ACKS2: Atom-Condensed Kohn-Sham DFT approximated to second order

T. Verstraelen,[1, a)] P.W. Ayers,[2] V. Van Speybroeck,[1] and M. Waroquier[1]

[1)] *Center for Molecular Modeling (CMM), Ghent University, 9000 Ghent, Belgium,*
(Member of the QCMM Ghent-Brussels Alliance Group)

[2)] *Department of Chemistry, McMaster University, 1280 Main Street West,*
Hamilton, Ontario, Canada

(Dated: 28 January 2013)

A new polarizable force field (PFF), namely ACKS2, is proposed for the efficient computation of atomic charges and linear response properties of extended molecular systems. It is derived from Kohn-Sham density functional theory (KS-DFT), making use of two novel ingredients in the context of PFFs: (i) constrained atomic populations and (ii) the Legendre transform of the Kohn-Sham kinetic energy. ACKS2 is essentially an extension of the Electronegativity Equalization Method (EEM) [J. Am. Chem. Soc. **108**, 4315 (1986)] in which two major EEM shortcomings are fixed: ACKS2 predicts a linear size-dependence of the dipole polarizability in the macroscopic limit and correctly describes the charge distribution when a molecule dissociates. All ACKS2 parameters are defined as atoms-in-molecules expectation values. The implementation of ACKS2 is very similar to that of EEM, with only a small increase in computational cost.

Keywords: polarizable force fields, electronegativity equalization, charge equilibration, Legendre transform, Kohn-Sham kinetic energy, constrained density functional theory, dissociation, linear response

---

[a)]Electronic mail: Toon.Verstraelen@UGent.be

I. INTRODUCTION

Sanderson's principle of electronegativity equalization states that, upon the formation of a molecule, electrons flow until all electronegativities are equalized.[1] Essentially the same principle is found in density functional theory (DFT), stating that the electronic ground state has a constant chemical potential.[1,2] Starting from basic DFT equations, Mortier et al. derived the electronegativity equalization method (EEM),[3,4] providing an elegant mathematical reformulation of Sanderson's principle. EEM assumes a quadratic model for the molecular energy as function of the atomic charges. A minimization of this energy (with a constraint on the total charge) leads to a set of linear electronegativity equations, whose solution yields *ab initio quality* atomic charges at minimal computational cost. After the introduction of EEM in the seminal paper of Mortier,[4] the model was extensively validated and applied to diverse chemical systems, including inorganic solids,[5–12] organic molecules,[13–29] biomolecular systems[30–33] and metal-organic frameworks.[34] These successes are the basis for the wide-spread adoption of EEM as a polarizable force field (PFF) used in high-throughput in-silico screening[35] and in molecular mechanics force fields.[9,13,15,20,21,24–26,31,36–38] Besides its practical utility, EEM is also of fundamental importance because it explains the molecular charge distribution with just a few simple equations and a set of transferable parameters.

Most of the EEM improvements that have been proposed over the past 25 years can be divided into two broad families. Early extensions of EEM (1986-2000) are *refinements* that introduced more details to obtain a more accurate model. More recent extensions (2000-2012) address the fundamental shortcomings of EEM through *alternative forms* of the model. Note that the calibration of EEM parameters for new elements or atom types is not considered as an extension in this context. One should also be aware that different names are commonly used for models that are very similar to EEM, e.g. charge equilibration (QE, QEq[39] or CHEQ[29]), fluctuating charges (FQ or FlucQ)[13], chemical potential equalization (CPE)[40], and so on.

The early adjustments of EEM (1986-2000) are straightforward approaches to obtain a more realistic model. Rappe and Goddard pioneered the use of distributed charge densities instead of atomic point charges[39] to achieve more accurate electrostatic interactions at short inter-atomic distances, in analogy with Thole's smeared atomic inducible dipoles.[41] Originally, Slater-type densities were used, but later works also employed Gaussian-type

densities[16,28,40] or other models.[8,15,30,42] The second important generalization was the introduction of atomic s-type and p-type density basis functions.[40,43] This allows a much more detailed description of electronic polarization, e.g. orthogonal to planar molecules, and is therefore widely used in PFFs.[44-49] These PFFs are often used in molecular dynamics (MD) simulations based on Born-Oppenheimer,[15] extended Lagrangian[20,36,50] (cfr. Car-Parrinello[51]) or predictor-corrector[52] approaches to compute atomic charges and/or dipoles at each time step. The applications of these early extension go beyond a fast computation of atomic charges and electrostatic interactions. IR and Raman intensities,[14] intermolecular charge transfer[53] and chemical reactivity indices[54,55] have also been modeled using improved EEM schemes.

Later model development papers (2000-2012) have focused on the more fundamental limitations of EEM, which become problematic in simulations of extended systems or when one studies atomic charges during chemical reactions. The first problem is that EEM always predicts a cubic scaling of the dipole polarizability with system size, while dielectric systems exhibit a linear scaling in the macroscopic limit.[56,57] The second problem is that EEM yields, in general, fractional molecular charges for a system with two or more molecules, even when these molecules are well separated.[58,59] For such systems, one expects integer-charged molecules because the energy of an isolated molecule is a piece-wise linear function of the molecular population with derivative discontinuities at integer populations.[60-62] Both problems boil down to one general weakness: EEM always allows long-range (even through-space) charge transfer, while this is only realistic inside conductor-like systems, e.g. metals or conjugated hydrocarbons.

Several ad hoc approaches were proposed to solve both EEM problems, e.g., with artificial constraints on molecular charges[20,57] or with harmonic restraints on molecular dipoles.[47] One can also suppress the impact of electronegativity differences at long distances.[59,63,64] A promising strategy to obtain correct dissociation limits, is the derivation of EEM variants from valence bond theory,[59,63-65] but to our knowledge no one has yet shown the transferability of parameters in these models within a broad class of molecular systems.

A turning point in the field was the atom-atom charge transfer (AACT) paper of Chelli et al.[56] In AACT, atomic charges are expressed in terms of split-charge[66] variables. Split charges (a.k.a. bond charges[30] or atom-atom charge transfers[56]) form a redundant basis for the atomic charges and are mathematically equivalent to bond-charge increments.[67,68] A

3

split charge is associated with a pair of atoms and determines the amount of charge that is transferred from one atom to the other. It can also be seen as a finite dipole, as opposed to a point dipole. In analogy with the inducible point dipole model,[41] the AACT energy contains quadratic split-charge terms with the bond hardness as a linear parameter. These terms guarantee a linear scaling of the dipole polarizability.[56,57] It is common to associate split charges only with pairs of atoms that are covalently bonded, which disables intermolecular charge transfer. Later, Nistor et al. proposed the split-charge equilibration (SQE),[66] which has EEM and AACT as limiting cases. SQE parameters were calibrated for organic[28] and inorganic[12,66] systems, and were found to be transferable to other molecules not included in the training set. The transferability indicates that the bond hardness parameter must have some physical interpretation. Nistor and Müser provided such interpretation by showing that the bond hardness correlates with the dielectric constant and the band gap of a solid in the macroscopic limit.[69,70] In analogy with EEM, SQE can also be extended with atomic inducible dipoles to refine the model.[71]

In spite of the valuable recent EEM extensions (2000-2012), many practical and fundamental questions about these improved EEM schemes remain unanswered, which impedes their broad application. Currently, the most pressing issues related to SQE are the following:

1. AACT, and by extension SQE, are mathematical generalizations of EEM for which there are merely some intuitive physical motivations. Unlike EEM, these improved models are not yet supported by a direct derivation from an underlying and generally accepted theory, such as DFT. The split-charge variable is the root of this problem. Split-charges are not defined as atoms-in-molecules (AIM)[72,73] quantities that can be computed directly from the electronic density or wavefunction. If such a definition would be available, one could try to approximate the DFT energy in terms of split charges and construct an EEM-like model. Currently split-charges can only be derived by first computing AIM charges, followed by a transformation of these charges into split charges. Furthermore, the transformation to split charges is not even uniquely defined, i.e. a single set of atomic charges corresponds in general to a manifold of split charges,[74] suggesting that the split charge is merely a mathematical construct instead of a tangible physical quantity.

2. AACT and SQE only provide an ad hoc solution to fix intermolecular charge transfer:

by convention, one excludes intermolecular split charges. Although some attempts were made to describe bond-breaking in SQE,[75,76] these are troublesome in numerical applications because they directly follow the suggestion of Cioslowsky[58] to let the bond hardness diverge to infinity as atoms separate. This is especially problematic when one tries to apply these models in the context of reactive force fields. Moreover, there are only empirical guidelines to determine which split charges should be excluded.

3. Parameters for EEM, AACT, SQE and many related models are determined with daunting calibration procedures so as to reproduce atomic charges and other properties for a large training set of molecules. Due to the large number of EEM parameters, these calibrations are afflicted by statistical correlations.[8,12,77] The more advanced the model, the more parameters it contains and the harder it becomes to determine the parameters. Keeping in mind that extended EEM models will be even more parameter-laden, one urgently needs an alternative to the conventional calibration of parameters. Although several authors have suggested direct expressions for the EEM parameters,[78,79] these did not gain a wide-spread adoption.

4. AACT and SQE can only describe neutral molecules. This is a consequence of the mathematical definition of the split charge. A recent extension, SQE+Q^0,[33,70] provided a solution on empirical grounds, but lacks a profound theoretical support.

The goal of this paper is to solve definitively the first and the second problems. For the third and fourth problems, this paper provides a solid foundation for future work. These goals are realized with the derivation of a new atomic charge model from Kohn-Sham DFT, namely "Atom-Condensed Kohn-Sham DFT approximated to second order" (ACKS2). The new model completely operates in atom-space,[74] e.g. split charges are no longer needed to reproduce the attractive features of AACT and SQE. Nevertheless, SQE is a limiting case of ACKS2 and the inverse of the bond hardness is still present as an off-diagonal parameter. We also show that this model can be used to correctly describe atomic charges when covalent bonds break or form. In the limit of large separations, ACKS2 predicts that the charges of molecular fragments become integers. Furthermore, our derivation also provides AIM expressions for all parameters in the model, enabling a direct computation of the parameters from a Kohn-Sham wavefunction. An extensive numerical validation of

these AIM parameter expressions will be published in a subsequent paper. Finally, ACKS2 can also describe charged species.

In Sec. II, ACKS2 is derived from Kohn-Sham DFT. A new derivation of EEM is proposed and the origins of its weaknesses are analyzed. The derivation of ACKS2 is based on this analysis. The solution of the ACKS2 model for a system that dissociates into two fragments is presented in Sec. III. Sec. IV describes the numerical implementation of ACKS2. Sec. V discusses potential extensions and applications and relates ACKS2 with other recent developments. The last section summarizes the main conclusions of this work. The equivalence of the ACKS2 and SQE+Q^0 models is shown in the Appendix. Atomic units are used unless noted otherwise.

II.   DERIVATION

The derivation of ACKS2 builds on a new connection between polarizable force fields (PFFs) and density functional theory (DFT), namely atom-condensed DFT. With this formalism, a new derivation of EEM is presented in which every essential approximation is made explicit. These approximations reveal the origins of the inherent weaknesses of EEM. Furthermore, this derivation shows how EEM parameters can be computed directly and how they depend on the molecular environment. The essential problem is that the EEM approximation cannot be justified for the non-local contributions to the energy functional in DFT, more specifically the kinetic energy. Therefore, a new second-order approximation is proposed for the Kohn-Sham kinetic energy, in which the atom-condensed DFT plays a central role. The ACKS2 energy is then defined as the sum of the new kinetic energy model and an EEM approximation for all other contributions to the DFT energy.

A.   Atom-Condensed DFT

In the conventional derivation of a PFF from DFT, one expands the changes in the electron density in a basis of atom-centered functions. By substituting this expansion in the DFT energy functional, one develops a model for the first- and second- order derivatives of the electronic energy towards the expansion coefficients.[40,79–82] Although the density basis expansion is also useful in this paper, the starting point of our derivation is constrained

DFT.[83,84] By imposing constraints on the electron density, people have already successfully studied charge transfer during bond dissociation,[58] long-range charge transfer,[85] and intermolecular charge-transfer and induction effects.[86] Note that these are exactly the phenomena that EEM cannot describe correctly. Therefore, we will use constraints on the atomic populations to define an atom-condensed DFT and use this as the starting point for the derivation of EEM and ACKS2.

In this work, the charge of an atom is defined as

$$q_A = Z_A - N_A \quad \forall A \in \{1, \dots, M\} \tag{1}$$

where Z_A is the nuclear charge, M is the number of atoms in the molecule and N_A is the atomic population defined as

$$N_A = \int \rho_A(\mathbf{r}) d\mathbf{r} \quad \text{with} \quad \rho_A(\mathbf{r}) = w_A(\mathbf{r})\rho(\mathbf{r}) \tag{2}$$

and where $w_A(\mathbf{r})$ is a weight function ($0 \leq w_A(\mathbf{r}) \leq 1$) that specifies which part of the total electron density is attributed to atom A. The function $\rho_A(\mathbf{r})$ is called the atoms-in-molecules (AIM) density of atom A. There are several schemes to derive atomic weight functions from the electronic density, e.g. Hirshfeld partitioning,[87] Becke partitioning,[88] Hirshfeld-I partitioning,[89] ISA,[90] and QTAIM.[91] In all these schemes, the weights add up to unity, so the sum of atomic populations is equal to the total molecular population:

$$\sum_{A=1}^{M} w_A(\mathbf{r}) = 1 \Leftrightarrow \sum_{A=1}^{M} N_A = N_{\text{tot}} \tag{3}$$

In this paper, the summation indexes A, B, C and D will be used for sums over atoms, which always go from 1 to M, unless noted otherwise.

This AIM partitioning can be coarse-grained towards molecular fragments by defining a fragment weight function as the sum of the weight functions of the constituting atoms. Similarly, one may refine the model with atomic dipoles, quadrupoles and so on. The inclusion of such degrees of freedom in the derivation below is not difficult but tedious.[40,43,92–94] In this work, the effect of inducible dipoles is neglected and (variations of) AIMs are approximated as spherically symmetric functions.

A PFF is essentially a model for the molecular electronic energy in terms of atomic populations (and higher multipoles). One can define an exact PFF energy by means of the

variational principle in DFT, combined with a constraint on each atomic population:

$$E_{\text{pop}}(N_1, \ldots, N_M) = \min_{\substack{\rho \\ N_A = \int w_A(\mathbf{r})\rho(\mathbf{r})d\mathbf{r}}} E_v[\rho] \tag{4}$$

The dependence of the energy on the nuclear coordinates is implicitly present through the external potential v. The constrained ground state corresponds to a stationary point of the Lagrangian:

$$\max_{\lambda_1, \ldots, \lambda_M} \min_{\rho} \left[E_v[\rho] - \sum_A \lambda_A \left(\int w_A(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} - N_A \right) \right] \tag{5}$$

Hereafter, the populations will be used as degrees of freedom, which — through the concept of population-constrained DFT — determine all other properties of the electronic system. The principal quantities we want to model in terms of populations are the energy:

$$E_{\text{pop}}(N_1, \ldots, N_M) \tag{6}$$

and the density

$$\rho_{\text{pop}}(N_1, \ldots, N_M; \mathbf{r}). \tag{7}$$

The conventional DFT energy functional, which depends on ρ, is hereby *condensed* in E_{pop}, which only depends on the atomic populations.

In this work, a PFF is constructed as an empirical model for a second-order expansion of E_{pop}. The primary application of such a PFF is to compute the charge distribution and linear response properties of extended molecular systems at a low computational cost. In the context of molecular mechanics force fields, one must also consider the dependence of E_{pop} on the nuclear coordinates and include the repulsion between the nuclei:

$$E_{\text{mol}} = E_{nn} + E_{\text{pop}}(N_1, \ldots, N_M) \tag{8}$$

It is insightful to compare population-constrained DFT with Levy's constrained search formulation of DFT.[95] The ground state populations can be found by minimizing E_{pop} with a constraint on the total charge. Since E_{pop} is in itself also defined through a constrained minimization, one may express the ground state energy as a double minimization:

$$\begin{aligned}
E_{\text{gs,pop}} &= \min_{\substack{N_1, \ldots, N_M \\ \sum_A N_A = N_{\text{tot}}}} E_{\text{pop}}(N_1, \ldots, N_M) \\
&= \min_{\substack{N_1, \ldots, N_M \\ \sum_A N_A = N_{\text{tot}}}} \min_{\substack{\rho \\ N_A = \int w_A(\mathbf{r})\rho(\mathbf{r})d\mathbf{r}}} E_v[\rho] \\
&= \min_{\substack{\rho \\ N_{\text{tot}} = \int \rho(\mathbf{r})d\mathbf{r}}} E_v[\rho]
\end{aligned} \tag{9}$$

The double minimization is presented schematically in Fig. 1. For each set of populations, $(N_1, \ldots, N_M)$, E_{pop} is the lowest possible energy of all densities with these populations. Hence, when the populations of the unconstrained DFT ground state are used, E_{pop} coincides with the DFT ground state energy, which justifies the last equality in Eq. (9), irrespective of the partitioning scheme used to define the weight functions, $w_A(\mathbf{r})$. This approach is reminiscent of Levy's constrained search formulation of DFT, which is also written as a double minimization. In Levy's work, the first minimization considers all possible N_{tot}-representable densities, while the second searches for the minimum energy wavefunction that corresponds to the given N_{tot}-representable density.

The derivation of ACKS2 makes use of the duality between the populations $(N_1, \ldots, N_M)$ and the Lagrange multipliers $(\lambda_1, \ldots, \lambda_M)$. It is assumed that the reader is familiar with the Legendre transform[96] that relates, in general, constrained variables with the corresponding Lagrange multipliers. Note that Eq. (5) is a non-standard Legendre transform of the following energy expression:

$$E_{\mathrm{pop}}(\lambda_1, \ldots, \lambda_M) = \min_{\rho} \left[E_v[\rho] - \sum_A \lambda_A \int w_A(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} \right] \tag{10}$$

This Legendre transform bears some similarities with Lieb's formulation of DFT.[97,98] One may also construct a PFF by expanding Eq. (10) to second order. This leads to the electronegativity equalization scheme of Cioslowski.[99,100] The fundamental novelty of ACKS2 is that it uses both the populations and a set of dual variables in a single model.

B.   Taylor expansion of the energy and the density in terms of atomic populations

Below we will introduce a truncated series expansion of the energy and the density in terms of atomic populations. Hence, we must select a reference point for the expansion, hereafter called *the reference populations*, denoted as $(N_1^0, \ldots, N_M^0)$. This state has a corresponding reference density and energy defined as:

$$\rho_{\mathrm{ref}}(\mathbf{r}) = \rho_{\mathrm{pop}}(N_1^0, \ldots, N_M^0; \mathbf{r}) \tag{11}$$

$$E_{\mathrm{ref}} = E_{\mathrm{pop}}(N_1^0, \ldots, N_M^0) = E_v[\rho_{\mathrm{ref}}] \tag{12}$$

Deviations from the reference populations will be called *relative populations*, denoted as $(\Delta_1, \ldots, \Delta_M)$, such that $N_A = N_A^0 + \Delta_A$. One can interpret the reference point as the *permanent* atomic monopoles, while the relative populations are the *induced* atomic monopoles.[40]

When the reference populations correspond to neutral atom populations, i.e. $N_A^0 = Z_A$, we will use the term *neutral reference*. EEM[4] and SQE,[66] use such a neutral reference, while York and Yang (CPE)[40] take the AIM populations of the ground state density as a reference. More recently, generalizations of the SQE were proposed, e.g. SQE+Q^0, that use integer reference populations corresponding to the oxidation states of the atoms.[33,70]

In line with the derivation of the CPE and related models,[40,79–82] the density is expanded to first order:

$$\rho_{\text{pop}}(N_1^0 + \Delta_1, \ldots, N_M^0 + \Delta_M; \mathbf{r}) \approx \rho_{\text{ref}}(\mathbf{r}) + \sum_A \Delta_A \left. \frac{\partial \rho_{\text{pop}}(\mathbf{r})}{\partial N_A} \right|_{N_* = N_*^0} \tag{13}$$

Similarly, the energy is expanded to second order:

$$E_{\text{pop}}(N_1^0 + \Delta_1, \ldots, N_M^0 + \Delta_M) \approx E_{\text{ref}} + \sum_A \Delta_A \left. \frac{\partial E_{\text{pop}}}{\partial N_A} \right|_{N_* = N_*^0}$$
$$+ \frac{1}{2} \sum_{AB} \Delta_A \Delta_B \left. \frac{\partial^2 E_{\text{pop}}}{\partial N_A \partial N_B} \right|_{N_* = N_*^0} \tag{14}$$

A compact notation with asterisks is used to indicate that the quantifier $\forall A \in \{1, \ldots, M\}$ is implicitly assumed, e.g. $N_* = N_*^0$ is a shorthand for $N_A = N_A^0 \;\; \forall A \in \{1, \ldots, M\}$.

The expansions of the density and the energy are only useful when one can derive (approximate) expressions for the reference energy and the first and the second-order derivatives. The chain rule, involving functional derivatives,[101] relates the expansion coefficients with standard DFT quantities:

$$\mu_A = \left. \frac{\partial E_{\text{pop}}}{\partial N_A} \right|_{N_* = N_*^0} = \int \mu[\rho_{\text{ref}}](\mathbf{r}) f_A(\mathbf{r}) d\mathbf{r} \tag{15}$$

$$\eta_{AB} = \left. \frac{\partial^2 E_{\text{pop}}}{\partial N_A \partial N_B} \right|_{N_* = N_*^0} = \iint \eta[\rho_{\text{ref}}](\mathbf{r}, \mathbf{r}') f_A(\mathbf{r}) f_B(\mathbf{r}') d\mathbf{r} d\mathbf{r}' \tag{16}$$

where

$$\mu[\rho_{\text{ref}}](\mathbf{r}) = \left. \frac{\delta E_v[\rho]}{\delta \rho(\mathbf{r})} \right|_{\rho = \rho_{\text{ref}}} \tag{17}$$

$$\eta[\rho_{\text{ref}}](\mathbf{r}, \mathbf{r}') = \left. \frac{\delta^2 E_v[\rho]}{\delta \rho(\mathbf{r}) \delta \rho(\mathbf{r}')} \right|_{\rho = \rho_{\text{ref}}} \tag{18}$$

$$f_A(\mathbf{r}) = \left. \frac{\partial \rho_{\text{pop}}(\mathbf{r})}{\partial N_A} \right|_{N_* = N_*^0} \tag{19}$$

Note that these equations bear many similarities with an earlier derivation of the CPE model,[79] but that they are now obtained through a constrained DFT formalism. The atomic chemical potential parameter, μ_A, represents the tendency of atom A to repel electrons. For example, when $\mu_A > \mu_B$ in a molecule with reference populations ($N_* = N_*^0$), the energy will decrease by transferring (a small amount of) electrons from atom A to atom B. The atomic hardness parameters, η_{AA}, and the off-diagonal hardness parameters, η_{AB}, can be interpreted as *force constants* that govern the resistance of the atoms to changes in populations. These parameters are the AIM counterparts of the electronegativity, $\chi_A(= -\mu_A)$, and hardness properties of isolated atoms,[102] except that (i) Parr and Pearson use finite differentiation and (ii) the *operational definition* of Parr and Pearson of the hardness includes the factor $\frac{1}{2}$ from the Taylor series. The function $f_A(\mathbf{r})$ will be referred to as the *AIM Fukui function* of atom A and will be discussed further below. It is similar (but not identical) to the well-known (molecular) Fukui function[103,104] or the condensed Fukui functions.[105] Note that the following term was not included in Eq. (16) because the density is only expanded to first order:

$$\int \mu[\rho_{\mathrm{ref}}](\mathbf{r}) \left. \frac{\partial^2 \rho_{\mathrm{pop}}(\mathbf{r})}{\partial N_A \partial N_B} \right|_{N_* = N_*^0} d\mathbf{r} \tag{20}$$

Using the notation introduced in Eqs. (15) to (19), the expansion of the density and the molecular energy, see Eq. (8), may be written compactly as follows:

$$\rho_{\mathrm{pop}}(\mathbf{r}) \approx \sum_A \rho_A^0(\mathbf{r}) + \Delta_A f_A(\mathbf{r}) \tag{21}$$

$$E_{\mathrm{mol}} \approx E_{nn} + E_{\mathrm{ref}} + \sum_A \Delta_A \mu_A + \frac{1}{2} \sum_{AB} \Delta_A \Delta_B \eta_{AB} \tag{22}$$

where $\rho_A^0(\mathbf{r})$, *the AIM reference density* of atom A, is obtained by a straightforward partitioning of the reference density, i.e. $\rho_A^0(\mathbf{r}) = w_A(\mathbf{r})\rho_{\mathrm{ref}}(\mathbf{r})$. In Sec. II C and II D, the second-order expansion of the molecular energy in Eq. (22) is the basis for the derivation of the EEM and ACKS2 models.

The AIM Fukui function is a new concept introduced in this work and it is helpful for the remainder of the text to discuss some properties of the AIM Fukui function explicitly. Note that analogous concepts, such as *perturbations of atomic electron densities*, were used by other authors in the derivation of electronegativity equalization models.[79] The AIM Fukui

function may be nonzero throughout the entire molecule:

$$f_A(\mathbf{r}) = \left. \frac{\partial \rho_{\text{pop}}(\mathbf{r})}{\partial N_A} \right|_{N_* = N_*^0} = \sum_B \left. \frac{\partial \rho_B(\mathbf{r})}{\partial N_A} \right|_{N_* = N_*^0} \tag{23}$$

However, because the atomic populations are independent variables, the changes in density of other atoms B integrate to zero. Similarly, the change in density of atom A integrates to one:

$$\int \left. \frac{\partial \rho_B(\mathbf{r})}{\partial N_A} \right|_{N_* = N_*^0} d\mathbf{r} = \left. \frac{\partial N_B}{\partial N_A} \right|_{N_* = N_*^0} = 0 \quad \forall B \neq A \tag{24}$$

$$\int \left. \frac{\partial \rho_A(\mathbf{r})}{\partial N_A} \right|_{N_* = N_*^0} d\mathbf{r} = \left. \frac{\partial N_A}{\partial N_A} \right|_{N_* = N_*^0} = 1 \tag{25}$$

This means that a change in population of atom A leads to density changes mainly near atom A. The electrons in the surrounding atoms may be polarized, but such changes in density integrate to zero. This is schematically represented in Fig. 2. Because this work only considers atomic monopolies, the following two approximations are regularly used below:

$$\left. \frac{\partial \rho_B(\mathbf{r})}{\partial N_A} \right|_{N_* = N_*^0} \approx 0 \quad \forall B \neq A \tag{26}$$

$$\left. \frac{\partial w_A(\mathbf{r}) \rho_{\text{pop}}(\mathbf{r})}{\partial N_A} \right|_{N_* = N_*^0} = \left. \frac{\partial \rho_A(\mathbf{r})}{\partial N_A} \right|_{N_* = N_*^0} \approx f_A(\mathbf{r}) \tag{27}$$

C. EEM

In this subsection, we start from Hohenberg-Kohn (HK) Density Functional Theory (DFT).[106] Step by step, the HK energy will be approximated until it is reduced to the well-known EEM energy. The primary motivation to analyze each step carefully, is to understand the origin of the well-known EEM limitations. Furthermore, this detailed derivation also reveals how one can estimate EEM parameters as AIM expectation values and how atomic parameters are affected by their molecular environment.

Condensed second-order approximation of HK-DFT energy. The HK electronic energy is:

$$E_v^{\text{HK}}[\rho] = \frac{1}{2} \iint \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' + E_{\text{txc}}[\rho] + \int \rho(\mathbf{r}) v(\mathbf{r}) d\mathbf{r} \tag{28}$$

where $E_{\text{txc}}[\rho]$ is the functional for the sum of the kinetic, exchange and correlation energy. This is slightly different from the conventional notation where the first two terms are combined in the so-called universal functional, $F[\rho]$. The first term, i.e. the Hartree term, plays

a special role in the EEM derivation and is therefore written explicitly. The last term is the interaction of the electron density with the external potential. The superscript HK will be omitted in the remainder of this subsection for the sake of compactness.

As a first step, one constructs a second-order expansion of the molecular energy in the HK formalism, as in Eq. (22). The constant term, the first- and the second-order coefficients are computed by applying the definitions in Eqs. (12), (15) and (16), respectively, to the HK energy:

$$E_{\text{ref}} = \frac{1}{2} \iint \frac{\rho_{\text{ref}}(\mathbf{r})\rho_{\text{ref}}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + E_{\text{txc}}[\rho_{\text{ref}}] + \int \rho_{\text{ref}}(\mathbf{r})v(\mathbf{r})d\mathbf{r} \tag{29}$$

$$\mu_A = \int \left(\int \frac{\rho_{\text{ref}}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + v_{\text{txc}}[\rho_{\text{ref}}](\mathbf{r}) + v(\mathbf{r}) \right) f_A(\mathbf{r})d\mathbf{r} \tag{30}$$

$$\eta_{AB} = \iint \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} + \eta_{\text{txc}}[\rho_{\text{ref}}](\mathbf{r}, \mathbf{r}') \right) f_A(\mathbf{r})f_B(\mathbf{r}')d\mathbf{r}d\mathbf{r}' \tag{31}$$

where $f_A(\mathbf{r})$ is defined in Eq. (19) and where the following functions were used:

$$v_{\text{txc}}[\rho_{\text{ref}}](\mathbf{r}) = \left. \frac{\delta E_{\text{txc}}[\rho]}{\delta\rho(\mathbf{r})} \right|_{\rho=\rho_{\text{ref}}} \tag{32}$$

$$\eta_{\text{txc}}[\rho_{\text{ref}}](\mathbf{r}, \mathbf{r}') = \left. \frac{\delta^2 E_{\text{txc}}[\rho]}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')} \right|_{\rho=\rho_{\text{ref}}} \tag{33}$$

The constant term (E_{ref}) is not considered in EEM because it is irrelevant for the computation of the atomic charges. Nevertheless, it does have a large contribution to the dependence of the energy on the nuclear coordinates, which is potentially valuable for the development of molecular mechanics models. Therefore, it is treated in this paper at the same level as the first- and second-order coefficients.

When one has empirical models for the AIM reference densities, $\rho_A^0(\mathbf{r}) = w_A(\mathbf{r})\rho_{\text{ref}}(\mathbf{r})$, and AIM Fukui functions, $f_A(\mathbf{r})$, Eqs. (29), (30) and (31) can be used to compute EEM parameters for any given configuration of the nuclei. Still, these equations do not explain two appealing EEM results: (i) the transferability of the atomic chemical potential, μ_A, and the atomic hardness, η_{AA}, between different molecular geometries and (ii) the approximation of $\eta_{AB} \approx 1/r_{AB}$ with $r_{AB} = |\mathbf{r}_A - \mathbf{r}_B|$. Such insights can only be obtained through additional approximations. First, we will introduce a set of approximations that leads to a more advanced *spherical atom* version of EEM. Later, the spherical atom EEM will be further simplified to obtain the traditional EEM equations.

Spherical atom and two-body approximations. In order to convert the second-order expansion of the HK energy into a transferable model, we will rely on two assumptions: (i)

AIM densities are approximately spherical and (ii) the HK energy can be approximated as a sum of one- and two-body terms. These assumptions allow us to approximate also E_{ref}, μ_A and η_{AB} as sums of one- and two-body terms, where the one-body terms are constants and the two-body terms only depend on the distance between atoms.

First, we assume that the atoms-in-molecules (AIM) densities are spherical (centered at the corresponding nuclei) to avoid the dependence of the HK energy on the orientation of the AIM densities:

$$\rho_A^0(\mathbf{r}) \approx \tilde{\rho}_A^0(|\mathbf{r} - \mathbf{r}_A|) \tag{34}$$

$$f_A(\mathbf{r}) \approx \tilde{f}_A(|\mathbf{r} - \mathbf{r}_A|) \tag{35}$$

The second assumption is that the HK energy can be approximated as a sum of one-body and two-body terms as follows:

$$E_v[\rho] \approx \sum_C E_{v_C}\big[w_C\rho\big] + \frac{1}{2}\sum_{CD}{}'\Big(E_{v_C+v_D}\big[(w_C+w_D)\rho\big] - E_{v_C}\big[w_C\rho\big] - E_{v_D}\big[w_D\rho\big]\Big) \tag{36}$$

where v_C is the external potential due to the nucleus of atom C and the primed double sum excludes terms for which $C = D$. Indices C and D are used instead of A and B to avoid confusion when taking derivatives towards N_A and N_B. The first sum contains the one-body terms, which are in general much larger than the two-body (interaction) terms in the second summation. The two-body terms are defined such that they converge to zero as atoms C and D separate. Higher many-body terms are neglected. For the Hartree term and the interaction with the external potential in Eq. (28), the two-body expansion is exact. One can easily verify this by substituting $\rho = \sum_C w_C\rho$ and $v = \sum_C v_C$ in the first and last term of Eq. (28). However, one can in general not justify the two-body approximation for the functional $E_{\mathrm{txc}}[\rho]$. One must assume that the functional $E_{\mathrm{txc}}[\rho]$ is semi-local and that the overlap between AIM densities remains small. A semi-local functional only predicts an interaction energy when the densities of two or more species overlap. When the AIM densities are well-localized, there are no regions in a molecule where more than two atoms contribute significantly to the density. Under these conditions, one can neglect the higher many-body terms.

Using the spherical atom and two-body approximations introduced above, we will now rewrite the second-order expansion of the HK energy. The reference energy – see Eq. (12) –

of the two-body expansion Eq. (36) becomes:

$$E_{\text{ref}} \approx \sum_C E_{v_C}[\rho_C^0] + \frac{1}{2} \sum_{CD}{}' \left(E_{v_C+v_D}[\rho_C^0 + \rho_D^0] - E_{v_C}[\rho_C^0] - E_{v_D}[\rho_D^0] \right) \tag{37}$$

The one-body terms are constants. Because we assume that the atoms are spherical, each two-body term becomes a function of only the corresponding inter-atomic distance and one can write:

$$E_{\text{ref}} \approx \sum_C E_{\text{ref},C} + \frac{1}{2} \sum_{CD}{}' E_{\text{ref},CD}(r_{CD}) \tag{38}$$

This form is similar to the frozen density (or frozen electron gas) approximation in the Kim-Gordon model, which can be used to estimate intermolecular interactions between closed-shell systems.[107,108] The main difference is that we are using spherically averaged AIM densities instead of densities of isolated systems. Like the frozen density approximation, Eq. (38) does not properly describe covalent interactions (bond stretch, angle bending, torsion and so on) because rearrangements of the density due to bond formation are neglected. Furthermore, one often relies on local approximations for the kinetic energy, e.g. the Thomas-Fermi model, which cannot describe covalent bonding.[109] This model for the reference energy can still be used to construct a force field for intermolecular Pauli repulsion interactions, provided one uses reasonable GGA approximations for the kinetic, exchange and correlation functionals.[110,111] When a non-neutral reference state is used, this term will also include an important electrostatic contribution.

The derivation of the first-order coefficients of the two-body energy – see Eq. (15) – is carried out in the same way. Making use of the chain rule and the approximations in Eqs. (26) and (27), one obtains the chemical potential parameter in the following compact form:

$$\mu_A = \mu_A^0 + \sum_B{}' \tilde{\mu}_A^B(r_{AB}) \tag{39}$$

where the primed sum excludes the term for which $B = A$ and where

$$\mu_A^0 = \int \left. \frac{\delta E_{v_A}[\rho]}{\delta \rho(\mathbf{r})} \right|_{\rho=\rho_A^0} f_A(\mathbf{r}) d\mathbf{r} \tag{40}$$

$$\tilde{\mu}_A^B(r_{AB}) = \int \left. \frac{\delta E_{v_A+v_B}[\rho]}{\delta \rho(\mathbf{r})} \right|_{\rho=\rho_A^0+\rho_B^0} f_A(\mathbf{r}) d\mathbf{r} - \mu_A^0 \tag{41}$$

Note that $\tilde{\mu}_A^B(r_{AB}) \neq \tilde{\mu}_B^A(r_{AB})$.

For the second-order coefficients – see Eq. (16) – one obtains:

$$\eta_{AA} \approx \eta_A^0 + \sum_B{}' \tilde{\eta}_{AA}^B(r_{AB}) \tag{42}$$

$$\eta_{AB} \approx \tilde{\eta}_{AB}(r_{AB}) \qquad \forall A \neq B \tag{43}$$

with

$$\eta_A^0 = \iint \left. \frac{\delta^2 E_{v_A}[\rho]}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')} \right|_{\rho=\rho_A^0} f_A(\mathbf{r})f_A(\mathbf{r}')d\mathbf{r}d\mathbf{r}' \tag{44}$$

$$\tilde{\eta}_{AA}^B(r_{AB}) = \iint \left. \frac{\delta^2 E_{v_A+v_B}[\rho]}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')} \right|_{\rho=\rho_A^0+\rho_B^0} f_A(\mathbf{r})f_A(\mathbf{r}')d\mathbf{r}d\mathbf{r}' - \eta_A^0 \tag{45}$$

$$\tilde{\eta}_{AB}(r_{AB}) = \iint \left. \frac{\delta^2 E_{v_A+v_B}[\rho]}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')} \right|_{\rho=\rho_A^0+\rho_B^0} f_A(\mathbf{r})f_B(\mathbf{r}')d\mathbf{r}d\mathbf{r}' \tag{46}$$

The diagonal coefficient, η_{AA}, consists of a constant term, the *intrinsic* atomic hardness of atom A,[4] and a sum of corrections due to interactions with neighboring atoms. In analogy with the atomic chemical potential, each correction term only depends on an inter-atomic distance. The off-diagonal coefficient, η_{AB}, only depends on the distance between atom A and B.

The parameters in Eqs. (40), (41), (44), (45) and (46) bear many similarities with Eqs. (15) and (16), but have a fundamental difference. In the spherical atom EEM, these parameters are purely based on AIM properties ρ_A^0 and f_A, whereas Eqs. (15) and (16) depend on the total molecular electronic density. In analogy with Eqs. (29), (30) and (31), one may substitute the HK functional form in the parameter expressions in Eqs. (40), (41), (44), (45) and (46). For a given GGA approximation of the kinetic, exchange and correlation functionals, one can use these expressions to pre-compute all the constants and pairwise functions in the EEM parameter expressions, which can then be applied to any configuration of the nuclei. Hence, the problem of defining a transferable second-order expansion for the HK energy is reduced to finding transferable models for ρ_A^0 and f_A. The transferability of these quantities will depend in the first place on the choice of the AIM scheme.[112] Nevertheless, some general remarks can be made. An atom in a molecule is to some extent always similar to an atom in vacuum, which implies that it should be possible to find a general ρ_A^0 for each element. Changes in atomic population are mainly situated in the valence shell of an atom. Therefore, $4\pi r^2 \tilde{f}_A(r)$ will only be significantly nonzero in a shell around the nucleus with the size of the covalent radius.[79] Given that the covalent radius is a well-defined property of each element, it should also be possible to find a transferable f_A for each element.

16

So far, we have introduced the spherical atom EEM. For the remainder of this paper, the most important result is that its derivation, through the two-body expansion in Eq. (36), depends on the assumption that $E_{\text{txc}}[\rho]$ is semi-local. This assumption is essential to obtain approximate expression for the parameters E_{ref}, μ_A and η_{AB} that depend only on transferable AIM properties and inter-atomic distances.

Besides this principal result, the spherical atom EEM also shows that it is possible to compute all EEM parameters as AIM properties. This result can be seen as a generalization of the expressions proposed by Itskowitz et al.[79] Conventionally, EEM parameters are obtained through daunting least-squares fits, in which correlations between chemical potential and hardness parameters are practically unavoidable.[8,12,77] The spherical atom EEM provides an alternative for these cumbersome calibration procedures. In analogy with the seminal paper of Mortier,[4] we also observe that the intrinsic chemical potential and hardness of an atom in a molecule must be corrected for the molecular environment. Similar corrections to the atomic parameters were already proposed and modeled earlier.[66,78,79,113,114] The expression for the off-diagonal matrix element $\tilde{\eta}_{AB}$ will converge to $1/r_{AB}$ at large distances, but at short distances several deviations are present. First of all, because the AIM Fukui functions, f_A and f_B, are smooth, the function $\tilde{\eta}_{AB}$ will remain finite, which is reminiscent of the distributed atomic charge densities in Qeq.[39] Furthermore, all the contributions to the pairwise functions due to non-classical terms such as kinetic, exchange and correlation functionals are similar to the exchange-polarization coupling[53] and Hückel-like corrections[40] introduced earlier. All these EEM corrections, of which most were proposed empirically by various authors, are all formally justified with the spherical atom EEM.

Finally, this derivation also reveals the characteristics of a suitable AIM scheme in the context of EEM. It is reported in the literature that EEM can reproduce atomic charges with different levels of accuracy, depending on which AIM scheme that was used to compute the charges for a given training set.[18,28] Based on the approximations introduced above, a suitable AIM scheme partitions the molecular density into atoms that are as spherical as possible and as local as possible. In practice one must find a compromise between these two criteria. For example, QTAIM densities are very local, but also very non-spherical.[91] The other extreme is ISA,[90] in which the sphericity of the atoms is optimized at the expense of their locality.[115,116] In our experience, Hirshfeld-I partitioning provides a reasonable compromise. On the one hand, this scheme yields charges that are competitive with electrostatic potential

(ESP) charges for the reproduction of the molecular electrostatic potential (MEP),[28,33,117] indicating that Hirsfheld-I AIMs are close to spherical. On the other hand, computational benchmarks also point out that Hirshfeld-I charges are only minimally sensitive to conformational changes (especially internal rotations), indicating the AIMs are well-localized.[33,118] Despite the encouraging computational results for the Hirshfeld-I method, a word of caution seems appropriate: most computational assessments of the Hirshfeld-I method consider organic molecules. A recent calibration of EEM and SQE parameters for silicates revealed that Hirshfeld-I charges systematically overestimate (in absolute value) the MEP of silica clusters.[12] Similar observations were made during the development of a forcefield for MIL-53(AL).[119] It is clear that there is still room for improved partitioning schemes that enhance the trade off between the sphericity and locality of the AIMs.

Point atom approximation. We will now introduce one more assumption to obtain the conventional EEM, namely that the overlap between AIM densities of different atoms can be neglected entirely. This allows several simplifications in the pairwise functions in Eqs. (39), (42) and (43): (i) all contributions due to the functional $E_{\mathrm{txc}}[\rho]$ can be discarded and (ii) the contributions from the Hartree term and the external potential are reduced to the Coulomb interaction between point charges placed at the nuclei. At this level of approximation, we have:

$$E_{\mathrm{ref}} \approx \sum_A E_{\mathrm{ref},A} + \frac{1}{2} \sum_{AB}{}' \frac{N_A^0 N_B^0}{r_{AB}} - \sum_{AB}{}' \frac{N_A^0 Z_B}{r_{AB}} \tag{47}$$

$$\mu_A \approx \mu_A^0 + \sum_B{}' \frac{N_B^0 - Z_B}{r_{AB}} \tag{48}$$

$$\eta_{AA} \approx \eta_A^0 \quad \text{and} \quad \eta_{AB} \approx \frac{1}{r_{AB}} \quad \text{if} \quad A \neq B \tag{49}$$

When the atomic reference energies are left out, the point atom approximation coincides with EEM as it was originally proposed by Mortier.[4]

$$E_{\mathrm{EEM}} = \sum_A \Delta_A \mu_A^0 + \sum_A \frac{\Delta_A^2}{2} \eta_A^0 + \frac{1}{2} \sum_{AB}{}' \frac{q_A q_B}{r_{AB}} \tag{50}$$

with $q_A = Z_A - N_A^0 - \Delta_A$. This expression also shows that a change in reference population can always be absorbed into the parameters μ_A^0 and η_A^0. In other words, an EEM model using the neutral reference is fully isomorphic with any other EEM model that is based on a non-neutral reference. Therefore, EEM is in practice always used with the neutral reference,

which is conventionally written as:

$$E_{\text{EEM}} = \sum_A q_A \chi_A^0 + \sum_A \frac{q_A^2}{2} \eta_A^0 + \frac{1}{2} \sum_{AB} {}' \frac{q_A q_B}{r_{AB}} \tag{51}$$

where the intrinsic atomic electronegativity parameters, χ_A^0, is equal to $-\mu_A^0$. Although the choice of the reference populations is not relevant for EEM, it will play an important role in the ACSK2 model.

Analysis of the approximations. Recently, it became clear that EEM has some fundamental limitations. First, EEM predicts that the dipole polarizability of a chain molecule grows cubically with the chain length, while one expects a linear trend in the macroscopic limit for dielectric molecules.[56,57] Second, one obtains fractional charges when a molecule dissociates,[58,59] while one expects integer-charged fragments.[60,62] These errors limit the applicability of EEM to isolated small molecules where an incorrect polarizability is acceptable. For other systems, one must introduce ad-hoc constraints to limit the impact of both errors.[20,57]

Several approximations were introduced in our EEM derivation, of which some must be responsible for the well-known EEM limitations. One could hope that the spherical atom EEM solves these issues, but similar empirical EEM modifications were already proposed earlier and have little effect on the polarizability scaling or the dissociation limits.[100,120] Moreover, the spherical atom EEM does not contain a new energy term that is similar to the bond-hardness term, which is responsible for the correct polarizability scaling limit of SQE.[57]

The fundamental problem is that there is no satisfactory semi-local approximation for $E_{\text{txc}}[\rho]$. Especially the kinetic energy is problematic. Accurate DFT computations always rely on the Kohn-Sham model for the kinetic energy, which is not a local functional of the density. It is also observed that the Kohn-Sham contribution to the linear response kernel already describes the larger part of the static electronic polarizability.[121–123] It was even suggested by Cioslowski that one could use the atom-condensed Kohn-Sham response kernel to construct an alternative EEM-like model with improved linear response properties.[99]

We conclude that the derivation of the (spherical atom) EEM is only accurate for certain parts of the energy functional, i.e. the Hartree term, the interaction with the external potential and those parts of $E_{\text{txc}}[\rho]$ for which good semi-local approximations are available. For all other contributions, a new approach is inevitable to obtain a significantly improved

model.

D. ACKS2

Based on the analysis in Sec. II C, we will now start from Kohn-Sham DFT:[124]

$$E_v^{\text{KS}}[\rho] = \frac{1}{2} \iint \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' + E_{\text{xc}}[\rho] + T_s[\rho] + \int \rho(\mathbf{r})v(\mathbf{r})d\mathbf{r} \tag{52}$$

with

$$\rho(\mathbf{r}) = \sum_{i \in \text{Occ.}} |\phi_i(\mathbf{r})|^2 \tag{53}$$

$$T_s[\rho] = -\frac{1}{2} \int \sum_{i \in \text{Occ.}} \phi_i(\mathbf{r})\nabla^2\phi_i(\mathbf{r})d\mathbf{r} \tag{54}$$

In these equations, $T_s[\rho]$ is the Kohn-Sham model for the kinetic energy. $E_{\text{xc}}[\rho]$ contains mainly contributions from exchange and correlation functionals, but also from the difference between the exact kinetic energy functional and the Kohn-Sham approximation to it. Reasonable semi-local models for the exchange-correlation functional are available, e.g. GGA approximations such as PBE[125] and BLYP.[126,127] Therefore, all terms but the Kohn-Sham kinetic energy can be approximated with the spherical atom EEM from Sec. II C. In this subsection, a novel second-order expansion is developed for the Kohn-Sham kinetic energy. As a result of the derivation below, ACKS2 will be defined as a new approximation of the molecular energy, see Eq. (22):

$$\begin{aligned} E_{\text{mol}} \approx E_{\text{ACKS2}}(\Delta_1, \ldots, \Delta_M) = E_{\text{ref}} + E_{nn} + \sum_A \Delta_A\mu_{\text{e},A} \\ + \frac{1}{2}\sum_{AB} \Delta_A\Delta_B\eta_{\text{e},AB} \\ + T_{\text{ACKS2}}(\Delta_1, \ldots, \Delta_M) \end{aligned} \tag{55}$$

The first term only depends on the nuclear coordinates (and not on relative populations), so it can be omitted if one is only interested in the charge distribution. The second, third and fourth term are the spherical EEM approximation of the KS-DFT energy functional without the kinetic energy contribution. Hence, the parameters $\mu_{\text{e},A}$ and $\eta_{\text{e},AB}$ are obtained with Eqs. (39), (42) and (43) applied to $E_v^{\text{KS}}[\rho] - T_s[\rho]$ instead of applying them to the HK energy functional in Eq. (28). The last term is the new model for the kinetic energy that will be derived below.

20

In order to construct a second-order expansion of the kinetic energy in terms of atomic populations, we consider the energy of non-interacting fermions in a Kohn-Sham-like external potential u.

$$E_u^\circ[\rho] = T_s[\rho] + \int \rho(\mathbf{r})u(\mathbf{r})d\mathbf{r} \tag{56}$$

where the superscript $^\circ$ is used to denote the energy of non-interacting fermions. For a given Kohn-Sham potential, u, and a fixed number of non-interacting fermions, N, the ground state energy is lower or equal to the energy of any N-particle trial density, $\tilde{\rho}$:

$$E_{gs}^\circ[u, N] \leq T_s[\tilde{\rho}] + \int \tilde{\rho}(\mathbf{r})u(\mathbf{r})d\mathbf{r} \tag{57}$$

Conversely, for a fixed N-particle ground state density ρ, the same inequality holds for any trial potential $\tilde{u}$. After rearranging some terms we obtain a dual inequality

$$T_s[\rho] \geq E_{gs}^\circ[\tilde{u}, N] - \int \rho(\mathbf{r})\tilde{u}(\mathbf{r})d\mathbf{r}, \tag{58}$$

which can also be written as

$$T_s[\rho] = \sup_{\tilde{u}} \left[E_{gs}^\circ[\tilde{u}, N] - \int \rho(\mathbf{r})\tilde{u}(\mathbf{r})d\mathbf{r} \right]. \tag{59}$$

A supremum must be used instead of a maximum because the density ρ may not be v-representable. The Kohn-Sham kinetic energy can thus also be interpreted as the Legendre transform[96] of the ground state energy of the N non-interacting fermions. This approach is analogous to Lieb's formulation of DFT.[97,98] The Legendre transform of the energy of a system of non-interacting fermions was already introduced in the adiabatic connection in DFT[128–130] and was recently also implemented in a numerical algorithm by Wu and Yang.[131] We are interested in the Kohn-Sham kinetic energy of the population-constrained density, ρ_{pop}, which is always v-representable. One may thus write:

$$T_s[\rho_{pop}] = \max_{\tilde{u}} \left[E_{gs}^\circ[\tilde{u}, N] - \int \rho_{pop}(\mathbf{r})\tilde{u}(\mathbf{r})d\mathbf{r} \right] \tag{60}$$

We will now introduce the essential approximation: the trial potential is expanded in a limited basis that is compatible with the AIM partitioning scheme in Eq. (2):

$$\tilde{u}(\mathbf{r}) = u_{ref}(\mathbf{r}) + \sum_A U_A w_A(\mathbf{r}) \tag{61}$$

where $u_{ref}(\mathbf{r})$ is the reference Kohn-Sham potential, i.e. the one that leads to the reference density $\rho_{ref}(\mathbf{r}) = \rho_{pop}(N_1^0, \ldots, N_M^0; \mathbf{r})$, and the basis functions, $w_A(\mathbf{r})$, are the dimensionless

atomic weight functions from the partitioning scheme. The scalar variable U_A can be interpreted as the amplitude of the change in Kohn-Sham potential felt by the electrons in atom A. This is an approximation because one cannot guarantee that this limited potential basis set is sufficient to reproduce exactly $\rho_{\text{pop}}(\mathbf{r})$ of the interacting fermion system. However, the basis is sufficient to reproduce the essential characteristics of ρ_{pop}: by tuning the coefficients U_A, one can control the atomic populations of the non-interacting ground state, in the same way as the Lagrange multipliers in Eq. (5) affect the populations of the density ρ_{pop}. The limited basis for the potential leads to a systematic underestimation of the kinetic energy. The density ρ_{pop} is also expanded to first order, in analogy with the derivation of the EEM model:

$$\rho_{\text{pop}}(\mathbf{r}) = \rho_{\text{ref}}(\mathbf{r}) + \sum_A \Delta_A f_A(\mathbf{r}) \tag{62}$$

With this ansatz we may approximate the kinetic energy as a maximization:

$$T_{\text{s}}[\rho_{\text{pop}}] \approx \max_{\substack{U_* \\ \sum_A U_A = 0}} \left[E_{\text{gs}}^{\circ}[\tilde{u}, N] - \int \rho_{\text{pop}}(\mathbf{r})\tilde{u}(\mathbf{r})d\mathbf{r} \right] \tag{63}$$

A constraint is added to fix the sum of the coefficients U_A. A change in the sum of U_A would only lead to a trivial change of the reference value of the Kohn-Sham-like potential $\tilde{u}$ because the sum of all weight functions is unity. The ACKS2 kinetic energy is defined by expanding the part between square brackets to second order in Δ_A and U_A:

$$T_{\text{ACKS2}}(\Delta_1, \ldots, \Delta_M) = \max_{\substack{U_* \\ \sum_A U_A = 0}} \left[\frac{1}{2} \sum_{AB} U_A X_{\text{s},AB} U_B + \sum_A (\mu_{\text{s},A} - U_A)\Delta_A \right] \tag{64}$$

where Kohn-Sham contribution to the atomic electronegativity parameter is

$$\mu_{\text{s},A} = - \int f_A(\mathbf{r})u_{\text{ref}}(\mathbf{r})d\mathbf{r} \tag{65}$$

and the matrix of second order coefficients

$$X_{\text{s},AB} = \iint X_{\text{s}}[u_{\text{ref}}](\mathbf{r}, \mathbf{r}')w_A(\mathbf{r})w_B(\mathbf{r}')d\mathbf{r}d\mathbf{r}' \tag{66}$$

is the condensed form of the the linear response kernel of the non-interacting fermion system in a fixed Kohn-Sham potential, u_{ref}. This kernel is also called the *non-interacting* or *Kohn-Sham* linear response kernel.[123,132,133] The simple form cross term, $-U_A \Delta_A$, originates from the expansion of the trial potential in AIM weight functions in Eq. (61).

22

When the ground state density is used as a reference for the second-order expansion, the total atomic chemical potential parameter, $\mu_A = \mu_{e,A} + \mu_{s,A}$, is equal to the molecular chemical potential. Hence, when using the ground state reference, there is little reason to compute the parameters $\mu_{e,A}$ and $\mu_{s,A}$ separately. For any other reference, e.g. a neutral reference, these parameters can no longer be ignored. One may construct transferable values for $\mu_{e,A}$ as outlined in the derivation of the spherical atom EEM (vide supra). There is no analogous approach to model the parameters $\mu_{s,A}$ in a transferable way. However, one can compute these parameters for a given molecular reference density, ρ_{ref}. The potential u_{ref} can be derived from ρ_{ref} using the the constrained search method of Wu and Yang,[131] which can then be used to evaluate Eq. (65).

The non-interacting linear response matrix can be derived from a conventional KS-DFT computation by applying perturbation theory on the Kohn-Sham wavefunction. In terms of Kohn-Sham orbitals, ϕ_i, and orbital energies, ϵ_i, of the reference system, one writes:[123,133]

$$X_{s,AB} = \sum_{\substack{i \in \text{Occ.} \\ j \in \text{Virt.}}} \frac{\langle \phi_i | w_A | \phi_j \rangle \langle \phi_j | w_B | \phi_i \rangle}{\epsilon_i - \epsilon_j} + \text{c.c.} \tag{67}$$

where "c.c." stands for the complex conjugate of the preceding term. Cioslowski showed several properties of this matrix, but used a different sign convention and a slightly unusual name, i.e. *atom-condensed softness matrix*.[99] We will refer to this matrix as the KS response matrix. This matrix is negative semidefinite and one can also show that:

$$\sum_A X_{s,AB} = 0 \tag{68}$$

Hence, the response matrix has (at least) one zero eigenvalue with corresponding eigenvector d (with $d_A = 1$). This means that the density of the reference system does not respond to a constant shift of the external potential.

In order to use the ACKS2 kinetic energy as an empirical model, one should at least propose convenient expressions for the KS response matrix elements that are (simple) functions of the molecular internal coordinates. In the literature, the dependence of the KS response matrix on the molecular structure is not yet extensively studied for a diverse set of chemical compounds or solid state systems. Nevertheless, some relations with molecular geometry are available. Based on a simple model of orbital overlap, it was shown that the off-diagonal matrix elements must decay exponentially with inter-atomic distance in the limit

of large separations.[58] One has also derived an approximate relation between the Wiberg bond order[134,135] and the KS linear response matrix, which is based on Unsöld's theorem:[136]

$$X_{\mathrm{s},AB} \approx -\frac{\mathrm{BO}_{AB}}{\langle \Delta E_{AB} \rangle} \tag{69}$$

with

$$\mathrm{BO}_{AB} = \sum_{\substack{i\in\mathrm{Occ.}\\j\in\mathrm{Occ.}}} \langle \phi_i | w_A | \phi_j \rangle \langle \phi_j | w_B | \phi_i \rangle + \mathrm{c.c.} \tag{70}$$

The factor $\langle \Delta E_{AB} \rangle$ is a weighted average of the denominators in Eq. (67). This relation implies that an off-diagonal KS response element will be nonzero when it corresponds to a pair of covalently bonded atoms. The opposite, i.e. that distant pairs of atoms have a negligible corresponding KS response matrix element, is not always valid when the atoms are in the same molecule. A detailed analysis of the KS response matrix[99,137,138] shows that the matrix elements decay exponentially with the number of bonds between the atoms in alkanes. However, in conjugated alkenes, the matrix elements show an oscillatory (and slowly decaying) trend as function of the number of bonds between the atoms. Also in aromatic species, the relation between the KS response elements and the inter-atomic distance is non-trivial. Based on these observations, Sablon et al. interpret the KS response matrix as a measure for electron delocalization.[137,138] At least for some systems, e.g. linear alkanes, there is numerical evidence that the KS response matrix is near-sighted. In the remainder of the paper, we will only rely on the exponential decay of the KS response matrix elements for pairs of atoms in different molecules. Nevertheless, it is clear that more insight is needed in the relation between molecular structure and the properties of KS response matrix.

In the remainder of the paper, the following definition of *the ACKS2 energy*, in line with Eq. (55), will be used:

$$E_{\mathrm{ACKS2}}(\Delta_1,\ldots,\Delta_M) = E_{\mathrm{ref}} + E_{nn} + \sum_A \Delta_A \mu_A + \frac{1}{2}\sum_{AB}\Delta_A\Delta_B\eta_{\mathrm{e},AB}$$
$$+ \max_{\substack{U_*\\\sum_A U_A=0}}\left[-\sum_A U_A\Delta_A + \frac{1}{2}\sum_{AB}U_A X_{\mathrm{s},AB}U_B \right] \tag{71}$$

where the atomic chemical potential parameter, $\mu_A = \mu_{\mathrm{e},A} + \mu_{\mathrm{s},A}$, consists of an EEM-like contribution ($\mu_{\mathrm{e},A}$) and a contribution due to the Kohn-Sham kinetic energy ($\mu_{\mathrm{s},A}$). This definition includes the maximization over the variables U_A. The ACKS2 ground state

is found by minimizing the ACKS2 energy with a constraint on the sum of the relative populations:

$$E_{\text{ACKS2,gs}} = E_{\text{ref}} + E_{nn} + \min_{\substack{\Delta_* \\ \sum_A \Delta_A=0}} \max_{\substack{U_* \\ \sum_A U_A=0}} \left[\sum_A \Delta_A \mu_A - \sum_A U_A \Delta_A \right. $$
$$\left. + \frac{1}{2} \sum_{AB} \Delta_A \Delta_B \eta_{\text{e},AB} + \frac{1}{2} \sum_{AB} U_A X_{\text{s},AB} U_B \right] \quad (72)$$

The sum of all Δ_A is constrained to zero. This means that the total electronic population of the system is constrained to the sum of the reference populations. The main motivation for this choice is that the Kohn-Sham kinetic energy as function of the total population contains derivative discontinuities. Quadratic models like EEM or ACKS2 can never reproduce these discontinuities, and therefore any deviation of the total population from the reference point will yield erratic results. As will be shown in Sec. III, enforcing $\sum_A \Delta_A = 0$ is also essential to obtain proper dissociation limits.

III. DISSOCIATION LIMITS

The dissociation limits will be studied with a coarse-grained model for the molecular charge distribution. Consider a molecule that dissociates into two fragments. Only the total population of each fragment is modeled as the molecule dissociates. This approach is used to study the fragment populations at large separation with both EEM and ACKS2.

A. ACKS2 in terms of fragment populations.

The fragment-condensed ACKS2 model is first developed. An EEM variant is obtained by omitting the kinetic energy term from the ACKS2 model. The two fragments are defined through sets of atoms α and β. The ACKS2 energy in terms of the relative fragment populations, see Eq. (71), is defined as:

$$E_{\text{ACKS2}}(\Delta_\alpha, \Delta_\beta) = \min_{\substack{\Delta_* \\ \left(\sum_{A\in\alpha} \Delta_A=\Delta_\alpha\right) \\ \left(\sum_{B\in\beta} \Delta_B=\Delta_\beta\right)}} E_{\text{ACKS2}}(\Delta_1, \dots, \Delta_M) \quad (73)$$

The fragment-condensed ACKS2 model can be derived in the same way as the atom-condensed ACKS2 model, based on fragment weight functions instead of atomic weight

functions:

$$w_\alpha(\mathbf{r}) = \sum_{A \in \alpha} w_A(\mathbf{r}) \tag{74}$$

$$w_\beta(\mathbf{r}) = \sum_{B \in \beta} w_B(\mathbf{r}) \tag{75}$$

The only limitation is that one cannot approximate molecular fragments as spherical charge distributions. Therefore, the fragment-constrained ACKS2 is only useful for fragments that are well separated, such that they do not overlap and only the monopole term of the multipole expansion of each fragment is significant for the electrostatic interactions. Based on Eq. (67), the linear response matrix elements of the fragments are simple sums over atomic matrix elements:

$$X_{\mathrm{s},\alpha\beta} = \sum_{A \in \alpha} \sum_{B \in \beta} X_{\mathrm{s},AB} \tag{76}$$

In contrast with the final paragraph of Sec. II D, $\Delta_\alpha + \Delta_\beta = \Delta_{\mathrm{tot}}$ is initially not forced to be zero, just to show that nonzero values of Δ_{tot} may lead to undesirable properties of the ACKS2 model.

B. Reaction coordinate.

The distance between the centers of mass of both fragments, x, is used as the reaction coordinate for the dissociation. For large values of x (compared to the size of the fragments), one can approximate all relevant parameters in the ACKS2 model as simple functions of x. The EEM parameters take the following form:

$$E_{\mathrm{nn}} \approx \frac{q_\alpha^0 q_\beta^0}{x} \quad \text{where} \quad q_\alpha^0 = \sum_{A \in \alpha}(Z_A - N_A^0)$$

$$\text{and} \quad q_\beta^0 = \sum_{B \in \beta}(Z_B - N_B^0) \tag{77}$$

$$\mu_\alpha \approx \mu_\alpha^0 - \frac{q_\beta^0}{x} \quad \text{where} \quad \mu_\alpha^0 = \text{const.} \tag{78}$$

$$\mu_\beta \approx \mu_\beta^0 - \frac{q_\alpha^0}{x} \quad \text{where} \quad \mu_\beta^0 = \text{const.} \tag{79}$$

$$\eta_{\mathrm{e},\alpha\alpha} \approx \eta_\alpha^0 = \text{const.} \tag{80}$$

$$\eta_{\mathrm{e},\beta\beta} \approx \eta_\beta^0 = \text{const.} \tag{81}$$

$$\eta_{\mathrm{e},\alpha\beta} \approx \frac{1}{x} \tag{82}$$

The reference energy is not considered here, because it does not affect the charge distribution. Hence, one obtains only the charge-transfer and electrostatic contributions to the interaction energy in this section.

As discussed extensively in Sec. II D, there are strong indications in the literature that, at large separations, the off-diagonal Kohn-Sham linear response matrix element of the two fragments decays exponentially with the separation, x:[58,139,140]

$$X_{s,\alpha\beta} \approx X_0 \exp(-x/\tau) \tag{83}$$

where the rate of decay, τ, and the amplitude, X_0, are constant parameters. It will be shown that this model for the long-range behavior of $X_{s,\alpha\beta}$ is essential to obtain the integer dissociation limits for the populations. Because of Eq. (68) the following equalities hold for any two-fragment system:

$$X_{s,\alpha\alpha} = X_{s,\beta\beta} = -X_{s,\alpha\beta} = -X_{s,\beta\alpha} \tag{84}$$

C. Dissociation in EEM.

For EEM it is sufficient to consider only a neutral reference, i.e. $q_\alpha^0 = q_\beta^0 = 0$. It was shown in Sec. II C that the EEM energy with a non-neutral reference can always be rewritten as an EEM energy with a neutral reference. The EEM ground state energy of the two separated fragments becomes:

$$E_{\text{EEM,gs}} = \min_{\substack{\Delta_\alpha, \Delta_\beta \\ \Delta_\alpha + \Delta_\beta = \Delta_{\text{tot}}}} \left[\Delta_\alpha \mu_\alpha^0 + \Delta_\beta \mu_\beta^0 + \frac{1}{2} \left(\Delta_\alpha^2 \eta_\alpha^0 + \Delta_\beta^2 \eta_\beta^0 + \frac{2\Delta_\alpha \Delta_\beta}{x} \right) \right] \tag{85}$$

The total charge constraint can be satisfied implicitly through the following substitutions:

$$\Delta_\alpha = \frac{\Delta_{\text{tot}}}{2} + \Delta_d \qquad\qquad \Delta_\beta = \frac{\Delta_{\text{tot}}}{2} - \Delta_d \tag{86}$$

where Δ_d is an unconstrained degree of freedom. After substitution and solving for Δ_d, one obtains the following:

$$\Delta_d = -\frac{\mu_\alpha^0 - \mu_\beta^0 + \frac{\Delta_{\text{tot}}}{2}(\eta_\alpha^0 - \eta_\beta^0)}{\eta_\alpha^0 + \eta_\beta^0 - \frac{2}{x}} \tag{87}$$

where the term $\frac{2}{x}$ vanishes in the limit of large separation. In general, the relative population in the dissociation limit is a fractional number, as often reported in the

27

literature,[59,63,64,141–143] while molecules are known to dissociate into integer-charged fragments.[60,62,144] Even if the parameters would be crafted such that the limit becomes integer, the convergence to the limit goes like $1/x$. Such slowly decaying intermolecular charge transfer is not found in ab initio computations.[58] EEM is clearly not capable of describing bond breaking/formation correctly.

D. Dissociation in ACKS2.

In the derivation below we must also consider a non-neutral reference. The ACKS2 ground state energy (see Eq. (72)) of the two-fragment system becomes:

$$E_{\text{ACKS2,gs}} - E_{\text{ref}} = \min_{\substack{\Delta_\alpha,\Delta_\beta \\ \Delta_\alpha+\Delta_\beta=\Delta_{\text{tot}}}} \max_{\substack{U_\alpha,U_\beta \\ U_\alpha+U_\beta=0}} \left[E_{\text{nn}} + \Delta_\alpha\mu_\alpha + \Delta_\beta\mu_\beta - \Delta_\alpha U_\alpha - \Delta_\beta U_\beta \right.$$
$$\left. + \frac{1}{2}\left(\Delta_\alpha^2\eta_\alpha^0 + \Delta_\beta^2\eta_\beta^0 + \frac{2\Delta_\alpha\Delta_\beta}{x}\right) - \frac{1}{2}(U_\alpha^2 + U_\beta^2 - 2U_\alpha U_\beta)X_0\exp(-x/\tau) \right] \quad (88)$$

Both constraints can be satisfied implicitly through the following substitutions:

$$\Delta_\alpha = \frac{\Delta_{\text{tot}}}{2} + \Delta_d \qquad\qquad \Delta_\beta = \frac{\Delta_{\text{tot}}}{2} - \Delta_d \qquad\qquad (89)$$

$$U_\alpha = U_d \qquad\qquad U_\beta = -U_d \qquad\qquad (90)$$

After some trivial rearrangements, one obtains solutions that converge exponentially in the separation limit:

$$U_d = \frac{\tilde{\mu}}{2 + 2\tilde{\eta}X_0\exp(-x/\tau)} \qquad\qquad (91)$$

$$\Delta_d = -\frac{\tilde{\mu}}{(X_0\exp(-x/\tau))^{-1} + \tilde{\eta}} \qquad\qquad (92)$$

where

$$\tilde{\eta} = \eta_\alpha^0 + \eta_\beta^0 - 2/x \qquad\qquad (93)$$

$$\tilde{\mu} = \mu_\alpha^0 - \mu_\beta^0 + \frac{q_\alpha^0 - q_\beta^0}{x} + \frac{\Delta_{\text{tot}}}{2}(\eta_\alpha^0 - \eta_\beta^0) \qquad\qquad (94)$$

For large x, U_d may be approximated as:

$$\lim_{x\to\infty} U_d = \frac{\tilde{\mu}}{2} \qquad\qquad (95)$$

which is a very intuitive result: at large separation the difference in the chemical potential ($\tilde{\mu}$) is compensated by the difference in the Kohn-Sham potential ($2U_d$), leaving no incentive for charge transfer. One should keep in mind that the difference in electronegativity of two systems at infinite separation becomes ill-defined.[145]

At large separation, Δ_d may be approximated as:

$$\lim_{x\to\infty} \Delta_d = \lim_{x\to\infty} -\tilde{\mu} X_0 \exp(-x/\tau) = 0 \tag{96}$$

The populations become:

$$N_\alpha = N_\alpha^0 + \Delta_\alpha = N_\alpha^0 + \frac{\Delta_{\text{tot}}}{2} - \frac{\tilde{\mu}}{(X_0 \exp(-x/\tau))^{-1} + \tilde{\eta}} \tag{97}$$

$$N_\beta = N_\beta^0 + \Delta_\beta = N_\beta^0 + \frac{\Delta_{\text{tot}}}{2} + \frac{\tilde{\mu}}{(X_0 \exp(-x/\tau))^{-1} + \tilde{\eta}} \tag{98}$$

For large x, the last term vanishes and the populations become constant. In order to guarantee integer fragment populations in the dissociation limit, it is in principle sufficient to choose parameters such that $N_\alpha^0 + \frac{\Delta_{\text{tot}}}{2}$ and $N_\beta^0 + \frac{\Delta_{\text{tot}}}{2}$ are integers. However, that would only work when a molecule is broken into two fragments. Proper dissociation limits in general may only work when $\Delta_{\text{tot}} = 0$ and when the reference populations of the fragments are integers. A correct dissociation behavior is of fundamental importance when studying reactions. Therefore, one should always respect these rules for the reference populations and total population when using ACKS2.

When $\Delta_{\text{tot}} = 0$ and the reference populations are integers, we can derive a simple approximation for the inter-fragment potential due to charge-transfer and electrostatic interactions. The remaining terms in the ground state energy are:

$$E_{\text{ACKS2,gs}} - E_{\text{ref}} = \min_{\Delta_d} \max_{U_d} \left[E_{\text{nn}} + \Delta_d \tilde{\mu} - 2\Delta_d U_d + \frac{\Delta_d^2}{2} \tilde{\eta} - 2U_d^2 X_0 \exp(-x/\tau) \right] \tag{99}$$

Because we are only interested in the long-range behavior, the following approximate solution for the ground state,

$$U_d \approx \frac{\tilde{\mu}}{2} \tag{100}$$

$$\Delta_d \approx -\tilde{\mu} X_0 \exp(-x/\tau), \tag{101}$$

and Eqs. (77), (78) and (79) for the parameters are substituted. The ground state energy becomes:

$$E_{\text{ACKS2,gs}} - E_{\text{ref}} \approx \frac{q_\alpha^0 q_\beta^0}{x} + \frac{\tilde{\mu}^2 X_0}{2} \left(-\exp(-x/\tau) + \tilde{\eta} X_0 \exp(-2x/\tau) \right) \tag{102}$$

29

At sufficiently large distances the repulsive exponential term can be neglected:

$$E_{\text{ACKS2,gs}} - E_{\text{ref}} \approx \frac{q_\alpha^0 q_\beta^0}{x} - \frac{\tilde{\mu}^2 X_0}{2} \exp(-x/\tau) \tag{103}$$

This result shows that the binding energy due to charge-transfer is proportional to the Kohn-Sham linear response matrix element and the difference in intrinsic chemical potential squared. This is analogous to the charge-transfer interaction first described by Murrel,[146] which was later modeled empirically with force fields,[147] and which is now also one of the contributions in the SIBFA model.[148] When the dissociated fragments are not neutral, a trivial Coulomb term is also present. When the atomic structure of each fragment is described explicitly, the interaction potential will also contain terms due to the higher-order multipoles of each fragment.

The dissociation of two fragments in ACKS2 can also be compared with earlier studies of the energy due to charge transfer through a chemical bond as the bond breaks. It was observed that the bond hardness of a diatomic molecule, i.e. the second-order derivative of the energy towards the charge transfer through the bond, diverges exponentially as the bond is broken.[58] One can obtain a similar result by elimination of the variable U_d from Eq. (99). One obtains the following energy as function of the charge transfer, Δ_d:

$$E_{\text{ACKS2}}(\Delta_d) - E_{\text{ref}} = E_{\text{nn}} + \Delta_d \tilde{\mu} + \frac{\Delta_d^2}{2} \left(\tilde{\eta} + (X_0 \exp(-x/\tau))^{-1} \right) \tag{104}$$

The second-order coefficient also increases exponentially as the atoms are separated.

E.   Numerical example.

Fig. 3 compares the dissociation properties of the EEM and ACKS2 models with reference data from population constrained CASSCF/6-311++G** computations for hydrogen fluoride taken from Ref. 58. The plotted properties are (a) the charge on the hydrogen atom [see Eq. (92)], (b) the energy due to charge transfer [substitution of Eq. (92) in Eq. (104)], and (c) the second-order derivative of the energy with respect to charge transfer [see second-order coefficient in Eq. (104)]. The EEM expressions for these parameters are formally recovered by taking the limit of the ACKS2 expressions for X_0 towards infinity. CASSCF data is used for comparison because this method is capable of describing the electronic structure along the entire bond dissociation path, as opposed to conventional single-reference DFT computations.

The ACKS2 parameters were estimated by minimizing the mismatch with the CASSCF results for distances larger than 2 Å. The reference populations are set to zero, as these are the expected populations at infinite separation. In the limit of large separations, ACKS2 agrees quantitatively with the CASSCF reference data. It is impossible to estimate the EEM parameters in a similar fashion. The optimal EEM parameters are completely unphysical and still result in a large disagreement between the EEM results and the CASSCF data. Instead, experimental chemical potential and hardness values of isolated atoms[102] were used as EEM parameters. All parameters are given in Table I. The estimated ACKS2 chemical potential and hardness parameters are close to the experimental values of Parr and Pearson.[102] This is remarkable because earlier studies revealed that calibrated EEM parameters have no profound physical meaning: they may vary widely for a given element, depending on the reference data, and hence they do not correlate with the numbers of Parr and Pearson.[18,77] This numerical example demonstrates that the ACKS2 model can, unlike EEM, quantitatively describe charge transfer between dissociated molecular fragments, using parameters that are physically reasonable.

It is also noteworthy that the assumption of a simple exponential form for the $X_{s,\alpha\beta}$ starts to fail as soon as bond length becomes too short, i.e. close the Coulson-Fischer point where the triplet state is no longer the only relevant contribution to the CASSCF wavefunction. At distances shorter than the Coulson-Fischer point, the singlet state becomes dominant and a different exponential ansatz may be used for the parameter $X_{s,\alpha\beta}$.

IV. IMPLEMENTATION

This section discusses the set of equations that must be solved to obtain the ACKS2 ground state. The similarity with the conventional EEM equations and some technical aspects of their implementation are also briefly discussed. The following matrix notation is used to keep the equations compact.

- $\eta \in \mathbb{R}^{M \times M}$ is the matrix with second-order EEM parameters,

- $\eta_e \in \mathbb{R}^{M \times M}$ is the matrix with second-order EEM parameters excluding the kinetic energy contribution,

- $\Delta \in \mathbb{R}^{M \times 1}$ is a column vector with the relative populations ($\Delta = N - N^0$),

- $\mu \in \mathbb{R}^{M \times 1}$ is a column vector with the first-order EEM parameters.

- $X_{\mathrm{s}} \in \mathbb{R}^{M \times M}$ is the linear response matrix of the reference system,

- $U \in \mathbb{R}^{M \times 1}$ is a column vector with expansion coefficients of the Kohn-Sham potential,

- $d \in \mathbb{R}^{M \times 1}$ is a column vector with all elements equal to 1, and

- I_M is the identity matrix of size M,

The equations below are written in terms of relative populations. Equivalent equations in terms of atomic charges are provided as supplementary material.[149]

A.   EEM

Atomic populations can be computed with EEM by minimizing the EEM energy with a constraint on the total charge. In this paper, we will write this total-charge constraint as a constraint on the sum of all Δ_A variables. The bound minimum is a stationary point of the following Lagrangian:

$$L_{\mathrm{EEM}} = E_{\mathrm{EEM}} - \mu_{\mathrm{mol}} \left(\sum_A \Delta_A - \Delta_{\mathrm{tot}} \right) \tag{105}$$

where the Lagrange multiplier is also known as the equalized chemical potential. Setting the derivatives of the Lagrangian towards Δ_C and μ_{mol} equal to zero, yields the so-called EEM equations. In block matrix notation these equations become:

$$- \begin{bmatrix} \eta & -d \\ \hline -d^T & 0 \end{bmatrix} \begin{bmatrix} \Delta \\ \hline \mu_{\mathrm{mol}} \end{bmatrix} = \begin{bmatrix} \mu \\ \hline \Delta_{\mathrm{tot}} \end{bmatrix} \tag{106}$$

where the square matrix in the left-hand side ($\in \mathbb{R}^{M+1 \times M+1}$) is called the *EEM matrix*. These equations can either be solved directly with standard routines from the LAPACK library or with an iterative algorithm. Because the solution is a saddle point[150] (minimization over relative populations and maximization over the Lagrange multiplier) a simple conjugate-gradient algorithm is not a suitable. A biconjugate gradient solver or more advanced generalized minimum residual method should be used instead.[151]

B. ACKS2

The computation of atomic charges with ACKS2 is completely analogous. Note that the sum of the variables Δ_A is constrained to zero in order to guarantee proper dissociation limits. The ACKS2 energy already contains a bound maximization problem, leading to the following Lagrangian with two Lagrange multipliers.

$$
\begin{aligned}
L_{\text{ACKS2}} = &\sum_A \left(\mu_A \Delta_A - U_A \Delta_A \right) \\
&+ \frac{1}{2} \sum_{AB} \left(\Delta_A \Delta_B \eta_{\text{e},AB} + U_A U_B X_{\text{s},AB} \right) \\
&- \mu_{\text{mol}} \sum_A \Delta_A - \lambda_U \sum_A U_A
\end{aligned}
\tag{107}
$$

Setting the derivatives of this Lagrangian towards Δ_C, μ_{mol}, U_C and λ_U equal to zero, yields a set of linear equations that must be solved to determine the ACKS2 ground state. This linear system can be written in block matrix notation as follows:

$$
-\begin{bmatrix}
\eta_{\text{e}} & -d & -I_M & 0 \\
-d^T & 0 & 0 & 0 \\
-I_M & 0 & X_{\text{s}} & -d \\
0 & 0 & -d^T & 0
\end{bmatrix}
\begin{bmatrix}
\Delta \\
\mu_{\text{mol}} \\
U \\
\lambda_U
\end{bmatrix}
=
\begin{bmatrix}
\mu \\
0 \\
0 \\
0
\end{bmatrix}
\tag{108}
$$

The total size of the *ACKS2 matrix* is $(2M + 2 \times 2M + 2)$. One may, just as with the EEM equations, either use a LAPACK routine or an iterative solver. However, the latter is computationally especially attractive because at least two large block matrices in the ACKS2 equations, namely $-I_M$, are sparse. As discussed above, the Kohn-Sham response matrix is near-sighted for some systems. In that case the hardness matrix (without kinetic energy contributions), η_{e}, is the only dense submatrix in these equations.

V. DISCUSSION

Based on the results obtained so far, one can already discuss potential extensions and applications of ACKS2. This section briefly highlights the main perspectives, but also some limitations of the new model.

In the Appendix it is shown that SQE is a limiting case of ACKS2. SQE was extensively validated with the calibration of transferable SQE parameters for organic and inorganic systems.[12,28] These studies showed that SQE successfully describes both the equilibrium charge distribution and the response to an external field. Furthermore, the clear improvements of SQE compared to EEM were presented. Simply because ACKS2 is a generalization of SQE, these conclusions are also valid for ACKS2. However, one should be aware that ACKS2, just like any other PFF with quadratic energy terms, has some intrinsic limitations. A second-order expansion is only valid for small deviations from the reference point. As a safe guideline, one should not trust quadratic PFFs when the charge on an atom or a molecule changes by more than one electron. In such cases, one should observe a piecewise discontinuity of the energy derivative, which is not present in a quadratic model. This guideline implies that one cannot use quadratic polarizable models (with a single set of parameters) to describe redox reactions or different ions of a given molecule. While such guidelines may not be immediately obvious in the EEM context, ACKS2 imposes them explicitly: for every atom one must define a reference charge and the total charge of an isolated molecule is always equal to the sum of these reference charges. These rules are essential for the correct dissociation limits and linear response properties of ACKS2. We expect that future improved PFFs will attempt to combine all the beneficial ACKS2 properties without imposing these restrictions. Efforts along these lines can be found in the literature, which are mostly based on valence bond theory.[59,63,64,141–143] One has also proposed to model redox reactions within the SQE+Q^0 as discrete transitions of the reference populations.[70] This idea is also applicable to the ACKS2 model, provided that the discrete transition of a reference population is accompanied by an update of the related atomic parameters and reference energies.

Several attempts were made to get correct bond dissociation limits in electronegativity equalization model, which can be divided into two seemingly very different categories: some authors tried to model how the electronegativity difference between two atoms goes to zero

as they separate,[59,63,64] while others have tried to let the bond hardness diverge as atoms separate.[58,75,76] Based on our analysis in Sec. III, where it is shown that the difference in electronegativity is canceled out by a contribution from the kinetic energy term, one could argue that ACKS2 belongs in the first category. However, as soon as one eliminates the potential variables from the ACKS2 energy (see Eqs. (104) and (A.16)), the second category seems more appropriate: the correct dissociation limit is completely captured by the long-range behavior of the KS response matrix. We conclude that both visions are valid and that they are only different mathematical formulations of the same concept. In the first case one writes the Legendre transform of the kinetic energy as a maximization, while in the second, the solution of this maximization is substituted in the energy expression.

VI. CONCLUSIONS

A new semi-empirical model is proposed to estimate atomic partial charges and molecular linear response properties, i.e. "Atom-Condensed Kohn-Sham DFT approximated to second order" (ACKS2). The derivation is based on a new atom-condensed DFT formalism, which provides a novel connection between polarizable force fields (PFFs) and DFT. This formalism is first used to derive the EEM equations, starting from the Hohenberg-Kohn (HK) formalism. The approximations required for the derivation of EEM clearly reveal the origins of its well-known weaknesses: they are unavoidable consequences of the semi-local approximation for the non-classical terms in the HK energy, i.e. the kinetic, exchange, and correlation functionals. Especially for the kinetic energy, this is a severe over-simplification. Based on this analysis, we propose a new model for the kinetic energy, which is derived by expanding the Legendre transform of the Kohn-Sham kinetic energy to second order in the atomic populations and a new set of dual atomic variables, the relative atomic Kohn-Sham potentials. The second order coefficients associated with the new variables are matrix elements of the Kohn-Sham (non-interacting) response kernel. The ACKS2 model is then defined as the sum of the new quadratic kinetic energy model and an EEM approximation for all other terms in the KS-DFT energy functional.

ACKS2 solves two major shortcomings of EEM. Obviously, one must first acknowledge that EEM has been successfully applied to understand and predict the molecular charge distribution in terms of a minimal number of atomic parameters and a set of basic linear

equations. However, EEM also has two principal drawbacks: it assigns fractional charges to dissociated molecules and it predicts a cubic scaling of the dipole polarizability with system size. Both EEM weaknesses are extensively discussed in the literature and several ad hoc solutions were proposed. By means of analytical derivations, we show that ACKS2 surmounts both limitations. Furthermore, it is also shown that SQE is a limiting case of the ACKS2 model. The implementation of ACKS2 is very similar to the well-known EEM equations.

It is to be expected that future developments of semi-empirical atomic charge models will focus on the remaining limitations of the ACKS2 model. In analogy to similar extensions of EEM and SQE, one may include atomic inducible dipoles or higher-order multipoles. Beyond these trivial variations, one also faces more fundamental challenges. ACKS2 cannot describe redox reactions or different ions of a given molecule with a single set of parameters. These are inherent limitations of all quadratic PFFs.

The derivation of ACKS2 also provides expressions to estimate all the parameters from a KS-DFT computation as atoms-in-molecules (AIM) expectation values. These expression provide an alternative to the daunting calibration procedures that are commonly used to estimate parameters in PFFs. It will be tested in a future paper to what extent this new parameterization strategy leads to a quantitatively accurate ACKS2 model. The major road-block is that the structure of the Kohn-Sham linear response matrix is not yet extensively analyzed for a large variety of substances and materials. Nevertheless, thorough calibrations and validations of SQE for organic and inorganic systems have already shown that SQE, and by extension ACKS2, are both accurate models for the prediction of atomic charges and linear response properties.

ACKNOWLEDGMENTS

the reviewers for their helpful comments.

Appendix

Two new concepts are needed to show that SQE is a limiting case of ACKS2: (i) split-charge variables and (ii) an alternative form of the linear response matrix. The matrix notation from Sec. IV is used below to keep the notation compact.

Split charges. We will first introduce the split-charge variables as a set of redundant degrees of freedom for the relative populations:

$$\Delta_A = \sum_B{}' P_{AB} \tag{A.1}$$

where P_{AB} is the amount of electrons transfered from B to A. By convention we require $P_{AB} = -P_{BA}$, which guarantees that $\sum_A \Delta_A = 0$. This definition is actually a generalization of the original split-charge concept,[56,66] which becomes clear when the atomic charges are written in terms of split charges:

$$q_A = Z_A - N_A^0 - \sum_B{}' P_{AB} = q_A^0 + \sum_B{}' P_{BA} \tag{A.2}$$

The reference charge, q_A^0, is always zero in the SQE model, see e.g. Eq. (1) in Ref. 66. This restriction is not imposed here in analogy with the SQE+Q^0.[33] A similar generalization of the SQE was also proposed elsewere.[70] Reference charges are similar to precharges, which are used to augment bond-charge increments in fixed-charge models and are also needed in that context to allow local deviations from neutrality.[68]

In SQE applications, split charges are only allowed to be nonzero for a small subset of atom pairs, e.g. covalently bonded atom pairs:

$$L = \{(A, B) \mid \forall (A, B) \text{ if } P_{AB} \neq 0\} \tag{A.3}$$

The main argument for this restriction is that direct through-space charge transfer between non-bonded atoms is energetically infeasible.[58] The mathematical origin of this restriction will become obvious later in the derivation. The relation between relative populations and split charges becomes:

$$\Delta_A = \sum_{B \text{ if } (A,B) \in L} P_{AB} \tag{A.4}$$

which may be written in matrix notation as follows:

$$\Delta = -\mathcal{T}P \tag{A.5}$$

where P is a column vector containing all *allowed* $((A, B) \in L)$ and *independent* $(A > B)$ split charges, P_{AB}. Let K be the length of this split-charge vector. If split charges would be allowed between all atom pairs, we would have $K = M(M-1)/2$. In general, some split charges may be excluded and one has $0 < K \leq M(M-1)/2$. We will use the compound index k_{AB} for the elements of the split-charge vector, just to show that each element of P corresponds to a pair of atoms. The transfer matrix, $\mathcal{T} \in \mathbb{R}^{M \times K}$, is defined as:

$$\mathcal{T}_{C,k_{AB}} = \delta_{AC} - \delta_{BC} \tag{A.6}$$

Chen et al. introduced the transfer matrix in the context electronegativity equalization to transform split-charge models into atomic charge models and vice versa.[74]

Alternative form of the linear response matrix. Because of Eq. (68), we may decompose the KS response matrix into $M(M-1)/2$ terms, where each term is an $M \times M$ matrix with only four nonzero elements at positions (A, A), (A, B), (B, A) and (B, B):

$$X_{\mathrm{s}} = \sum_{A>B} \begin{bmatrix} \ddots & \vdots & & \vdots & \\ \cdots & -X_{\mathrm{s},AB} & \cdots & +X_{\mathrm{s},AB} & \cdots \\ & \vdots & \ddots & \vdots & \\ \cdots & +X_{\mathrm{s},AB} & \cdots & -X_{\mathrm{s},AB} & \cdots \\ & \vdots & & \vdots & \ddots \end{bmatrix} \tag{A.7}$$

This decomposition is inspired by Eq. (30) from Ref. 99. All terms corresponding to pairs (A, B) for which $X_{\mathrm{s},AB}$ is zero, may be omitted from Eq. (A.7). We will use this as a new (more formal) criterion to define the set L:

$$L = \{(A, B) \mid \forall (A, B) \text{ if } X_{\mathrm{s},AB} \neq 0\} \tag{A.8}$$

Using the transfer matrix, $\mathcal{T}$, we may now write the response matrix as:

$$X_{\mathrm{s}} = -\mathcal{T}R^2\mathcal{T}^T \tag{A.9}$$

where $R \in \mathbb{R}^{K \times K}$ is a diagonal matrix containing the root of the nonzero sub-diagonal elements of the response matrix:

$$R^2_{k_{AB},k_{AB}} = X_{\mathrm{s},AB} \tag{A.10}$$

In order to facilitate the derivation below, one must introduce an affine basis transformation for the split charges:

$$\tilde{P} = R^{-1}P \tag{A.11}$$

The expression for the relative populations and the response matrix can now be written as:

$$\Delta = \mathcal{T}R\tilde{P} = \mathcal{S}\tilde{P} \tag{A.12}$$

$$X_\mathrm{s} = -\mathcal{S}\mathcal{S}^T \tag{A.13}$$

Link between SQE+Q^0 and ACKS2. Using the ingredients introduced above, we will now show that the ACKS2 ground state is equivalent to the SQE+Q^0 ground state. Using matrix notation, the ACKS2 energy, defined in Eq. (71), minus the reference energy, becomes:

$$
\begin{aligned}
E_\mathrm{ACKS2}(\Delta) - E_\mathrm{ref} = E_\mathrm{nn} + \mu^T\Delta + \frac{1}{2}\Delta^T\eta_\mathrm{e}\Delta \\
+ \max_{U,d^TU=0}\left[\frac{1}{2}U^TX_\mathrm{s}U - U^T\Delta\right]
\end{aligned}
\tag{A.14}
$$

In analogy with the derivation of Eq. (51), the first three terms correspond to the EEM energy expression and will therefore be compactly written as $E_\mathrm{EEM}(\Delta)$. Because the SQE energy equals the EEM energy plus a sum of quadratic terms in the split charges,[66] we only have to show that the last term in Eq. (A.14) is equivalent to the bond hardness term of the SQE model. A particular solution U that maximizes the part between brackets of Eq. (A.14) may formally be written as:

$$U = X_\mathrm{s}^+\Delta \tag{A.15}$$

where X_s^+ is the Moore-Penrose pseudoinverse of X_s. After substitution of this solution into the energy expression, one gets:

$$E_\mathrm{ACKS2}(\Delta) - E_\mathrm{ref} = E_\mathrm{EEM}(\Delta) - \frac{1}{2}\Delta^TX_\mathrm{s}^+\Delta \tag{A.16}$$

This shows explicitly that the ACKS2 model is, just like EEM, a quadratic function of the relative populations. Next, the relative populations and the linear response matrix are substituted, using Eqs. (A.12) and (A.13):

$$E_\mathrm{ACKS2}(\tilde{P}) - E_\mathrm{ref} = E_\mathrm{EEM}(-\mathcal{S}\tilde{P}) + \frac{1}{2}\tilde{P}^T\mathcal{S}^T\left(\mathcal{S}\mathcal{S}^T\right)^+\mathcal{S}\tilde{P} \tag{A.17}$$

which may be written as

$$E_{\text{ACKS2}}(P) - E_{\text{ref}} = E_{\text{EEM}}(-\mathcal{S}\tilde{P}) + \frac{1}{2}\tilde{P}^T(\mathcal{S}^+\mathcal{S})^T(\mathcal{S}^+\mathcal{S})\tilde{P} \tag{A.18}$$

where $\mathcal{S}^+$ is the Moore-Penrose pseudoinverse of $\mathcal{S}$ and $\mathcal{S}^+\mathcal{S}$ is a projection on the image of $\mathcal{S}$. Let us introduce a separation of the split charge vector into two terms, $\tilde{P} = \tilde{P}_\parallel + \tilde{P}_\perp$, such that:

$$\tilde{P}_\parallel = (\mathcal{S}^+\mathcal{S})\tilde{P} \tag{A.19}$$

$$\tilde{P}_\perp = (I_K - \mathcal{S}^+\mathcal{S})\tilde{P} \tag{A.20}$$

Making use of the identity $\mathcal{S}\mathcal{S}^+\mathcal{S} = \mathcal{S}$, one may show that the relative populations are only determined by $\tilde{P}_\parallel$, and not by $\tilde{P}_\perp$:

$$\Delta = -\mathcal{S}\tilde{P} = -\mathcal{S}\mathcal{S}^+\mathcal{S}\tilde{P} = -\mathcal{S}\tilde{P}_\parallel \tag{A.21}$$

Hence, one can show that also the ACKS2 energy only depends on $\tilde{P}_\parallel$:

$$E_{\text{ACKS2}}(\tilde{P}_\parallel) - E_{\text{ref}} = E_{\text{EEM}}(-\mathcal{S}\tilde{P}_\parallel) + \frac{1}{2}\tilde{P}_\parallel^T\tilde{P}_\parallel \tag{A.22}$$

In order to make the connection with SQE+Q^0, one must consider the ground state energy:

$$E_{\text{ACKS2,gs}} - E_{\text{ref}} = \min_{\tilde{P}_\parallel}\left[E_{\text{EEM}}(-\mathcal{S}\tilde{P}_\parallel) + \frac{1}{2}\tilde{P}_\parallel^T\tilde{P}_\parallel\right] \tag{A.23}$$

and add additional terms that are zero, such that the ground state can be written as a minimization over the full split charge vector, $\tilde{P}$, as in the SQE model.[66] We make use of the following identities:

$$\mathcal{S}\tilde{P}_\perp = 0 \tag{A.24}$$

$$\tilde{P}_\perp^T\tilde{P}_\parallel = 0 \tag{A.25}$$

$$\min_{\tilde{P}_\perp}\left[\frac{1}{2}\tilde{P}_\perp^T\tilde{P}_\perp\right] = 0 \tag{A.26}$$

and use these to obtain the following form for the ACKS2 ground state energy:

$$
\begin{aligned}
E_{\text{ACKS2,gs}} - E_{\text{ref}} &= \min_{\tilde{P}_\parallel, \tilde{P}_\perp}\left[E_{\text{EEM}}(-\mathcal{S}\tilde{P}_\parallel) + \frac{1}{2}\tilde{P}_\parallel^T\tilde{P}_\parallel + \tilde{P}_\perp^T\tilde{P}_\parallel + \frac{1}{2}\tilde{P}_\perp^T\tilde{P}_\perp\right] \\
&= \min_{\tilde{P}}\left[E_{\text{EEM}}(-\mathcal{S}\tilde{P}) + \frac{1}{2}\tilde{P}^T\tilde{P}\right]
\end{aligned}
\tag{A.27}
$$

It is straightforward to rewrite the last term in scalar notation because the matrix R is diagonal. This leads to the familiar bond hardness term from the SQE model, e.g. see the first term of Eq. (16) in Ref. 66:

$$\frac{1}{2}\tilde{P}^T\tilde{P} = \frac{1}{2}P^T R^{-2} P = \frac{1}{2}\sum_{(A,B)\in L}\frac{P_{AB}^2}{R_{k_{AB},k_{AB}}^2} \tag{A.28}$$

The main difference with the original SQE model, is the generalized relation between the split charges and the atomic charges in Eq. (A.2).

This derivation also reveals the interpretation of the bond hardness parameters, $\kappa_{k_{AB}}$, in the SQE model. (See Eq. (16) in Ref. 66.)

$$\kappa_{k_{AB}} = \frac{1}{R_{k_{AB},k_{AB}}^2} = \frac{1}{X_{s,AB}} \tag{A.29}$$

Using Eq. (67), the bond hardness can be derived from a standard DFT computation, provided one has a method to construct the atomic weight functions. This expression for the bond hardness also shows why it is numerically infeasible to break bonds in the SQE model: based on Eq. (83), the bond hardness must diverge exponentially as atoms separate, as has been proposed in the literature.[58,75,76] This problem does not arise in the ACKS2 formalism because it does not explicitly use the inverse of the KS response matrix elements. Finally, by combining Eqs. (A.29) and (67), one can show that the bond hardness has a leading term that is proportional to the energy gap between the overlapping occupied and virtual orbitals at atom A and B. This is in line with the observation of Müser that the bond hardness parameter correlates with the band gap when the SQE model is used to describe solids.[70] The main result of this Appendix, Eq. (A.27), is not just applicable to ground states. It remains valid when perturbations like a uniform electric field are added to the external potential. Hence, both models will also predict the same linear response properties. A simple implementation of the ACKS2 and SQE+Q^0 model with a set of example input and output files is provided as supplementary material to offer additional numerical evidence that both models are equivalent for the computation of equilibrium charges and linear response properties.[152]

REFERENCES

[1]R. T. Sanderson, Science **114**, 670 (1951).

[2] W. Kohn, A. D. Becke, and R. G. Parr, J. Phys. Chem. **100**, 12974 (1996).

[3] W. Mortier, K. Van Genechten, and J. Gasteiger, J. Am. Chem. Soc. **107**, 829 (1985).

[4] W. J. Mortier, S. K. Ghosh, and S. Shankar, J. Am. Chem. Soc. **108**, 4315 (1986).

[5] K. A. Van Genechten, W. J. Mortier, and P. Geerlings, J. Chem. Phys. **86**, 5063 (1987).

[6] G. O. A. Janssens, B. G. Baekelandt, H. Toufar, W. J. Mortier, and R. A. Schoonheydt, J. Phys. Chem. **99**, 3251 (1995).

[7] R. Heidler, G. O. A. Janssens, W. J. Mortier, and R. A. Schoonheydt, Micropor. Mater. **12**, 1 (1997).

[8] S. L. Njo, J. Fan, and B. van de Graaf, J. Mol. Cat. A **134**, 79 (1998).

[9] A. C. T. van Duin, A. Strachan, S. Stewman, Q. Zhang, X. Xu, and W. A. Goddard, III, J. Phys. Chem. A **107**, 3803 (2003).

[10] K. S. Smirnov and D. Bougeard, Chem. Phys. **292**, 53 (2003).

[11] A. Hallil, R. Tétot, F. Berthier, I. Braems, and J. Creuze, Phys. Rev. B **73**, 165406 (2006).

[12] T. Verstraelen, S. V. Sukhomlinov, V. Van Speybroeck, M. Waroquier, and K. S. Smirnov, J. Phys. Chem. C **116**, 490 (2012).

[13] S. W. Rick, S. J. Stuart, and B. J. Berne, J. Chem. Phys. **101**, 6141 (1994).

[14] R. Chelli, S. Ciabatti, G. Cardini, R. Righini, and P. Procacci, J. Chem. Phys. **111**, 4218 (1999).

[15] A. C. T. van Duin, S. Dasgupta, F. Lorant, and W. A. Goddard, III, J. Phys. Chem. A **105**, 9396 (2001).

[16] R. Chelli and P. Procacci, J. Chem. Phys. **117**, 9175 (2002).

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

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

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

[20] S. Patel and C. L. Brooks, III, J. Comput. Chem. **25**, 1 (2004).

[21] S. Patel and C. L. Brooks, III, J. Chem. Phys. **122**, 024508 (2005).

[22] R. S. Vařeková, Z. Jiroušková, J. Vaněk, v. Suchomel, and J. Koča, Int. J. Mol. Sci. **8**, 572 (2007).

[23]I. Berente, E. Czinki, and G. Náray-szabó, J. Comput. Chem. **28**, 1936 (2007).

[24]J. E. Davis, G. L. Warren, and S. Patel, J. Phys. Chem. B **112**, 8298 (2008).

[25]B. A. Bauer and S. Patel, J. Mol. Liq. **142**, 32 (2008).

[26]Y. Zhong, G. L. Warren, and S. Patel, J. Comput. Chem. **29**, 1142 (2008).

[27]Z. Jiroušková, R. S. Varcařeková, J. Vaněk, and J. Koča, J. Comput. Chem. **30**, 1174 (2009).

[28]T. Verstraelen, V. Van Speybroeck, and M. Waroquier, J. Chem. Phys. **131**, 044127 (2009).

[29]Y. Zhong and S. Patel, J. Phys. Chem. B **114**, 11076 (2010).

[30]J. L. Banks, G. A. Kaminski, R. Zhou, D. T. Mainz, B. J. Berne, and R. A. Friesner, J. Chem. Phys. **110**, 741 (1999).

[31]S. Patel, A. D. Mackerell, Jr., and C. L. Brooks, III, J. Comput. Chem. **25**, 1504 (2004).

[32]Q. Yang and K. A. Sharp, J. Chem. Theory Comput. **2**, 1152 (2006).

[33]T. Verstraelen, E. Pauwels, F. De Proft, V. Van Speybroeck, P. Geerlings, and M. Waroquier, J. Chem. Theory Comput. **8**, 661 (2012).

[34]C. E. Wilmer and R. Q. Snurr, Chem. Eng. J. **171**, 775 (2011).

[35]J. S. Puranen, M. J. Vainio, and M. S. Johnson, J. Comput. Chem. **31**, 1722 (2010).

[36]Y.-P. Liu, K. Kim, B. J. Berne, R. A. Friesner, and S. W. Rick, J. Chem. Phys. **108**, 4739 (1998).

[37]T. T. Järvi, A. C. T. van Duin, K. Nordlund, and W. A. Goddard, III, J. Phys. Chem. A **115**, 10315 (2011).

[38]H. Manzano, S. Moeini, F. Marinelli, A. C. T. van Duin, F.-J. Ulm, and R. J.-M. Pellenq, J. Am. Chem. Soc. **134**, 2208 (2012).

[39]A. K. Rappe and W. A. Goddard, III, J. Phys. Chem. **95**, 3358 (1991).

[40]D. M. York and W. Yang, J. Chem. Phys. **104**, 159 (1996).

[41]B. T. Thole, Chem. Phys. **59**, 341 (1981).

[42]J. Chaves, J. M. Barroso, P. Bultinck, and R. Carbó-Dorca, J. Chem. Inf. Model. **46**, 1657 (2006).

[43]R. Chelli, R. Righini, S. Califano, and P. Procacci, J. Mol. Liq. **96-97**, 87 (2002).

[44]H. A. Stern, G. A. Kaminski, J. L. Banks, R. Zhou, B. J. Berne, and R. A. Friesner, J. Phys. Chem. B **103**, 4730 (1999).

[45]G. A. Kaminski, H. A. Stern, B. J. Berne, R. A. Friesner, Y. X. Cao, R. B. Murphy,

R. Zhou, and T. A. Halgren, J. Comput. Chem. **23**, 1515 (2002).

[46]G. A. Kaminski, H. A. Stern, B. J. Berne, and R. A. Friesner, J. Phys. Chem. A **108**, 621 (2004).

[47]R. Chelli, M. Pagliai, P. Procacci, G. Cardini, and V. Schettino, J. Chem. Phys. **122**, 074504 (2005).

[48]C. M. MacDermaid and G. A. Kaminski, J. Phys. Chem. B **111**, 9036 (2007).

[49]I. Lazić and B. J. Thijsse, Comput. Mater. Sci. **53**, 483 (2012).

[50]G. Tabacchi, C. J. Mundy, J. Hutter, and M. Parrinello, J. Chem. Phys. **117**, 1416 (2002).

[51]R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).

[52]J. Kolafa, J. Comput. Chem. **25**, 335 (2004).

[53]R. Chelli, V. Schettino, and P. Procacci, J. Chem. Phys. **122**, 234107 (2005).

[54]B. G. Baekelandt, W. J. Mortier, J. L. Lievens, and R. A. Schoonheydt, J. Am. Chem. Soc. **113**, 6730 (1991).

[55]P. Bultinck, W. Langenaeker, R. Carbó-Dorca, and J. P. Tollenaere, J. Chem. Inf. Comput. Sci. **43**, 422 (2003).

[56]R. Chelli, P. Procacci, R. Righini, and S. Califano, J. Chem. Phys. **111**, 8569 (1999).

[57]G. L. Warren, J. E. Davis, and S. Patel, J. Chem. Phys. **128**, 144110 (2008).

[58]J. Cioslowski and B. B. Stefanov, J. Chem. Phys. **99**, 5151 (1993).

[59]J. Chen and T. J. Martínez, Chem. Phys. Lett. **438**, 315 (2007).

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

[61]W. Yang, Y. Zhang, and P. W. Ayers, Phys. Rev. Lett. **84**, 5172 (2000).

[62]P. W. Ayers, J. Math. Chem. **43**, 285 (2008).

[63]J. Morales and T. J. Martínez, J. Phys. Chem. A **105**, 2842 (2001).

[64]J. Morales and T. J. Martínez, J. Phys. Chem. A **108**, 3076 (2004).

[65]S. M. Valone and S. R. Atlas, J. Chem. Phys. **120**, 7262 (2004).

[66]R. A. Nistor, J. G. Polihronov, M. H. Müser, and N. J. Mosey, J. Chem. Phys. **125**, 094108 (2006).

[67]T. A. Halgren, J. Comput. Chem. **17**, 520 (1996).

[68]B. L. Bush, C. I. Bayly, and T. A. Halgren, J. Comput. Chem. **20**, 1495 (1999).

[69]R. A. Nistor and M. H. Müser, Phys. Rev. B **79**, 104303 (2009).

[70] M. H. Müser, Eur. Phys. J. B **85**, 135 (2012).

[71] H. S. Smalø, P.-O. Åstrand, and L. Jensen, J. Chem. Phys. **131**, 044101 (2009).

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

[73] C. F. Matta and R. F. W. Bader, J. Phys. Chem. A **110**, 6365 (2006).

[74] J. Chen, D. Hundertmark, and T. J. Martínez, J. Chem. Phys. **129**, 214113 (2008).

[75] D. Mathieu, J. Chem Phys. **127**, 224103 (2007).

[76] P. T. Mikulski, M. T. Knippenberg, and J. A. Harrison, J. Chem. Phys. **131**, 241105 (2009).

[77] T. Verstraelen, P. Bultinck, V. Van Speybroeck, P. W. Ayers, D. Van Neck, and M. Waroquier, J. Chem. Theory Comput. **7**, 1750 (2011).

[78] F. De Proft, W. Langenaeker, and P. Geerlings, J. Mol. Struct. **339**, 45 (1995).

[79] P. Itskowitz and M. L. Berkowitz, J. Phys. Chem. A **101**, 5687 (1997).

[80] R. Winkler and S. T. Pantelides, J. Chem. Phys. **106**, 7714 (1997).

[81] G. Tabacchi, J. Hutter, and C. J. Mundy, J. Chem. Phys. **123**, 074108 (2005).

[82] T. J. Giese and D. M. York, Theor. Chem. Acc. **131**, 1145 (2012).

[83] P. H. Dederichs, S. Blügel, R. Zeller, and H. Akai, Phys. Rev. Lett. **53**, 2512 (1984).

[84] B. Kaduk, T. Kowalczyk, and T. Van Voorhis, Chem. Rev. **112**, 321 (2012).

[85] Q. Wu and T. Van Voorhis, J. Chem. Theory Comput. **2**, 765 (2006).

[86] Q. Wu, P. W. Ayers, and Y. Zhang, J. Chem. Phys. **131**, 164112 (2009).

[87] F. L. Hirshfeld, Theor. Chem. Acc. **44**, 129 (1977).

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

[89] P. Bultinck, C. Van Alsenoy, P. W. Ayers, and R. Carbó-Dorca, J. Chem. Phys. **126**, 144111 (2007).

[90] T. C. Lillestolen and R. J. Wheatley, Chem. Commun. **45**, 5909 (2008).

[91] R. F. W. Bader, *Atoms in Molecules: A Quantum Theory* (Oxford University Press, USA, 1994).

[92] P. Itskowitz and M. L. Berkowitz, J. Phys. Chem. A **102**, 4808 (1998).

[93] P. Itskowitz and M. L. Berkowitz, J. Chem. Phys. **109**, 10142 (1998).

[94] A. Aguado, L. Bernasconi, and P. A. Madden, Chem. Phys. Lett. **356**, 437 (2002).

[95] M. Levy, Proc. Natl. Acad. Sci. USA **76**, 6062 (1979).

[96] R. K. P. Zia, E. F. Redish, and S. R. McKay, Am. J. Phys. **77**, 614 (2009).

[97] E. H. Lieb, Int. J. Quantum Chem. **24**, 243 (1983).

[98] W. Kutzelnigg, J. Mol. Struct. Theochem **768**, 163 (2006).

[99] J. Cioslowski and M. Martinov, J. Chem. Phys. **101**, 366 (1994).

[100] J. Cioslowski and M. Martinov, J. Phys. Chem. **100**, 6156 (1996).

[101] W. Parr, Robert G. ann Yanf, in *Density-Functional Theory of Atoms and Molecules* (Oxford University Press, 200 Madison Avenue, New York, NY 10016, 1989) pp. 246–254.

[102] R. G. Parr and R. G. Pearson, J. Am. Chem. Soc. **105**, 7512 (1983).

[103] K. Fukui, T. Yonezawa, and H. Shingu, J. Chem. Phys. **20**, 722 (1952).

[104] R. G. Parr and W. Yang, J. Am. Chem. Soc. **106**, 4049 (1984).

[105] W. Yang and W. J. Mortier, J. Am. Chem. Soc. **108**, 5708 (1986).

[106] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).

[107] R. G. Gordon and Y. S. Kim, J. Chem. Phys. **56**, 3122 (1972).

[108] Y. S. Kim and R. G. Gordon, J. Chem. Phys. **60**, 1842 (1974).

[109] E. H. Lieb and B. Simon, Phys. Rev. Lett. **31**, 681 (1973).

[110] B. Rotenberg, M. Salanne, C. Simon, and R. Vuilleumier, Phys. Rev. Lett. **104**, 138301 (2010).

[111] S. Laricchia, E. Fabiano, L. A. Constantin, and F. Della Sala, J. Chem. Theory Comput. **7**, 2439 (2011).

[112] P. W. Ayers, J. Chem. Phys. **113**, 10886 (2000).

[113] J. Cioslowski and S. T. Mixon, J. Am. Chem. Soc. **115**, 1084 (1993).

[114] H. Toufar, K. Nulens, G. O. A. Janssens, W. J. Mortier, R. A. Schoonheydt, F. De Proft, and P. Geerlings, J. Phys. Chem. **100**, 15383 (1996).

[115] P. Bultinck, D. L. Cooper, and D. Van Neck, Phys. Chem. Chem. Phys. **11**, 3424 (2009).

[116] T. Verstraelen, P. W. Ayers, V. Van Speybroeck, and M. Waroquier, Chem. Phys. Lett. **545**, 138 (2012).

[117] S. Van Damme, P. Bultinck, and S. Fias, J. Chem. Theory Comput. **5**, 334 (2009).

[118] S. Catak, M. D'hooghe, T. Verstraelen, K. Hemelsoet, A. Van Nieuwenhove, H.-J. Ha, M. Waroquier, N. De Kimpe, and V. Van Speybroeck, J. Org. Chem. **75**, 4530 (2010).

[119] L. Vanduyfhuys, T. Verstraelen, V. Van Speybroeck, and M. Waroquier, J. Chem. Theory Comput. **8**, 3217 (2012).

[120] T. J. Giese and D. M. York, J. Chem. Phys. **120**, 9903 (2004).

[121] A. Savin, F. Colonna, and M. Allavena, J. Chem. Phys. **115**, 6827 (2001).

[122] H. Nakano, T. Yamamoto, and S. Kato, J. Chem. Phys. **132**, 044106 (2010).

[123]N. Sablon, F. De Proft, P. W. Ayers, and P. Geerlings, J. Chem. Theory Comput. **6**, 3671 (2010).

[124]W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[125]J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[126]A. D. Becke, Phys. Rev. A **38**, 3098 (1988).

[127]C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).

[128]F. Colonna and A. Savin, J. Chem. Phys. **110**, 2828 (1999).

[129]A. M. Teale, S. Coriani, and T. Helgaker, J. Chem. Phys. **130**, 104111 (2009).

[130]A. M. Teale, S. Coriani, and T. Helgaker, J. Chem. Phys. **132**, 164115 (2010).

[131]Q. Wu and W. Yang, J. Chem. Phys. **118**, 2498 (2003).

[132]E. K. U. Gross and W. Kohn, Phys. Rev. Lett. **55**, 2850 (1985).

[133]P. W. Ayers, Theor. Chem. Acc. **106**, 271 (2001).

[134]K. Wiberg, Tetrahedron **24**, 1083 (1968).

[135]I. Mayer, J. Comput. Chem. **28**, 204 (2007).

[136]J. G. Ángyán, J. Mol. Struct. Theochem **501-502**, 379 (2000).

[137]N. Sablon, F. De Proft, and P. Geerlings, J. Phys. Chem. Lett. **1**, 1228 (2010).

[138]N. Sablon, F. De Proft, and P. Geerlings, Chem. Phys. Lett. **498**, 192 (2010).

[139]E. Prodan and W. Kohn, Proc. Natl. Acad. Sci. **102**, 11635 (2005).

[140]R. Resta and D. Vanderbilt, *Physics of Ferroelectrics: A Modern Perspective* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2007) pp. 31–68.

[141]S. M. Valone and S. R. Atlas, Phys. Rev. Lett. **97**, 256402 (2006).

[142]S. M. Valone, J. Li, and S. Jindal, Int. J. Quantum Chem. **108**, 1452 (2008).

[143]S. M. Valone, J. Chem. Theory Comput. **7**, 2253 (2011).

[144]R. F. Nalewajski, Int. J. Quantum Chem. **69**, 591 (1998).

[145]P. W. Ayers, Theor. Chem. Acc. **118**, 371 (2007).

[146]J. N. Murrell, M. Randic, and D. R. Williams, Proc. R. Soc. London **284**, 566 (1965).

[147]N. Gresh, P. Claverie, and A. Pullman, Int. J. Quantum Chem. **29**, 101 (1986).

[148]N. Gresh, G. A. Cisneros, T. A. Darden, and J.-P. Piquemal, J. Chem. Theory Comput. **3**, 1960 (2007).

[149]See supplementary material at [URL will be inserted by AIP] for a reformulation of the EEM and ACKS2 equations in terms of atomic charges.

[150]M. Benzi, G. H. Golub, and J. Liesen, Acta Numer. **14**, 1 (2005).

[151]W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing, Second Edition*, 2nd ed. (Cambridge University Press, 1992) pp. 59–71.

[152]See supplementary material at [URL will be inserted by AIP] for a simple implementation of the SQE and ACKS2 models.

TABLES AND FIGURES

Parameter	EEM (Parr & Pearson)	ACKS2 (fitted)
$\mu_\mathrm{H} - \mu_\mathrm{F}$ [eV/e]	5.04	3.24
$\eta_\mathrm{H} + \eta_\mathrm{F}$ [eV/e^2]	25.60	26.86
$1/X_0$ [eV/e^2]	–	0.0672
τ [Å]	–	0.3280

TABLE I. EEM and ACKS2 parameters for the computation of several properties during the dissociation of hydrogen fluoride. (See Fig. 3.) The EEM parameters are based on experimental atomic properties.[102] The ACKS2 parameters were fitted to CASSCF reference data. (See text for more details.)

Space of n-representable densities
with N_{tot} electrons

Subspace of n-representable densities
with populations $(N_1, \ldots, N_M)$.

$\rho_{\text{pop}}(N_1, \ldots, N_M; \mathbf{r})$
with lowest energy in this subspace

$(N_1, \ldots, N_M)$

$(N_1', \ldots, N_M')$ $(N_1'', \ldots, N_M'')$
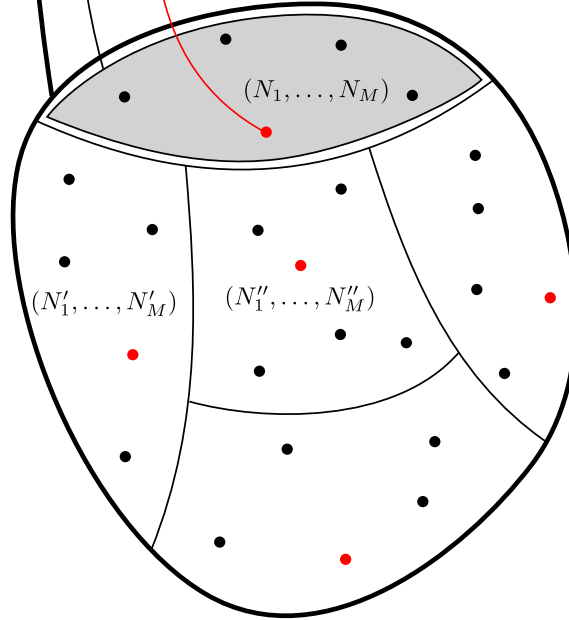
$$N_{\text{tot}} = \sum_A N_A = \sum_A N_A' = \sum_A N_A'' = \ldots$$

FIG. 1. Schematic presentation of population-constrained DFT. Each subspace corresponds to all n-representable densities with a pre-defined set of populations. Within each set, $\rho_{\text{pop}}(N_1, \ldots, N_M)$ (red dot) minimizes the DFT energy over all densities that have populations $(N_1, \ldots, N_M)$.
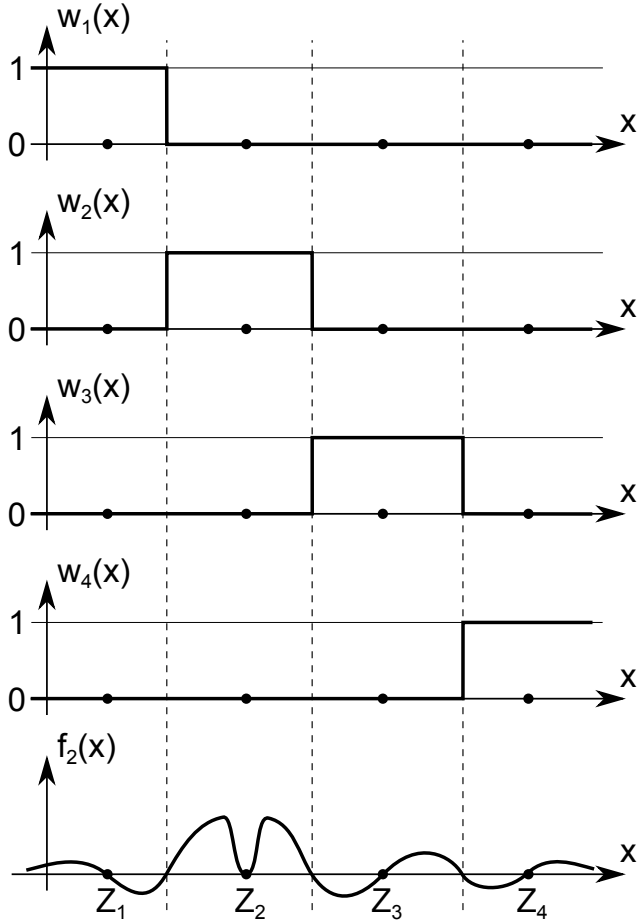
50

FIG. 2. Schematic representation of the AIM Fukui function for the second atom, $f_2(x)$, in a linear molecule consisting of four atoms. For the sake of a clear visualization, the figure uses binary weight functions, $w_A(x)$, like the ones used in the QTAIM scheme.[91] The properties of the AIM Fukui function discussed in the text are also applicable to more general weight functions. In atoms 1, 3 and 4, the AIM Fukui function contains positive and negative contributions that cancel each other out. The AIM Fukui function in atom 2 integrates to unity and is small in the region of the core electrons, if those would be present.
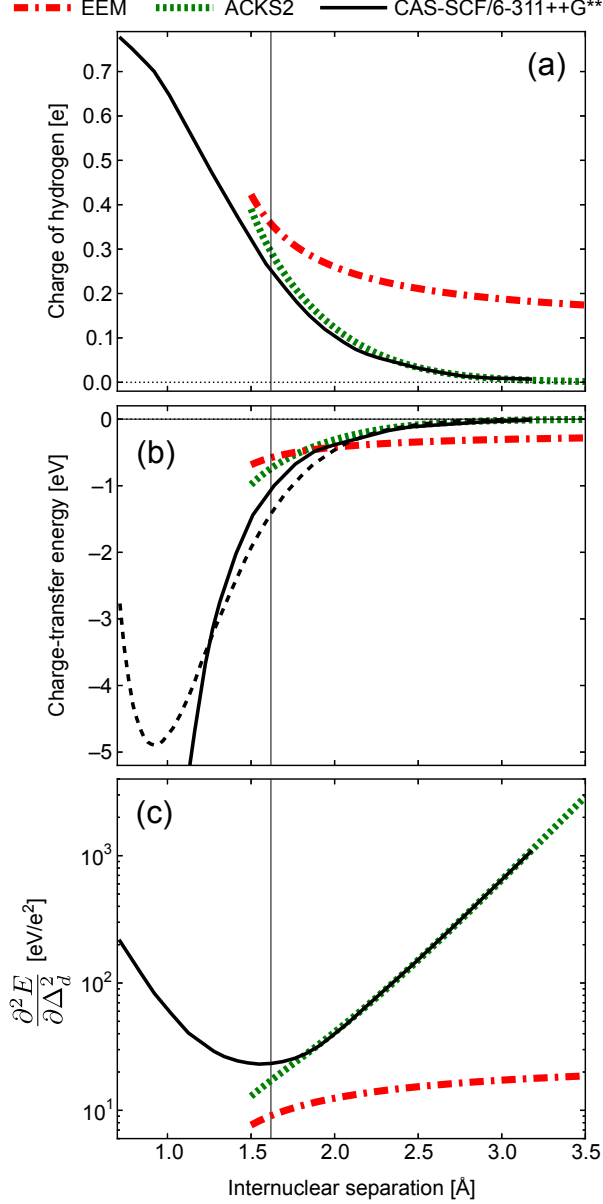
FIG. 3. Numerical results for the dissocation of hydrogen fluoride. The following properties are plotted as function of the internuclear separation: (a) the charge on the hydrogen atom, (b) the energy due to charge transfer and (c) the hardness of the system with respect to charge transfer. The reference data (solid black lines) are CASSCF/6-311++G** results taken from Ref. 58. In part (b), the total CASSCF interaction energy of the diatomic systems is also included (black dashed line). The reference data are compared with results obtained with EEM (red dash-dot line) and ACKS2 (green dotted line). The Coulson-Fischer point is indicated by a think black vertical line. The ACKS2 parameters were fitted to reproduce the CASSCF data, leading to reasonable parameters shown in Table I. No satisfactory parameters were found for the EEM model. For this figure, the experimental isolated-atom properties from Ref. 102 were used as EEM parameters.