J. Cioslowski, P. R. Surján, *J. Mol. Struct.* **1992**, *255*, 9–33.
- [34] P. Löwdin, *J. Chem. Phys.* **1950**, *18*, 365–375.
- [35] P. Ros, G. C. A. Schuit, *Theor. Chim. Acta* **1966**, *4*, 1–12.
- [36] F. M. Bickelhaupt, N. J. R. van Eikema Hommes, C. Fonseca Guerra, E. J. Baerends, *Organometallics* **1996**, *15*, 2923–2931.
- [37] J. A. Montgomery, M. J. Frisch, J. W. Ochterski, G. A. Petersson, *J. Chem. Phys.* **2000**, *112*, 6532–6542.
- [38] L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi, S. Grimme, *Phys. Chem. Chem. Phys.* **2017**, *19*, 32184–32215.
- [39] A. Najibi, L. Goerigk, *J. Chem. Theory Comput.* **2018**, *14*, 5725–5738.
- [40] N. Mehta, M. Casanova-Páez, L. Goerigk, *Phys. Chem. Chem. Phys.* **2018**, *20*, 23175–23194.
- [41] G. Santra, N. Sylvetsky, J. M. L. Martin, *J. Phys. Chem. A* **2019**, *123*, 5129–5143.
- [42] J. M. L. Martin, G. Santra, *Isr. J. Chem.* **2020**, *60*, ijch.201900114.
- [43] E. Caldeweyher, S. Ehlert, A. Hansen, H. Neugebauer, S. Spicher, C. Bannwarth, S. Grimme, *J. Chem. Phys.* **2019**, *150*, 154122.
- [44] W. C. Lu, C. Z. Wang, M. W. Schmidt, L. Bytautas, K. M. Ho, K. Ruedenberg, *J. Chem. Phys.* **2004**, *120*, 2638–2651.
- [45] R. F. Nalewajski, R. G. Parr, *Proc. Mont. Acad. Sci.* **2000**, *97*, 8879–8882.
- [46] P. Salvador, E. Ramos-Cordoba, *J. Chem. Phys.* **2013**, *139*, 071103.
- [47] A. Karton, N. Sylvetsky, J. M. L. Martin, *J. Comput. Chem.* **2017**, *38*, 2063–2075.
- [48] E. Maslen, M. Spackman, *Aust. J. Phys.* **1985**, *38*, 273.
- [49] M. A. Spackman, E. N. Maslen, *J. Phys. Chem.* **1986**, *90*, 2020–2027.
- [50] U. C. Singh, P. A. Kollman, *J. Comput. Chem.* **1984**, *5*, 129–145.
- [51] J. Wang, P. Cieplak, P. A. Kollman, *J. Comput. Chem.* **2000**, *21*, 1049–1074.
- [52] J. Cioslowski, *J. Am. Chem. Soc.* **1989**, *111*, 8333–8336.
- [53] T. Lu, F. Chen, *J. Theor. Comput. Chem.* **2012**, *11*, 163–183.
- [54] A. A. Voityuk, A. J. Stasyuk, S. F. Vyboishchikov, *Phys. Chem. Chem. Phys.* **2018**, *20*, 23328–23337.
- [55] S. F. Vyboishchikov, A. A. Voityuk, *J. Comput. Chem.* **2019**, *40*, 875–884.
- [56] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- [57] C. Adamo, V. Barone, *J. Chem. Phys.* **1999**, *110*, 6158–6170.
- [58] F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.
- [59] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J.



- Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, *Gaussian 16 Revision C.01*, Gaussian, Inc., Wallingford, CT, **2016**.
- [60] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, W. Györfy, D. Kats, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, S. J. Bennie, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Köppl, S. J. R. Lee, Y. Liu, A. W. Lloyd, Q. Ma, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, T. F. Miller III, M. E. Mura, A. Nicklaß, D. P. O'Neill, P. Palmieri, D. Peng, K. Pflüger, R. M. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, M. Wang, and M. Welborn, *MOLPRO, A Package of Ab Initio Programs, Version 2019.2*, Institute For Theoretical Chemistry, University Of Stuttgart, **2019**.
- [61] Y. Shao, Z. Gan, E. Epifanovsky, A. T. B. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng, D. Ghosh, M. Goldey, P. R. Horn, L. D. Jacobson, I. Kaliman, R. Z. Khaliullin, T. Kuś, A. Landau, J. Liu, E. I. Proynov, Y. M. Rhee, R. M. Richard, M. A. Rohrdanz, R. P. Steele, E. J. Sundstrom, H. L. Woodcock, P. M. Zimmerman, D. Zuev, B. Albrecht, E. Alguire, B. Austin, G. J. O. Beran, Y. a. Bernard, E. Berquist, K. Brandhorst, K. B. Bravaya, S. T. Brown, D. Casanova, C.-M. Chang, Y. Chen, S. H. Chien, K. D. Closser, D. L. Crittenden, M. Diedenhofen, R. A. DiStasio, H. Do, A. D. Dutoi, R. G. Edgar, S. Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-Zadorozhnaya, J. Gomes, M. W. D. Hanson-Heine, P. H. P. Harbach, A. W. Hauser, E. G. Hohenstein, Z. C. Holden, T.-C. Jagau, H. Ji, B. Kaduk, K. Khistyayev, J. Kim, J. Kim, R. A. King, P. Klunzinger, D. Kosenkov, T. Kowalczyk, C. M. Krauter, K. U. Lao, A. D. Laurent, K. V. Lawler, S. V. Levchenko, C. Y. Lin, F. Liu, E. Livshits, R. C. Lochan, A. Luenser, P. Manohar, S. F. Manzer, S.-P. Mao, N. Mardirossian, A. V. Marenich, S. A. Maurer, N. J. Mayhall, E. Neuscamman, C. M. Oana, R. Olivares-Amaya, D. P. O'Neill, J. A. Parkhill, T. M. Perrine, R. Peverati, A. Prociuk, D. R. Rehn, E. Rosta, N. J. Russ, S. M. Sharada, S. Sharma, D. W. Small, A. Sodt, T. Stein, D. Stück, Y.-C. Su, A. J. W. Thom, T. Tsuchimochi, V. Vanovschi, L. Vogt, O. Vydrov, T. Wang, M. A. Watson, J. Wenzel, A. White, C. F. Williams, J. Yang, S. Yeganeh, S. R. Yost, Z.-Q. You, I. Y. Zhang, X. Zhang, Y. Zhao, B. R. Brooks, G. K. L. Chan, D. M. Chipman, C. J. Cramer, W. A. Goddard, M. S. Gordon, W. J. Hehre, A. Klamt, H. F. Schaefer, M. W. Schmidt, C. D. Sherrill, D. G. Truhlar, A. Warshel, X. Xu, A. Aspuru-Guzik, R. Baer, A. T. Bell, N. A. Besley, J.-D. Chai, A. Dreuw, B. D. Dunietz, T. R. Furlani, S. R. Gwaltney, C.-P. Hsu, Y. Jung, J. Kong, D. S. Lambrecht, W. Liang, C. Ochsenfeld, V. A. Rassolov, L. V. Slipchenko, J. E. Subotnik, T. Van Voorhis, J. M. Herbert, A. I. Krylov, P. M. W. Gill, M. Head-Gordon, *Mol. Phys.* **2015**, *113*, 184–215.
- [62] S. Dasgupta, J. M. Herbert, *J. Comput. Chem.* **2017**, *38*, 869–882.
- [63] T. Lu, F. Chen, *J. Comput. Chem.* **2012**, *33*, 580–592.
- [64] A. D. Becke, *J. Chem. Phys.* **1988**, *88*, 2547–2553.
- [65] E. W. Stout, P. Politzer, *Theor. Chim. Acta* **1968**, *12*, 379–386.
- [66] T. Verstraelen, P. Tecmer, F. Heidar-Zadeh, C. E. González-Espinoza, M. Chan, T. D. Kim, K. Boguslawski, S. Fias, S. Vandenbrande, D. Berrocal, P. W. Ayers, *HORTON 3*, **2018**.
- [67] S. F. Vyboishchikov, *ACP and I-ACP*, Institute For Computational Chemistry And Catalysis, University Of Girona, Girona, **2018**.
- [68] S. A. Ghasemi, A. Hofstetter, S. Saha, S. Goedecker, *Phys. Rev. B - Condens. Matter Mater. Phys.* **2015**, *92*, 1–6.
- [69] I. T. Jolliffe, *Principal Component Analysis*, Springer-Verlag, New York, **2002**.
- [70] J. Cadima, J. O. Cerdeira, M. Minhoto, *Comput. Stat. Data Anal.* **2004**, *47*, 225–236.
- [71] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation For Statistical Computing, Vienna, Austria, **2013**.
- [72] F. Herstein, *J. Appl. Crystallogr.* **1991**, *24*, 1080–1081.

Manuscript received: January 15, 2020

Revised manuscript received: February 12, 2020

Accepted manuscript online: February 13, 2020

Version of record online: March 23, 2020