# Direct Computation of Parameters for Accurate Polarizable Force Fields

Toon Verstraelen,[1, a)] Steven Vandenbrande,[1] and Paul W. Ayers[2]

[1)] *Center for Molecular Modeling (CMM), Ghent University, Technologiepark 903, B9000 Ghent, Belgium, (Member of the QCMM Ghent–Brussels Alliance)*

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

(Dated: 1 November 2014)

We present an improved electronic linear response model to incorporate polarization and charge-transfer effects in polarizable force fields. This model is a generalization of the Atom-Condensed Kohn-Sham DFT, approximated to second order (ACKS2): it can now be defined with any underlying variational theory (next to KS-DFT) and it can include atomic multipoles and off-center basis functions. Parameters in this model are computed efficiently as expectation values of an electronic wavefunction, obviating the need for their calibration, regularization and manual tuning. In the limit of a complete density and potential basis set in the ACKS2 model, the linear response properties of the underlying theory for a given molecular geometry are reproduced exactly. A numerical validation with a test set of 110 molecules shows that very accurate models can already be obtained with fluctuating charges and dipoles. These features greatly facilitate the development of polarizable force fields.

Keywords: polarizable force fields, Kohn-Sham kinetic energy, linear response

[a)] Electronic mail: toon.verstraelen@ugent.be

## I. INTRODUCTION

A wide variety of molecular systems is studied with force-field (FF) simulations because these methods enable an efficient sampling of the molecular configurations. Force-field applications usually do not involve chemical reactions but rather focus on supramolecular processes, which are dominated by non-covalent interactions. Hence, an accurate model for the non-covalent energy is indispensable for reliable predictions. Whereas many force-field models are limited to additive two-, three or four-body terms in the energy expression, it is well established that some types of interactions have a strong many-body character. Metallic bonding is a classical example of a many-body interaction, for which specific force fields were developed, e.g. the Embedded-Atom Method.[1] Also non-covalent interactions contain considerable many-body contributions, which are largely due to electronic polarization.[2–7] The mutual polarization of two molecular fragments results in an attractive force that is hard to capture with few-body terms.[8–11] To improve the accuracy of non-covalent interaction energies, in the 1980s scientists started incorporating explicit treatments of electronic polarization into force fields, creating so-called polarizable force fields (PFFs).[2,12–16] Beginning in the mid-1990s, PFFs became a very active field of research, especially for biomolecular simulations.[17]

Many authors have assessed the importance and effectiveness of PFFs in their simulations. Often the simulation setup involves an inhomogeneous host-guest or solvent-solute system. One of the earliest successes of PFFs was the proper description of solvated ions.[3,18,19] For biomolecular simulations, the improved accuracy of PFFs is now well-established. For example, the error in simulated pKa shifts of carboxyl groups in a serine protease inhibitor decreases drastically (from 3.28 to 0.58 units) when switching from a regular FF to a PFF.[20] In a different benchmark, it was shown that PFFs improve the correspondence of theoretical protein-ligand structures with XRD reference data.[21] Another recent study demonstrates the improved structures and thermal equilibria of short solvated DNA helices obtained with PFF simulations.[22] The properties of guest molecules in porous materials are also actively investigated with PFFs.[23,24] For example, Cirera *et al.* studied the adsorption of water in MIL-53(Cr) and found that the hydrogen bonding network was significantly affected by the explicit treatment of polarization.[25] An impressive non-standard use of PFFs is the modeling of redox reactions in atomistic battery simulations.[26,27] Even for gas phase dimers

or oligomers, electronic polarization contributes significantly to the interaction energy,[9,28,29] and such reference data is therefore often used to validate PFFs.[18,21,30–32]

For homogeneous systems, PFFs improve the correspondence between simulated and experimental data. For example, PFF simulations of organic or ionic liquids showed an improved correspondence of thermodynamic and structural properties with experiment, while also reproducing gas-phase dimer properties.[24,30,31,33–40] Also in the field of materials science, PFFs are gaining attention, e.g. a polarizable Al-O potential was recently proposed that can reproduce crystal energies, lattice constants and deformation energies of different $Al_xO_y$ crystal phases.[41] Another nice example is a LiI PFF that was fitted to ab initio reference data and that reproduces properties of both the liquid and crystal phases.[42]

It is clear from the successes described above that PFFs have the potential to be significantly more accurate than non-polarizable FFs. Yet, in some cases, PFFs perform worse than more mature non-polarizable FFs, for a few of the thermodynamic and structural properties that are used to assess the FF.[38,43–46] Although the inclusion of electronic polarization improves the physics of the model, it becomes clearly more challenging to find the optimal values of all parameters.

This paper will exclusively focus on the development of the polarization part of the PFF, i.e. the part that approximates the electronic linear response of the system. For the sake of clarity, we will refer to these energy terms as the electronic response force field (ERFF). ERFFs form a research field on their own with specific challenges and methodologies. Below, two major difficulties in the development of accurate ERFFs are reviewed: (i) the non-trivial assignment of adjustable parameters and (ii) the systematic errors encountered in conventional ERFFs with variable charges. Our long-term objective is to eliminate both weaknesses by proposing improved analytic forms and by devising methodological tools that facilitate the parameterization process. In this paper, we present for the first time an ERFF that can correctly describe charge-transfer polarization in dielectric systems with all ERFF parameters directly computed as expectation values of an electronic wavefunction. In the limit of a complete ERFF basis set, the model exactly reproduces the DFT reference data. The ERFF proposed in this work is free from adjustable parameters and hence no iterative refinement or human intervention is needed to optimize them. Instead, the model defines how the parameters can be computed deterministically.

The iterative calibration of adjustable ERFF parameters, e.g. atomic dipole polarizabil-

ities, is well-known and becomes labor-intensive and ill-posed for complex systems with a large number of parameters. Such least-squares fits are usually carried out with ab initio data computed for a set of reference molecules of interest, e.g. the dipole polarizability tensor and/or the response of the electron density or electrostatic potential (ESP) to a probe such as a point charge.[21,24,30,31,39,47–53] An alternative is the dipole-matching technique,[19,54] which is the ERFF analogon of force-matching.[55] The main disadvantage of these iterative calibration techniques is that regularization must be used to prevent parameters from converging to nonphysical values during the calibration,[21,30,47,56] which necessitates a non-trivial human judgment to fix certain degrees of freedom in the parameters. When a large number of systems must be parameterized, this becomes intractable and, depending on the type of regularization, it may also undermine the reproducibility, the physical interpretation, or the optimality of the parameters.[57] To overcome such issues, several authors use physical arguments to fix some parameters *a priori* and to calibrate only the remainder of the ERFF parameters.[51,56] There are also a few attempts to eliminate these issues entirely with a direct *approximation* of all ERFF parameters as expectation values of density response basis functions, e.g. using a Jellium model[42,58] or by partitioning the non-interacting response in Hartree-Fock theory[59,60] (used in the NEMO method[61,62]). Besides the relatively direct calibration of ERFF parameters with ab initio linear response data, it is also common to fit all parameters in the whole PFF to more diverse experimental and/or ab initio reference data.[22,24,37,44,63–65]

To avoid the empirical calibration of ERFF parameters, we will compute them *exactly* from the electronic wavefunction obtained with an *underlying* variational electronic structure theory. In this approach, the ERFF is an electronic linear-response model in which all parameters are well-defined in terms of the underlying theory. In the limit of a complete basis, the ERFF exactly reproduces the response properties of the *underlying* theory. This choice seriously affects the perspective on force-field development. Usually, one considers the force-field model to be a set of equations in which adjustable parameters must be fixed *a posteriori*. In this work, the ERFF model defines both the equations and the values of the parameters for any given molecular geometry. Several authors have already followed this approach to construct an ERFF in the context of Density Functional Theory (DFT) but they still resorted to an approximate model for, or to an empirical calibration of, the parameters.[14,42,58,66–69]

While the direct computation of ERFF parameters is clearly an attractive feature, some challenges still remain. For example, there is no computational gain compared to ab initio molecular dynamics when the ERFF parameters must be recomputed from an ab initio theory at every time step. For a practically feasible force field, some additional effort is required to approximate the geometry dependence of the parameters with simple functional forms.[36,42,70] In this work, we mainly focus on a new ERFF model and the validity of the parameters derived from an underlying theory. Simple functional forms to approximate the geometry dependence of such parameters are the topic of ongoing research.

A second limitation of *conventional* ERFFs is that the inclusion of variable atomic charges leads to systematic errors, particularly for dielectric systems. By *conventional*, we mean that no special constraints or restraints are imposed on the charges and only pairwise (damped or screened) classical electrostatic interactions between the charges are in effect, e.g. as in the Electronegativity Equalization Method (EEM),[14,71] Charge Equilibration (QE, Qeq[15] or CHEQ[36]), Fluctuation Charges (FlucQ),[16] and the Chemical Potential Equalization (CPE) approaches.[66] A first systematic error is the cubic scaling of the dipole polarizability with system size, while for dielectric systems one expects a linear scaling. In other words, for extended systems, one always obtains a metallic behavior.[72–76] EEM also fails to reproduce other response properties of organic molecules, such as the dipole derivatives.[77] A second systematic error is the non-integer charge on dissociated fragments,[78,79] unlike the behavior one expects in the dissociation limit.[80] The most pragmatic solution is to relinquish variable charges altogether but this also excludes the possibility of describing non-local polarization (charge-transfer) effects[4,29,79,81–83] in an ERFF. When a molecular system is composed of a well-defined collection of molecules, e.g. a molecular liquid, one may impose constraints on the total charge of each molecule.[35,74,84] However, this is merely an ad hoc solution that is not applicable to many cases of interest, e.g. reactive force field simulations[85,86] or computations on extended and periodic systems.

To obtain qualitatively correct ERFFs with fluctuating charges for dielectric systems, it is well-known that one must somehow limit the distance over which charge can be transferred between atoms.[72–76] We recently showed that such restraints on charge-transfer can be justified with a quantum mechnical model: they effectively account for the non-local effects of the electronic kinetic energy.[79] We condensed the Kohn-Sham DFT equations[87] to atomic degrees of freedom, i.e. relative populations and relative Kohn-Sham potentials,

and approximated the energy to second order in these variables. This lead to the Atom-Condensed Kohn-Sham DFT model approximated to Second order (ACKS2),[79] which is a straightforward extension of the Electronegativity Equalization Method (EEM)[14] with an electronic kinetic energy term. We also showed that the ACKS2 model is isomorphic to the Split-Charge Equilibration (SQE).[73,75] Some authors suggested correcting for kinetic, exchange and correlation effects with local (short-ranged) Hückel-type corrections to the electrostatic interactions.[42,66,69] However, such corrections cannot alleviate the systematic errors found in ERFFs with variable charges, i.e. the metallic behavior of any large system and the erratic dissociation limits.[72–76] In this paper we will derive a generalization of the ACKS2 model for any underlying variational level of theory. We will also introduce more general density and potential basis functions as to allow for atomic multipoles[33,51,58,88] and variable off-center charges.[64,89,90] This second generalization introduces an additional matrix of parameters (the overlap between density and potential basis functions), which equaled the unity matrix in our previous work.[79]

Another new aspect of this paper is the numerical validation of the generalized ACKS2 model using a test set of 110 molecules with chemical formula $C_4O_2H_8$. In this test we maintain a backward compatibility with the original ACKS2 paper (and thus also the SQE model) in the sense that density and potential basis sets are kept bi-orthogonal. The accuracy of the ACKS2 model is tested by checking the dipole polarizability tensors against reference data obtained with the coupled-perturbed Kohn-Sham (CPKS) method.[91,92]

Unlike some previous work on PFFs, mostly related to EEM,[14,50,52,57,71,76,93–96] we do not consider first-order parameters in this work. Instead, the electronic ground state is used as the reference point for the second order expansion. Hence, the current approach does not predict the static atomic multipoles and only focuses on the multipoles induced by a perturbation of the external field. Although the direct computation of first-order parameters is also a topic of interest, it goes beyond the scope of this work.

The paper is structured as follows. In section II, we present a simplified and more general derivation of the ACKS2 model, including straightforward expressions to evaluate the parameters for a given Kohn-Sham wavefunction. Section III presents the numerical validation of the generalized ACKS2 model. Finally, section IV summarizes the main conclusions and provides an outlook on future work.

## II. THEORY

In this section, we develop a generalization of the ACKS2 model.[79] The derivation starts from DFT with a separation of the universal functional into an explicit and an implicit part,[97] which allows one to define an ACKS2-type ERFF based on any variational electronic structure theory. A linear response theory is developed in this formalism, followed by a condensation of the equations on finite basis sets. Finally, practical expressions for the ERFF parameters are given when Kohn-Sham DFT with a semi-local exchange-correlation functional is used as the underlying theory.

### A. Density Functional Theory with an Auxiliary Wavefunction

A fundamental insight from DFT is that the electronic energy of a molecular system can be written as the sum of a universal functional of the electron density and the interaction of the electrons with an external potential.[98,99] We consider models for the universal functional that consist of two parts: an explicit functional, $E^{\text{exp}}$, and an implicit functional, $E^{\text{imp}}$.

$$E_v[\rho] = E^{\text{exp}}[\rho] + E^{\text{imp}}[\rho] + \int \rho(\mathbf{r}) v(\mathbf{r}) d\mathbf{r} \tag{1}$$

The implicit functional uses an auxiliary $N$-fermion wavefunction through a constrained-search formulation:[99]

$$E^{\text{imp}}[\rho] = \min_{\Psi \to \rho} W[\Psi] \tag{2}$$

Implicit functionals are commonly used in Kohn-Sham DFT to obtain better approximations for the exchange-correlation energy such as the popular hybrid functionals that contain a fraction of Hartree-Fock exchange.[97,100,101] Here, we use the concept of an implicit functional in a more general sense, i.e. it is used to partition the *entire* universal functional into an explicit and implicit part.

For the derivation below, it is not essential to specify which contributions to the energy are assigned to the implicit or the explicit functional. For any reasonable ERFF, however, the Hartree energy will be included in the explicit term. If applicable, some additional density dependent terms may be included in the explicit functional as well, e.g. a GGA approximation[102–104] of the exchange-correlation energy. All the remaining terms can be included in the implicit functional, e.g. the (Kohn-Sham) kinetic energy or (fractional)

Hartree-Fock exchange. This formulation covers many approximations of the universal functional such as Kohn-Sham DFT with GGA,[102–104] MGGA,[105] or hybrid functionals.[100,101] Furthermore, one may also cast other electronic structure theories in the same form, such as Configuration Interaction methods[106] and novel theories based on the Density Matrix Renormalization Group (DMRG)[107–109] or a geminal wavefunction.[110–112]

The constrained search can be written out explicitly with the method of Lagrange multipliers:[99,113–121]

$$E^{\mathrm{imp}}[\rho] = \sup_{u} \left( E^{\circ}[u, N] - \int \rho(\mathbf{r}) u(\mathbf{r}) d\mathbf{r} \right) \tag{3}$$

$$E^{\circ}[u, N] = \min_{\Psi} \left( W[\Psi] + \int \rho[\Psi](\mathbf{r})\, u(\mathbf{r}) d\mathbf{r} \right) \tag{4}$$

where $u(\mathbf{r})$ is a function that specifies a Lagrange multiplier at every point in space and $\rho[\Psi](\mathbf{r})$ is the electron density of the auxiliary wavefunction. The expression is divided over two equations to emphasize that the implicit functional is conveniently written as a Legendre transform of a functional, $E^{\circ}[u, N]$, where $u(\mathbf{r})$ acts as an auxiliary potential and $N$ is the number of particles. Note that all terms that depend on the auxiliary wavefunction are collected in $E^{\circ}[u, N]$. This energy can be interpreted as the ground state energy of $W[\Psi]$ in a fixed auxiliary potential, $u(\mathbf{r})$.

Using the notation introduced above, the $N$-electron ground state for a given external potential, $v(\mathbf{r})$, is found by minimizing the following Lagrangian w.r.t. $\rho(\mathbf{r})$ and maximizing it w.r.t. $u(\mathbf{r})$ and $\mu$:

$$L_v[\rho, u, \mu] = E_v[\rho, u] - \mu \left( \int \rho(\mathbf{r}) d\mathbf{r} - N \right), \tag{5}$$

where the energy is now a functional of the density and the auxiliary potential:

$$E_v[\rho, u] = E^{\mathrm{exp}}[\rho] + E^{\circ}[u, N] + \int \rho(\mathbf{r}) \big( v(\mathbf{r}) - u(\mathbf{r}) \big) d\mathbf{r}. \tag{6}$$

The stationary point is defined by the following sets of equations, which can be seen as the analogue of the conventional DFT Euler-Lagrange equations when the universal functional is separated into an explicit and implicit term:

$$\frac{\delta E^{\mathrm{exp}}[\rho]}{\delta \rho(\mathbf{r})} + v(\mathbf{r}) - u(\mathbf{r}) = \mu \tag{7a}$$

$$\frac{\delta E^{\circ}[u, N]}{\delta u(\mathbf{r})} - \rho(\mathbf{r}) = 0 \tag{7b}$$

$$\int \rho(\mathbf{r}) d\mathbf{r} = N \tag{7c}$$

Note that these equations may not fully fix the auxiliary potential, $u(\mathbf{r})$, e.g. one may always add an arbitrary constant. Still, all solutions for $u(\mathbf{r})$ will lead to the same electron density and electronic energy.

### B. Linear Response

In analogy with the CPE method by York and Yang,[66] a reference is chosen for the external potential, $v_0(\mathbf{r})$. This choice also fixes the reference value of the density, $\rho_0(\mathbf{r})$, the auxiliary potential, $u_0(\mathbf{r})$, and the equalized chemical potential, $\mu_0$, through the Euler-Lagrange Eqs. (7). A polarizable force field is then developed by considering deviations from this reference state:

$$v(\mathbf{r}) = v_0(\mathbf{r}) + \Delta v(\mathbf{r}) \qquad \rho(\mathbf{r}) = \rho_0(\mathbf{r}) + \Delta\rho(\mathbf{r})$$
$$u(\mathbf{r}) = u_0(\mathbf{r}) + \Delta u(\mathbf{r}) \qquad \mu = \mu_0 + \Delta\mu$$

In ERFF terminology, $\Delta v(\mathbf{r})$ is a perturbation in the external potential and $\Delta\rho(\mathbf{r})$, $\Delta u(\mathbf{r})$ and $\Delta\mu$ are induced quantities due to this perturbation. For a given perturbation, the induced quantities are found by solving similar Euler-Lagrange equations:

$$\frac{\delta E^{\mathrm{exp}}[\rho_0 + \Delta\rho]}{\delta\rho(\mathbf{r})} - \frac{\delta E^{\mathrm{exp}}[\rho_0]}{\delta\rho(\mathbf{r})} + \Delta v(\mathbf{r}) - \Delta u(\mathbf{r}) = \Delta\mu \tag{8a}$$
$$\frac{\delta E^{\circ}[u_0 + \Delta u, N]}{\delta u(\mathbf{r})} - \frac{\delta E^{\circ}[u_0, N]}{\delta u(\mathbf{r})} - \Delta\rho(\mathbf{r}) = 0 \tag{8b}$$
$$\int \Delta\rho(\mathbf{r}) d\mathbf{r} = 0 \tag{8c}$$

In the limit of a small perturbation, the first two equations can be linearized:

$$\int \frac{\delta^2 E^{\mathrm{exp}}[\rho_0]}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')} \Delta\rho(\mathbf{r}') d\mathbf{r}' + \Delta v(\mathbf{r}) - \Delta u(\mathbf{r}) \approx \Delta\mu \tag{9a}$$
$$\int \frac{\delta^2 E^{\circ}[u_0, N]}{\delta u(\mathbf{r})\delta u(\mathbf{r}')} \Delta u(\mathbf{r}') d\mathbf{r}' - \Delta\rho(\mathbf{r}) \approx 0 \tag{9b}$$

The second order functional derivative in Eq. (9a) is the hardness kernel of the explicit functional, which will be denoted as $\eta^{\mathrm{exp}}(\mathbf{r}, \mathbf{r}')$. The second order functional derivative in Eq. (9b) is the response kernel of the implicit functional, which will be denoted as $\chi^{\mathrm{imp}}(\mathbf{r}, \mathbf{r}')$.

One approximates the change in electronic energy, see Eq. (6), to second order as follows:

$$\begin{aligned} &E_{v_0+\Delta v}[\rho_0+\Delta\rho, u_0+\Delta u] - E_{v_0}[\rho_0, u_0] \approx \\ &\frac{1}{2}\iint \eta^{\mathrm{exp}}(\mathbf{r},\mathbf{r}')\Delta\rho(\mathbf{r})\Delta\rho(\mathbf{r}')d\mathbf{r}d\mathbf{r}' + \frac{1}{2}\iint \chi^{\mathrm{imp}}(\mathbf{r},\mathbf{r}')\Delta u(\mathbf{r})\Delta u(\mathbf{r}')d\mathbf{r}d\mathbf{r}' \\ &- \int \Delta\rho(\mathbf{r})\Delta u(\mathbf{r})d\mathbf{r} + \int \Delta\rho(\mathbf{r})\Delta v(\mathbf{r})d\mathbf{r} + \int \rho_0(\mathbf{r})\Delta v(\mathbf{r})d\mathbf{r} \end{aligned} \tag{10}$$

The first three terms constitute the (positive) energy needed to polarize the molecule. The last two terms are the interaction of the perturbation with the polarized and permanent charge distributions, respectively. After substituting Eqs. (9) and Eq. (8c), one obtains the change in energy for a small perturbation $\Delta v(\mathbf{r})$:

$$\begin{aligned} &E_{v_0+\Delta v}[\rho_0+\Delta\rho, u_0+\Delta u] - E_{v_0}[\rho_0, u_0] \approx \\ &\int \rho_0(\mathbf{r})\Delta v(\mathbf{r})d\mathbf{r} + \frac{1}{2}\int \Delta\rho(\mathbf{r})\Delta v(\mathbf{r})d\mathbf{r} \end{aligned} \tag{11}$$

where $\Delta\rho(\mathbf{r})$ is found by solving Eqs. (9) and Eq. (8c). The second term, often called the polarization or induction energy, is quadratic in $\Delta v(\mathbf{r})$ because we consider the linear response of $\Delta\rho(\mathbf{r})$ to $\Delta v(\mathbf{r})$. One can also show that it is always negative.[122] A comparison of Eq. (10) with Eq. (11), confirms that, in the linear response regime, the energy required to polarize the molecule equals half the energy gained through the interaction of the perturbation with the induced $\Delta\rho$.[8]

### C. Expansion in a Finite Basis

To obtain a practical ERFF, one must expand density and auxiliary potential fluctuations in a finite basis:[42,66,67]

$$\Delta\rho(\mathbf{r}) = \sum_k C_k f_k(\mathbf{r}) \tag{12}$$

$$\Delta u(\mathbf{r}) = \sum_k U_k g_k(\mathbf{r}) \tag{13}$$

In principle, one may take any finite set of basis functions to obtain an approximate ERFF: the density and potential basis sets need not be related; they do not even have to contain the same number of basis functions. Clearly, this approach is general enough to extend the ACKS2 model to arbitrary atomic multipoles and/or off-center basis functions. In

practice, however, density and potential basis sets are always carefully constructed and well-balanced. As will be shown below, the choice of the basis functions completely fixes all ERFF parameters and hence determines the accuracy of the model and the robustness of the parameters. A proper choice of the potential basis should span the relevant eigenmodes of the implicit response kernel, $\chi^{\text{imp}}(\mathbf{r},\mathbf{r}')$.[122,123] The density basis should span the density deformations that correspond to the soft modes of the electron density.

Substitution of the basis-set expansions in Eq. (9a), followed by a multiplication with $f_k(\mathbf{r})$ and integration over $\mathbf{r}$, leads to the following algebraic equation:

$$\sum_\ell \eta_{k\ell} C_\ell + V_k - \sum_\ell O_{k\ell} U_\ell = \Delta\mu\, D_k \quad \forall\, k \tag{14}$$

with

$$\eta_{k\ell} = \iint \eta^{\text{exp}}(\mathbf{r},\mathbf{r}') f_k(\mathbf{r}) f_\ell(\mathbf{r}') d\mathbf{r} d\mathbf{r}', \tag{15}$$

$$O_{k\ell} = \int f_k(\mathbf{r}) g_l(\mathbf{r}) d\mathbf{r}, \tag{16}$$

$$V_k = \int f_k(\mathbf{r}) \Delta v(\mathbf{r}) d\mathbf{r}, \tag{17}$$

$$D_k = \int f_k(\mathbf{r}) d\mathbf{r}. \tag{18}$$

Similar manipulations of Eq. (9b) and Eq. (8c), lead to the remaining algebraic equations needed to approximate the response:

$$\sum_\ell \chi_{k\ell} U_\ell - \sum_\ell C_k O_{k\ell} = 0 \quad \forall\, k \tag{19}$$

$$\sum_\ell D_\ell C_\ell = 0 \tag{20}$$

with

$$\chi_{k\ell} = \iint \chi^{\text{imp}}(\mathbf{r},\mathbf{r}') g_k(\mathbf{r}) g_\ell(\mathbf{r}') d\mathbf{r} d\mathbf{r}'. \tag{21}$$

These working equations are a generalization of the original ACKS2 model:[79] they are valid for any underlying variational electronic structure theory and for any choice of density and potential basis functions. Obviously, the quality of an ERFF will crucially depend on a proper choice of underlying theory and basis sets.

In a conventional empirical force field, the distinction between "model" and "parameters" is obvious. Since the generalized ACKS2 model does not contain adjustable parameters, we

must clarify how the terms "model" and "parameter" are used in this work. The adjustable choices in this work are the underlying theory, the choice of the basis functions and the molecular geometry for which the model is constructed. These choices fix the matrix elements $\eta_{k\ell}$, $O_{k\ell}$ and $\chi_{k\ell}$. Hence we use the term "parameters" for the numerical values of all these matrix elements. The mathematical expression to obtain the parameters and the equations in which they occur are referred to as the "model". This may be different from more conventional ERFFs, e.g. off-diagonal elements are often simple functions of the molecular geometry and diagonal parameters are adjustable instead of being fixed by the model.

Finally, some remarks on the numerical implementation are in order. First, when the expansion coefficients of the changes in the auxiliary potential are not uniquely determined, one may include additional equations to just fix these degrees of freedom. Such additional equations will not affect the induced density or the polarization energy. Second, without making assumptions on the matrix elements $\eta_{k\ell}$, $O_{k\ell}$ and $\chi_{k\ell}$, the computational cost of solving Eqs. (14), (19) and (20) is $O(N^3)$ where $N$ is the number of density and potential basis functions of the entire system. For an efficient force-field implementation, the cost should ideally be $O(N \log N)$. This can only be achieved by exploiting the sparsity or the structure of the matrix elements. Although such efforts go beyond the scope of this paper, we will briefly summarize the potential advances that can be made in future work. One should avoid that all matrix elements are explicitly constructed as that would already require an $O(N^2)$ cost. Instead, the matrix-vector products for each matrix must be implemented with an $O(N \log N)$ or $O(N)$ cost. In practice, the explicit hardness kernel, $\eta_{k\ell}$, is mainly governed by classical electrostatic interactions for which $O(N)$ or $O(N \log N)$ methods exist, e.g. the fast multipole method.[124] When the density and potential basis functions are designed to be local, the overlap matrix, $O_{k\ell}$, becomes sparse, which results in an $O(N)$ cost of the matrix-vector product. The structure of the implicit response matrix, $\chi_{k\ell}$, is less obvious. For dielectric systems this matrix is found to be near-sighted[125] and thus sparse, leading again to an $O(N)$ cost of the matrix-vector product. Furthermore, these matrix-vector products must be used in an iterative solver that only requires $O(\log N)$ iterations to converge. However, without more profound insight in the conditioning of the equations, it is hard to estimate the rate of convergence that can be achieved in future work.

### D. Application to KS-DFT with a Semi-local XC Functional

For the numerical validation in this paper, we will focus on Kohn-Sham DFT with a semi-local exchange-correlation functional. In this case, the explicit part of the energy functional takes the following form:

$$E^{\text{exp}}[\rho] = \frac{1}{2}\iint \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + E_{xc}[\rho] \tag{22}$$

where the first term is the Hartree energy and the second term is the exchange-correlation functional. The auxiliary wavefunction, $\Psi$, is a single Slater determinant of Kohn-Sham orbitals and $W[\Psi]$ is the Kohn-Sham kinetic energy:

$$W[\Psi] = \sum_{i\in\text{Occ.}} \int \phi_i^*(\mathbf{r}) \left(-\frac{1}{2}\nabla^2\right) \phi_i(\mathbf{r}) d\mathbf{r} \tag{23}$$

where $\phi_i(\mathbf{r})$ are the occupied Kohn-Sham orbitals.

The expressions for the explicit hardness matrix and implicit response matrix take the following forms:

$$\eta_{kl} = \iint \left(\frac{1}{|\mathbf{r}-\mathbf{r}'|} + \frac{\delta^2 E_{xc}[\rho]}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')}\right) f_k(\mathbf{r}) f_l(\mathbf{r}') d\mathbf{r}d\mathbf{r}' \tag{24}$$

$$\chi_{kl} = \sum_{\substack{i\in\text{Occ.}\\ j\in\text{Virt.}}} \frac{1}{\epsilon_i - \epsilon_j} \left(\int \phi_i^*(\mathbf{r}) g_k(\mathbf{r}) \phi_j(\mathbf{r}) d\mathbf{r}\right) \left(\int \phi_j^*(\mathbf{r}) g_l(\mathbf{r}) \phi_i(\mathbf{r}) d\mathbf{r}\right) + \text{c.c.} \tag{25}$$

where $\phi_j(\mathbf{r})$ are the virtual Kohn-Sham orbitals. In practice, the Hartree term is the dominant contribution to the explicit hardness matrix,[126] especially when the basis functions $f_k(\mathbf{r})$ and $f_l(\mathbf{r}')$ are localized on distant atoms. In that case, the semi-local exchange-correlation contribution decreases exponentially with distance between the basis functions.

It is insightful to compare Eq. (24) and Eq. (25) with those of related techniques in the literature. Tabacchi *et al.* derived hardness-like second-order ERFF parameters, as in Eq. (24).[42,58] Instead of using an implicit functional for the kinetic energy, the hardness kernel of a semi-local kinetic energy approximation was included. Even though that yields the exact parameters in the context of orbital free semi-local DFT, i.e. the Jellium model, this DFT approximation is very coarse, only suitable for metallic systems,[127–129] and thus not applicable to the majority of PFF applications involving dielectric systems. Strangely, nearly all successful ERFFs can be justified with such a formalism. From a theoretical DFT perspective, however, this Jellium approach only works reasonably well for dielectric

systems when variable charges are deliberately kept fixed. Some methods (e.g. NEMO[61,62]) also suggest ERFF parameters of the form of Eq. (25) without including additional hardness parameters.[59,60] Because such parameters are only valid for systems of non-interacting fermions, the derived parameters must be rescaled empirically to obtain a quantitatively accurate ERFF.[59]

## III. NUMERICAL VALIDATION

This section presents the numerical validation of the generalized ACKS2 model using a set of 110 organic molecules. First, a computational protocol is presented to fix the basis functions needed to derive ACKS2 parameters from a KS-DFT wavefunction. A series of density and potential basis sets with increasing degree of completeness is defined, which allows us to show that the ACKS2 equations can quantitatively reproduce reference dipole polarizability tensors in the complete basis set limit.

### A. Set of Molecules and Reference Data

Initial Cartesian coordinates were downloaded from the PubChem database[133] for all available molecules with the chemical formula $C_4O_2H_8$. Because this set contains duplicates, InChIKeys[130] were computed for all molecules with OpenBabel.[134,135] The first 14 characters of an InChIKey are determined by the connectivity of the elements while the remainder also accounts for isotopes, chirality, etc. Therefore, only the first 14 characters of each InChIKey were used to make a subset of all unique molecules. Molecular dianions were discarded. This lead to a set of 110 neutral organic molecules (depicted in Fig. 1) with very diverse bonding patterns, heats of formation, polarizabilities, atomic partial charges and organic groups. The motivation for this set is twofold. First, using conventional calibration techniques, it is very tedious to construct polarizable force fields for all these molecules. Second, the results for all molecules are easily compared because they have the same size and composition, e.g. to test the transferability of parameters.

All molecular geometries were optimized with Gaussian09[136] using Kohn-Sham DFT with the BLYP exchange-correlation functional[102,103] and the 6-311++G(3df,3pd) basis set,[137–139] followed by a computation of the dipole polarizability tensor with the Coupled Perturbed

Kohn-Sham (CPKS) method[91,92] at the same level of theory. The results of these computations are stored in formatted checkpoint (FCHK) files to facilitate the post-processing.

### B. Basis Set and ACKS2 Parameters

In this subsection, a procedure is outlined to obtain density and potential basis sets with which the ACKS2 results converge systematically to the reference data. Our primary objective is to validate the ACKS2 model and hence we strictly focus on the accuracy of the model rather than the robustness of the parameters. The procedure in this section is the same for each molecule but the resulting basis functions may be different for similar atoms in different molecules. Hence we do not expect robust parameters *a priori*.

The Hirshfeld-E method[140] was used to define atoms-in-molecules (AIM) weight functions, $w_A(\mathbf{r})$. These are traditionally used to partition the molecular electron density into atomic contributions: $\rho_A(\mathbf{r}) = w_A(\mathbf{r})\rho(\mathbf{r})$. Moreover, one may use the same AIM weight functions to condense response kernels into atom-pair properties.[125,141–143] Inspired by the partitioning of response kernels, we use AIM weight functions to define the potential basis set:

$$g_{l(A,\ell,m)}(\mathbf{r}) = w_A(\mathbf{r}) R_\ell^m(\mathbf{r} - \mathbf{R}_A) \tag{26}$$

where $R_\ell^m$ is a real solid harmonic with angular quantum number $\ell$ and magnetic quantum number $m$. In practice, the series of spherical harmonics must be truncated at some value $\ell_{\max}$ to obtain a finite basis. In this work we considered $\ell_{\max} = 0$, 1, 2, 3, 4 or 5, corresponding to a PFF with charges, charges+dipoles, charges+dipoles+quadrupoles, and so on. We assume that this potential basis set is sufficient (in the limit of large $\ell_{\max}$) to span all relevant eigenmodes of the response kernel. The number of basis functions is denoted as $M(\ell_{\max}) = (\ell_{\max} + 1)^2$.

To construct the density basis set, the first-order density response of the non-interacting fermions to each potential basis function was computed as a starting point:

$$\tilde{f}_l(\mathbf{r}) = \int \chi(\mathbf{r}, \mathbf{r}') g_l(\mathbf{r}') d\mathbf{r}' \tag{27}$$

These density responses effectively capture the soft modes in the electron density and are therefore useful as a density basis.[42] (It would be even better to compute the density response of the interacting electrons but the current approach is easier to implement and already

sufficient for our purposes.) There is one drawback, however: one can construct a linear combination of potential basis functions that equals a constant because $\sum_{A=1}^{N} w_A = 1$. The (non-interacting) density response to this linear combination is zero and hence the functions $\tilde{f}_l(\mathbf{r})$ only span an $(M(\ell_{\max}) - 1)$-dimensional space. Therefore, the density responses are extended with the Fukui function of the molecule,[144] approximated as the average of the HOMO and LUMO density,[145] to obtain a set of $M(\ell_{\max}) + 1$ *initial* density basis functions that span a $M(\ell_{\max})$-dimensional space. Singular-value-decomposition was then used to derive a set of $M(\ell_{\max})$ basis functions bi-orthogonal to the potential basis set, i.e. such that $O_{kl} = \delta_{kl}$. Following this procedure, every density basis function, denoted as $f_l(\mathbf{r})$, can be associated with a corresponding potential basis function and thus also with a given atom and spherical harmonic. This implies that we can label the density basis functions as s-type, p-type, etc.

The bi-orthogonality of the potential and basis functions results in a set of ACKS2 equations that is backward compatible with our previous paper[79] (and the SQE model[73]), where the overlap matrix was an identity matrix by construction. This seems conceptually appealing because it precludes the need to model the overlap elements in ACKS2-PFF simulations. Unfortunately, the bi-orthogonality introduces many fluctuations in the density basis functions, especially for increasing $\ell_{\max}$. This is illustrated in Fig. 2 with the s-type basis function of the epoxy-oxygen of one of the 110 molecules. Due to the non-local features in the basis functions (mostly in the low-density region) we do not expect the ERFF parameters to be transferable. As stated earlier, a high degree of parameter transferability goes beyond the scope of this paper. We primarily focus on the quantitative accuracy of the model and the basis functions are especially designed for that purpose. Yet, it is already obvious from the current analysis that transferability can only be achieved in our approach when (i) well-behaved localized basis functions are used and (ii) the overlap matrix is modeled explicitly.

ACKS2 parameters were obtained by evaluating Eq. (24) and Eq. (25) with the basis functions introduced above. Integrals involving the hardness kernel of the explicit functional were evaluated numerically using a pruned Becke-Lebedev integration grid.[146] Integrals over the exchange-correlation kernel were evaluated with first-order symmetric finite differences

| $\ell_{\max}$ | RMSD [a.u.] | Rel. RMSD [%] | Mean Error [a.u.] |
|---|---|---|---|
| 0 | 37.9 | 57.6 | −37.7 |
| 1 | 1.02 | 1.54 | −0.97 |
| 2 | 0.452 | 0.687 | −0.168 |
| 3 | 0.210 | 0.321 | −0.184 |
| 4 | 0.111 | 0.169 | −0.089 |
| 5 | 0.0591 | 0.0899 | −0.0417 |

TABLE I. Comparison of the performance of the ACKS2 model derived using the different basis sets (defined by $\ell_{\max}$, see text). RMSD stands for the root-mean-square deviation between the eigenvalues of the reference BLYP dipole polarizability tensors of all 110 test molecules and the corresponding ACKS2 results. The relative RMSD equals the RMSD divided by the RMS value of all BLYP eigenvalues. The mean error is the average difference between the BLYP reference eigenvalues and the ACKS2 results.

using the BLYP[102,103] exchange-correlation potential implemented in LibXC:[147]

$$\iint f_k(\mathbf{r}) f_l(\mathbf{r}') \left.\frac{\delta^2 E_{xc}[\rho]}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')}\right|_{\rho_0} d\mathbf{r}d\mathbf{r}' \approx \frac{1}{2\epsilon}\int f_k(\mathbf{r}) \left( \left.\frac{\delta E_{xc}[\rho]}{\delta\rho(\mathbf{r})}\right|_{\rho_0+\epsilon f_l} - \left.\frac{\delta E_{xc}[\rho]}{\delta\rho(\mathbf{r})}\right|_{\rho_0-\epsilon f_l} \right) d\mathbf{r} \quad (28)$$

with $\epsilon = 10^{-4}$. Integrals over the Hartree kernel were implemented with a Becke-Poisson solver.[148] The same Becke-Lebedev grids were used to evaluate the matrix elements in the implicit response kernel in Eq. (25).

All computations described in this section were carried out with Horton.[149]

### C. Reproduction of the Dipole Polarizability Tensor Eigenvalues

To asses the accuracy of the ACKS2 model, Table I contains the root-mean-square deviation (RMSD) between the reference eigenvalues of the dipole polarizability tensor for the 110 molecules and the corresponding ACKS2 predictions with increasing basis sets. Also the relative RMSD, i.e. the RMSD divided by RMS value of the reference eigenvalues, is included in Table I. Systematic errors are revealed by the mean error, which is just the average of the difference between reference and model eigenvalues. The main result of this

paper is that all these errors decrease exponentially with basis set size to negligibly small values for the largest basis set. The errors on the tensor components and eigenvectors converge similarly. (Results not shown.) This confirms numerically that the ACKS2 equations and the expressions for the parameters reproduce ab initio response data to arbitrary accuracy. More importantly, this is achieved without any empirical calibration of parameters or manual interference. Instead, it is now possible to assign extremely accurate ERFF parameters in a black-box fashion for a given molecular geometry. This is a significant step forward in the efficient development of reliable PFFs. However, a feasible ERFF model also constitutes simple functional forms for the geometry dependence of the parameters. The present work only shows how to compute the parameters for a given molecular geometry, a choice of density and potential basis functions and a given underlying theory.

It is also encouraging that very small errors can already be obtained by just considering fluctuating charges and dipoles. The parity plot of the eigenvalues for $\ell_{\max} = 1$ in Fig. 3 shows that there are no outliers. This good correspondence may be partially due to the use of response functions in the definition of the density basis set. More research is needed to see how accurate a charge+dipole ACKS2 model will be when more well-behaved basis functions are used.

Another interesting result is that the model with only s-type functions captures the right trends in the dipole polarizability. Fig. 3 shows that the errors in the s-type ACKS2 model are mostly systematic. This reveals the importance of charge-transfer effects in PFFs, especially when changes in polarizability between conformers or isomers are of interest. We presume that the constant shift is due to a nearly additive contribution of the atomic dipole polarizabilities that is missing in the s-type model.

Finally, it is worth noting that the ACKS2 model never overestimates the dipole polarizability in the current numerical validation. In case of $\ell_{\max} \geq 4$, some eigenvalues are slightly overestimated (by less than 0.1 a.u.), which may be due to numerical integration errors. However, we do not expect the ACKS2 model to systematically underestimate response properties in general. For example, one can lower the eigenvalues of the explicit hardness matrix by using a less complete basis set, which may lead to an overestimate of the dipole polarizability.

### D. Robustness and Interpretation of Parameters

Despite our low expectations, it is still insightful to study the robustness of the parameters with the current choice of basis functions. We foresee that with our methodology, the development of a high-quality ERFF will require local, well-behaved, small and yet complete density and potential basis sets (and the selection of a proper underlying theory). In this work, the basis sets are small and relatively complete but very non-local and ill-behaved. We will show how this causes different issues with the robustness of the atomic parameters and the geometry dependence of off-diagonal matrix elements.

Table II shows the average and the standard deviation of the diagonal charge-charge hardness matrix elements for hydrogen, carbon and oxygen atoms and for different basis sets in this work. The most notable trend is that the values increase with $\ell_{\max}$. This is mainly a consequence of the bi-orthogonality: for higher $\ell_{\max}$ the s-type density functions are orthogonal to a larger number of potential basis functions. This causes sharper fluctuations and a higher maximum for every s-type density function, see Fig. 2, which in turn increase the electrostatic self-interaction of these basis functions, leading to larger diagonal hardness parameters. If transferability of these parameters is desired between different values of $\ell_{\max}$, the bi-orthogonality must be abandoned.

A second observation is that none of the atomic hardness parameters in Table II are close to the experimental atomic properties of Parr and Pearson:[150] 12.84 eV, 10.0 eV and 12.16 eV, for hydrogen, carbon and oxygen, respectively. (We have doubled the experimental values to make them consistent with Eq. (24).[151]) There are two probable causes for the discrepancy. First, the basis functions are non-local and are therefore not representative for the specific atom, but also for their environment. Second, the bi-orthogonality causes a lot of oscillations in the density basis functions, which leads to an increased electrostatic self-interaction.

Table II shows that the standard deviation on the atomic hardness parameters is at least 10%, which is not negligible. The use of bi-orthogonalized response functions leads to very different shapes of the density basis functions, even for the same elements in similar environments, which in turn causes significant variation in the derived parameters. This may seem contradictory to the well-behaved ACKS2 parameters for the dissociation of a HCl molecule reported earlier.[79] However, in the dissociation of HCl, the atoms barely overlap and the issues arising from the bi-orthogonality of density and potential basis functions were

| Element | $\ell_{max}$ | Mean [eV] | Std. Dev. [eV] |
|---|---|---|---|
| H | 0 | 56.0 | 6.0 |
| | 1 | 112.4 | 22.2 |
| | 2 | 180.8 | 34.1 |
| | 3 | 314.7 | 57.6 |
| | 4 | 534.8 | 130.6 |
| | 5 | 976.8 | 328.6 |
| C | 0 | 35.8 | 3.2 |
| | 1 | 40.2 | 3.7 |
| | 2 | 47.1 | 3.0 |
| | 3 | 77.5 | 4.4 |
| | 4 | 129.1 | 11.4 |
| | 5 | 223.1 | 24.3 |
| O | 0 | 38.8 | 3.3 |
| | 1 | 42.0 | 3.2 |
| | 2 | 45.5 | 4.0 |
| | 3 | 54.6 | 5.1 |
| | 4 | 69.0 | 8.4 |
| | 5 | 96.0 | 12.6 |

TABLE II. Comparison of diagonal elements of the explicit charge-charge hardness matrix, obtained for the different basis sets in this work (defined by $\ell_{max}$, see text). The average and standard deviation of the diagonal elements are given after grouping them per chemical element.

not relevant.

In most ERFF models, the off-diagonal charge-charge hardness parameters are modeled as the (screened) electrostatic interaction between spherically symmetric and decaying charge densities. Figure 4 compares these off-diagonal elements in the ACKS2 model (for $\ell_{max} = 0$ and 1) with a simple interaction between point charges. Also in this comparison, it becomes clear that the non-locality and variability of the density basis function makes it impossible to model the off-diagonal hardness parameters as simple functions of the inter-atomic distance.

(Encouragingly, the order of magnitude and the trend in the scatter is still reasonable.)

## IV. CONCLUSIONS AND PROSPECTIVES

This paper presents a new electronic response force field that can be used as the polarization model in a polarizable force field. The new model is a generalization of ACKS2 toward any underlying variational electronic structure theory and with completely general density and potential basis sets. The model is numerically validated with the reproduction of the ab initio dipole polarizability tensors of a set of 110 organic molecules. An accurate model can be obtained as soon as fluctuating charges and dipoles are considered. The errors further decrease exponentially as higher fluctuating atomic multipoles are included. In the limit of a complete basis set, ACKS2 reproduces the underlying theory exactly.

The parameters in the model are defined as expectation values of density and potential basis functions. For a given molecular geometry, basis sets and underlying theory, all parameters can be computed directly. No further iterative calibration or manual tuning is needed which clearly facilitates the development of an ERFF. However, a computationally feasible ERFF also constitutes simple functional forms for the geometry dependence of the parameters, which is a topic of ongoing research. The basis functions used for the numerical validation in this work focused exclusively on the accuracy of the corresponding ACKS2 model. Due to their non-local and irregular shapes, the obtained parameters are not simply transferable and have no trivial geometry dependence. Ongoing research focuses on methods to construct well-behaved and local basis functions, without sacrificing accuracy. Additionally, we will also work on the implementation of parameters for higher levels of theory and the inclusion of first-order parameters, so that the model can also reproduce permanent atomic multipoles.

## ACKNOWLEDGMENTS

T.V. and S.V. acknowledge the Foundation of Scientific Research - Flanders (FWO), the Research Board of Ghent University (BOF), and BELSPO in the frame of IAP/7/05 for their financial support. P.W.A. gratefully acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada. The computational resources and

services used were provided by Ghent University (Stevin Supercomputer Infrastructure).

## REFERENCES

[1]M. S. Daw, S. M. Foiles, and M. I. Baskes, Mater. Sci. Rep. **9**, 251 (1993).

[2]P. Barnes, J. L. Finney, J. D. Nicholas, and J. E. Quinn, Nature **282**, 459 (1979).

[3]L. Perera and M. L. Berkowitz, J. Chem. Phys. **95**, 1954 (1991).

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

[5]A. Warshel, M. Kato, and A. V. Pisliakov, J. Chem. Theory Comput. **3**, 2034 (2007).

[6]J.-P. Piquemal, R. Chelli, P. Procacci, and N. Gresh, *J. Phys. Chem. A*, J. Phys. Chem. A **111**, 8170 (2007).

[7]G. A. Cisneros, M. Karttunen, P. Ren, and C. Sagui, Chem. Rev. **114**, 779 (2013).

[8]H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, J. Phys. Chem. **91**, 6269 (1987).

[9]J. B. Foresman and C. L. Brooks, J. Chem. Phys. **87**, 5892 (1987).

[10]U. Góra, R. Podeszwa, W. Cencek, and K. Szalewicz, J. Chem. Phys. **135**, 224102 (2011).

[11]U. Góra, W. Cencek, R. Podeszwa, A. van der Avoird, and K. Szalewicz, J. Chem. Phys. **140**, 194101 (2014).

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

[13]G. V. Lewis and C. R. A. Catlow, J. Phys. C: Solid State Phys. **18**, 1149 (1985).

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

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

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

[17]T. A. Halgren and W. Damm, Curr. Opin. Struct. Biol. **11**, 236 (2001).

[18]A. Grossfield, P. Ren, and J. W. Ponder, J. Am. Chem. Soc. **125**, 15671 (2003).

[19]S. Tazi, J. J. Molina, B. Rotenberg, P. Turq, R. Vuilleumier, and M. Salanne, J. Chem. Phys. **136**, 114507 (2012).

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

[21]J. R. Maple, Y. Cao, W. Damm, T. A. Halgren, G. A. Kaminski, L. Y. Zhang, and R. A. Friesner, J. Chem. Theory Comput. **1**, 694 (2005).

[22]A. Savelyev and A. D. MacKerell, J. Comput. Chem. **35**, 1219 (2014).

[23]T. Pham, K. A. Forrest, P. Nugent, Y. Belmabkhout, R. Luebke, M. Eddaoudi, M. J. Zaworotko, and B. Space, J. Phys. Chem. C **117**, 9340 (2013).

[24]A. L. Mullen, T. Pham, K. A. Forrest, C. R. Cioce, K. McLaughlin, and B. Space, J. Chem. Theory Comput. **9**, 5421 (2013).

[25]J. Cirera, J. C. Sung, P. B. Howland, and F. Paesani, J. Chem. Phys. **137**, 054704 (2012).

[26]W. B. Dapp and M. H. Müser, J. Chem. Phys. **139**, 064106 (2013).

[27]W. B. Dapp and M. H. Müser, Euro. Phys. J. B **86**, 337 (2013).

[28]J. W. Caldwell and P. A. Kollman, J. Am. Chem. Soc. **117**, 4177 (1995).

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

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

[31]X. Li, S. Y. Ponomarev, Q. Sa, D. L. Sigalovsky, and G. A. Kaminski, J. Comput. Chem. **34**, 1241 (2013).

[32]P. Schyman and W. L. Jorgensen, J. Phys. Chem. Lett. **4**, 468 (2013).

[33]P. Ren and J. W. Ponder, J. Phys. Chem. B **107**, 5933 (2003).

[34]W. Xie, J. Pu, A. D. Mackerell, and J. Gao, J. Chem. Theory Comput. **3**, 1878 (2007).

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

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

[37]P. Tröster, K. Lorenzen, M. Schwörer, and P. Tavan, J. Phys. Chem. B **117**, 9486 (2013).

[38]B. Lin, X. He, and A. D. MacKerell, J. Phys. Chem. B **117**, 10572 (2013).

[39]X. He, P. E. M. Lopes, and A. D. MacKerell, Biopolymers **99**, 724 (2013).

[40]A. de Oliveira Cavalcante, M. C. C. Ribeiro, and M. S. Skaf, J. Chem. Phys. **140**, 144108 (2014).

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

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

[43]T. G. Kucukkal and S. J. Stuart, J. Phys. Chem. B **116**, 8733 (2012).

[44]P. E. M. Lopes, J. Huang, J. Shim, Y. Luo, H. Li, B. Roux, and A. D. MacKerell, J. Chem. Theory Comput. **9**, 5430 (2013).

[45]G. Kamath, S. A. Deshmukh, and S. K. R. S. Sankaranarayanan, J. Phys.: Condens. Matter **25**, 305003 (2013).

[46]F. Manzoni and P. Söderhjelm, J. Comput.-Aid. Mol. Des. **28**, 235 (2014).

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

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

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

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

[51]M. Salanne, B. Rotenberg, S. Jahn, R. Vuilleumier, C. Simon, and P. A. Madden, Theor. Chem. Acc. **131**, 1143 (2012).

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

[53]L. Huang and B. Roux, J. Chem. Theory Comput. **9**, 3543 (2013).

[54]P. A. Madden, R. Heaton, A. Aguado, and S. Jahn, J. Mol. Struct. THEOCHEM **771**, 9 (2006).

[55]A. Laio, S. Bernard, G. L. Chiarotti, S. Scandolo, and E. Tosatti, Science **287**, 1027 (2000).

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

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

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

[59]G. Karlström, Theoret. Chem. Acc. **60**, 535 (1982).

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

[61]O. Engkvist, P.-O. Åstrand, and G. Karlström, *Chem. Rev.*, Chem. Rev. **100**, 4087 (2000).

[62]D. Hagberg, G. Karlström, B. O. Roos, and L. Gagliardi, *J. Am. Chem. Soc.*, J. Am. Chem. Soc. **127**, 14250 (2005).

[63]K. Ansorg, M. Tafipolsky, and B. Engels, J. Phys. Chem. B **117**, 10093 (2013).

[64]W. Yu, P. E. M. Lopes, B. Roux, and A. D. MacKerell, J. Chem. Phys. **138**, 034508 (2013).

[65]O. M. Szklarczyk, S. J. Bachmann, and W. F. van Gunsteren, J. Comput. Chem. **35**, 789 (2014).

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

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

[68]T. Ishida, J. Phys. Chem. A **112**, 7035 (2008).

[69]I. V. Bodrenko and F. Della Sala, J. Chem. Phys. **139**, 144109 (2013).

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

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

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

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

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

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

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

[77]T. Verstraelen and P. Bultinck, Spectrochim. Acta A-M (2013), published online. DOI:10.1016/j.saa.2013.10.124.

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

[79]T. Verstraelen, P. W. Ayers, V. Van Speybroeck, and M. Waroquier, J. Chem. Phys. **138**, 074108 (2013).

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

[81]T. Cremer, C. Kolbeck, K. R. J. Lovelock, N. Paape, R. Wölfel, P. S. Schulz, P. Wasserscheid, H. Weber, J. Thar, B. Kirchner, F. Maier, and H.-P. Steinrück, Chem.-Eur. J. **16**, 9018 (2010).

[82]M. Soniat and S. W. Rick, J. Chem. Phys. **137**, 044511 (2012).

[83]A. J. Lee and S. W. Rick, J. Phys. Chem. Lett. **3**, 3199 (2012).

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

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

[86]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).

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

[88]S. M. Kandathil, T. L. Fletcher, Y. Yuan, J. Knowles, and P. L. A. Popelier, J. Comput. Chem. **34**, 1850 (2013).

[89]R. F. Borkman, G. Simons, and R. G. Parr, J. Chem. Phys. **50**, 58 (1969).

[90]R. Chaudret, N. Gresh, G. A. Cisneros, A. Scemama, and J.-P. Piquemal, Can. J. Chem.

**91**, 804 (2013).

[91]G. Diercksen and R. McWeeny, J. Chem. Phys. **44**, 3554 (1966).

[92]C. E. Dykstra and P. G. Jasien, Chem. Phys. Lett. **109**, 388 (1984).

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

[94]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).

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

[96]C.-M. Ionescu, S. Geidl, R. Svobodová Vařeková, and J. Koča, J. Chem. Inf. Model. **53**, 2548 (2013).

[97]E. Engel and R. M. Dreizler, J. Comput. Chem. **20**, 31 (1999).

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

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

[100]A. D. Becke, J. Chem. Phys. **98**, 1372 (1993).

[101]A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).

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

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

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

[105]J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria, Phys. Rev. Lett. **91**, 146401 (2003).

[106]A. Szabo and N. S. Ostlund, "Modern quantum chemistry: Introduction to advanced electronic structure theory," (Dover Publications, Mineola, NY, 1996) Chap. 4. Configuration Interaction, pp. 231–270, 1st (revised) ed.

[107]G. K.-L. Chan and M. Head-Gordon, J. Chem. Phys. **116**, 4462 (2002).

[108]G. K.-L. Chan and S. Sharma, Annu. Rev. Phys. Chem. **62**, 465 (2011).

[109]S. Wouters, W. Poelmans, P. W. Ayers, and D. Van Neck, Comput. Phys. Commun. **185**, 1501 (2014).

[110]P. R. Surján, in *Correlation and Localization*, Top. Curr. Chem., Vol. 203, edited by R. J. Bartlett, F. Bogár, D. L. Cooper, B. Kirtman, W. Klopper, W. Kutzelnigg, N. H. March, P. G. Mezey, H. Müller, J. Noga, J. Paldus, J. Pipek, M. Raimondi, I. Røeggen, J. Q. Sun, P. R. Surján, C. Valdemoro, and S. Vogtner (Springer Berlin Heidelberg, 1999) pp.

63–88.

[111]P. A. Limacher, P. W. Ayers, P. A. Johnson, S. De Baerdemacker, D. Van Neck, and P. Bultinck, J. Chem. Theory Comput. **9**, 1394 (2013).

[112]P. A. Johnson, P. W. Ayers, P. A. Limacher, S. D. Baerdemacker, D. Van Neck, and P. Bultinck, Comput. Theor. Chem. **1003**, 101 (2013).

[113]E. H. Lieb, I. J. Quant. Chem. **24**, 243 (1983).

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

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

[116]H. Eschrig, *The Fundamentals of Density Functional Theory*, 2nd ed. (Edition am Gutenbergplatz, Leipzig, Germany, 2003).

[117]P. Ayers and M. Levy, J. Chem. Sci. **117**, 507 (2005).

[118]P. W. Ayers and W. Yang, J. Chem. Phys. **124**, 224108 (2006).

[119]P. W. Ayers, S. Golden, and M. Levy, J. Chem. Phys. **124**, 054101 (2006).

[120]P. Ayers and P. Fuentealba, Phys. Rev. A **80**, 032510 (2009).

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

[122]S. Liu, T. Li, and P. W. Ayers, J. Chem. Phys. **131**, 114106 (2009).

[123]C. Cárdenas, P. W. Ayers, and A. Cedillo, J. Chem. Phys. **134**, 174103 (2011).

[124]L. Greengard and V. Rokhlin, J. Comput. Phys. **135**, 280 (1997).

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

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

[127]E. Smargiassi and P. A. Madden, Phys. Rev. B **49**, 5220 (1994).

[128]L. Hung and E. A. Carter, J. Phys. Chem. C **115**, 6269 (2011).

[129]J. Xia, C. Huang, I. Shin, and E. A. Carter, J. Chem. Phys. **136**, 084102 (2012).

[130]S. Heller, A. McNaught, S. Stein, D. Tchekhovskoi, and I. Pletnev, J. Cheminform. **5**, 7 (2013).

[131]W. Humphrey, A. Dalke, and K. Schulten, J. Mol. Graph. **14**, 33 (1996).

[132]J. Stone, *An Efficient Library for Parallel Ray Tracing and Animation*, Master's thesis, Computer Science Department, University of Missouri-Rolla (1998).

[133]E. Bolton, Y. Wang, P. A. Thiessen, and S. H. Bryant, in *Annual Reports in Computational Chemistry*, Vol. 4 (American Chemical Society, Washington, DC, 2008) Chap. 12. PubChem: Integrated Platform of Small Molecules and Biological Activities,

http://pubchem.ncbi.nlm.nih.go (last accessed: Aug 17, 2014).

[134]R. Guha, M. T. Howard, G. R. Hutchison, P. Murray-Rust, H. Rzepa, C. Steinbeck, J. Wegner, and E. L. Willighagen, J. Chem. Inf. Model. **46**, 991 (2006).

[135]N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison, J. Cheminform. **3**, 33 (2011).

[136]M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, *Gaussian 09 Revision D.01* (Gaussian Inc.: Wallingford CT, 2013).

[137]R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople, J. Chem. Phys. **72**, 650 (1980).

[138]T. Clark, J. Chandrasekhar, G. W. Spitznagel, and P. Von Ragué Schleyer, J. Comput. Chem. **4**, 294 (1983).

[139]M. J. Frisch, J. A. Pople, and J. S. Binkley, J. Chem. Phys. **80**, 3265 (1984).

[140]T. Verstraelen, P. W. Ayers, V. Van Speybroeck, and M. Waroquier, J. Chem. Theory Comput. **9**, 2221 (2013).

[141]J. Cioslowski, Mol. Phys. **88**, 621 (1996).

[142]J. G. Ángyán, J. Mol. Struct. THEOCHEM **501–502**, 379 (2000).

[143]A. Krishtal, P. Senet, M. Yang, and C. Van Alsenoy, J. Chem. Phys. **125**, 034312 (2006).

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

[145]W. Yang, R. G. Parr, and R. Pucci, J. Chem. Phys. **81**, 2862 (1984).

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

[147]M. A. L. Marques, M. J. T. Oliveira, and T. Burnus, Comput. Phys. Comm. **183**, 2272 (2012).

[148]A. D. Becke and R. M. Dickson, J. Chem. Phys. **89**, 2993 (1988).

[149](2014), t. Verstraelen, S. Vandenbrande, M. Chan, F. H. Zadeh, C. González, P. A. Limacher, A. Malek; Horton 1.2.1, http://theochem.github.com/horton/.

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

[151]R. G. Pearson, *J. Chem. Sci.*, J. Chem. Sci. **117**, 369 (2005).

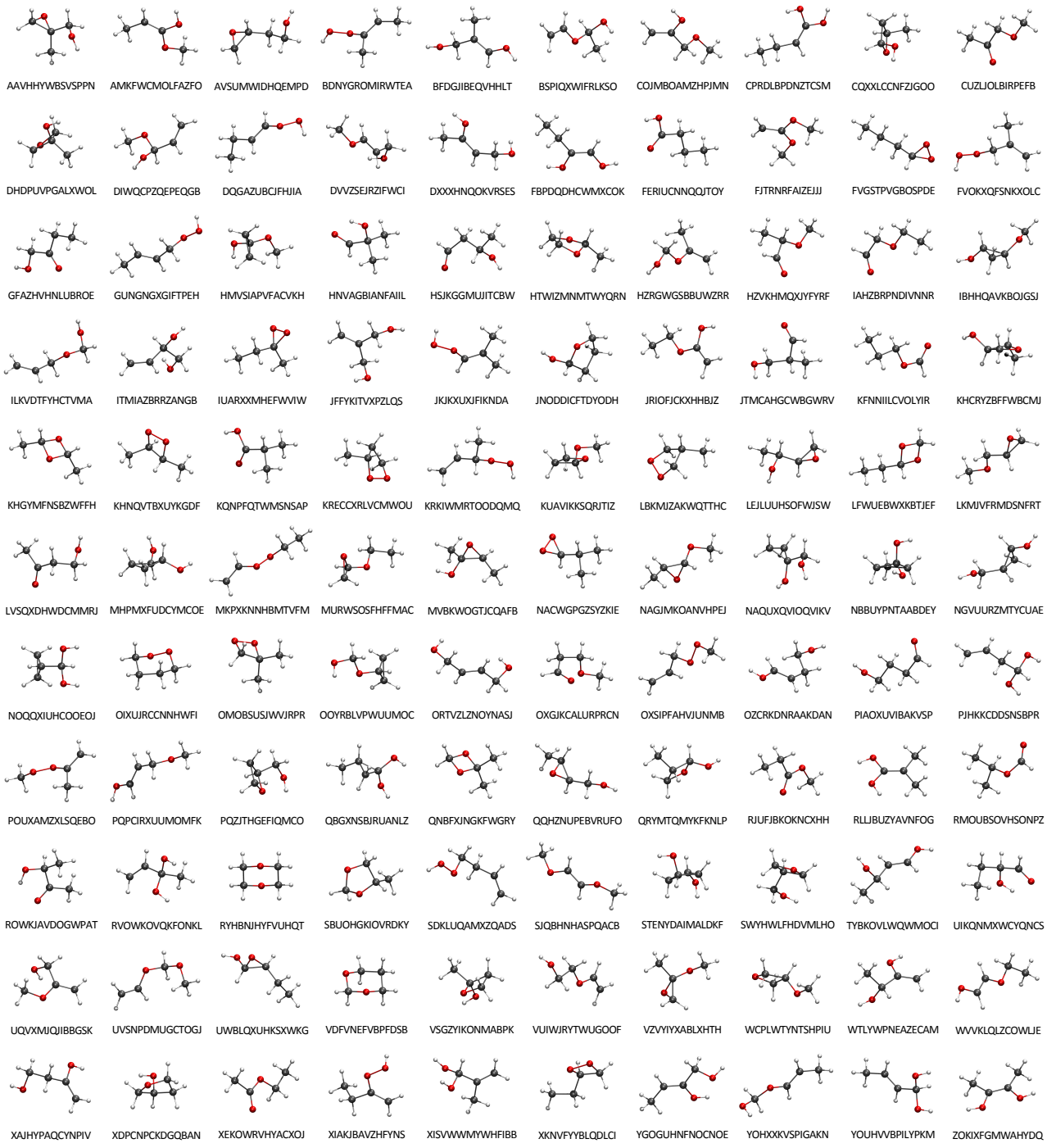


FIG. 1. Visualization of the 110 $C_4O_2H_8$ molecules in the dataset. The first 14 characters of the InChIKey[130] are printed below each image. 3D visualizations were made with VMD.[131,132]

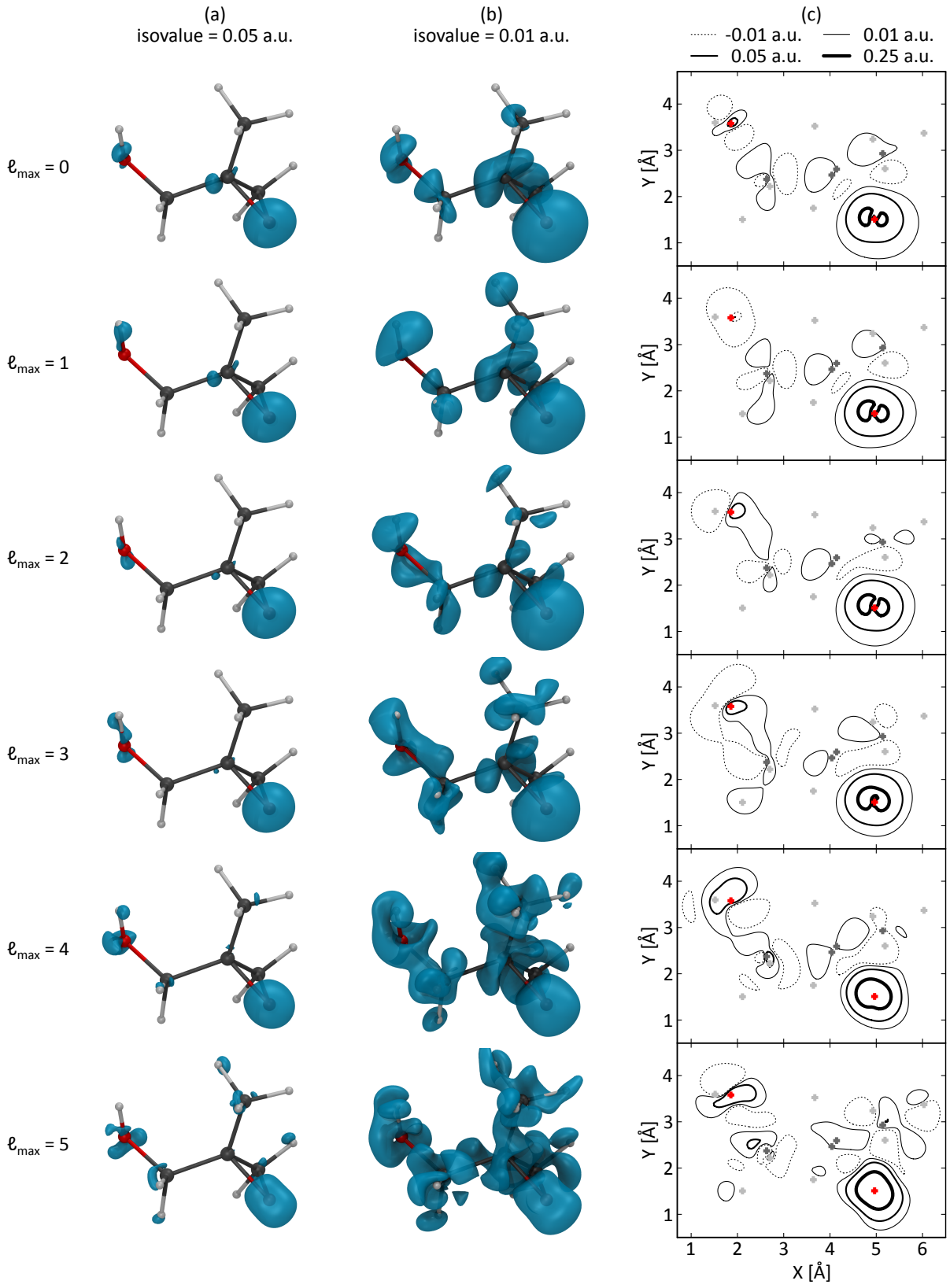


FIG. 2. Visualization of the s-type density basis function of the epoxy oxygen in 2-methyloxiranemethanal, one of the 110 $C_4O_2H_8$ molecules, for different values of $\ell_{\max}$: (a) isosurface at 0.05 a.u. (b) isosurface at 0.01 a.u. (c) Contour plots in a plane through the epoxy oxygen, including atom positions projected on this plane (crosses). 3D visualizations were made with VMD.[131,132]

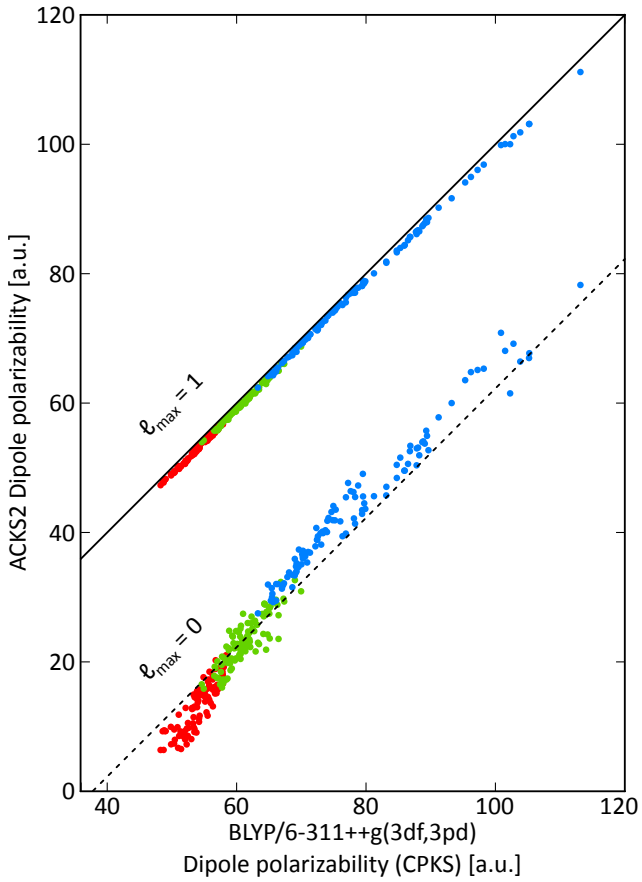


FIG. 3. Parity plot of the dipole polarizability eigenvalues of all 110 molecules: red=first, green=second and blue=third eigenvalue. Two datasets are included: the lowest set is obtained with the basis and potential basis set generated with $\ell_{\max} = 0$ (fluctuating charges only), while the higher set is generated with $\ell_{\max} = 1$ (fluctuating charges and dipoles). The first bisector is plotted as a black solid line. The dashed line is parallel to the first bisector, i.e. shifted down by the mean error for $\ell_{\max} = 0$. (See Table I).

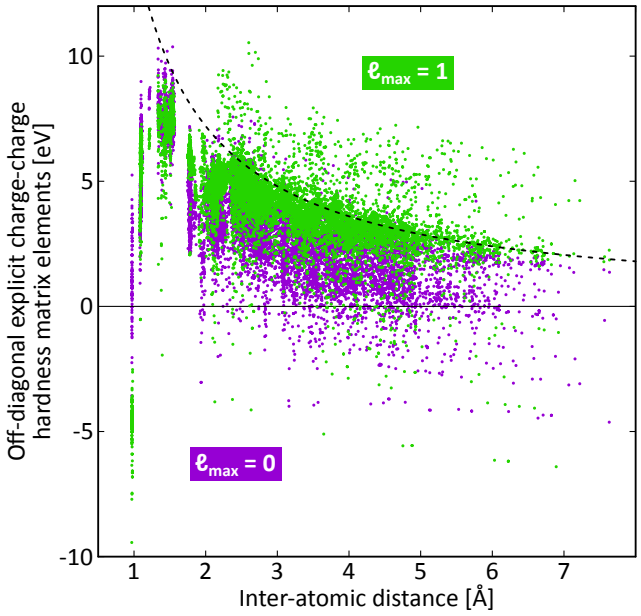


FIG. 4. Comparison of the off-diagonal elements of the explicit charge-charge hardness matrix, obtained for two different basis sets (defined by $\ell_{\max}$, see text). The conventional $1/4\pi\epsilon_0 r$ relation from the Electronegativity Equalization Method (EEM) is added as a dashed curve for comparison.