

RESEARCH ARTICLE | JUNE 21 2024

A polarizable valence electron density based force field for high-energy interactions between atoms and molecules

Special Collection: [2024 JCP Emerging Investigators Special Collection](#)

José Romero ; Paulo Limão-Vieira ; Thana Maihom ; Kersti Hermansson ; Michael Probst  



J. Chem. Phys. 160, 235101 (2024)

<https://doi.org/10.1063/5.0210949>



The Journal of Chemical Physics

Special Topics Open
for Submissions

[Learn More](#)

A polarizable valence electron density based force field for high-energy interactions between atoms and molecules

Cite as: J. Chem. Phys. 160, 235101 (2024); doi: 10.1063/5.0210949

Submitted: 28 March 2024 • Accepted: 29 May 2024 •

Published Online: 21 June 2024



View Online



Export Citation



CrossMark

José Romero,^{1,2,a)}  Paulo Limão-Vieira,²  Thana Maihom,^{3,4}  Kersti Hermansson,⁵ 
and Michael Probst^{1,3,b)} 

AFFILIATIONS

¹Institute of Ion Physics and Applied Physics, University of Innsbruck, Technikerstraße 25, 6020 Innsbruck, Austria

²Atomic and Molecular Collisions Laboratory, CEFITEC, Department of Physics, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal

³School of Molecular Science and Engineering, Vidyasirimedhi Institute of Science and Technology, Rayong 21210, Thailand

⁴Department of Chemistry, Faculty of Liberal Arts and Science, Kasetsart University, Kamphaeng Saen Campus, Nakhon Pathom 73140, Thailand

⁵Department of Chemistry-Ångström, Uppsala University, Box 538, SE-75121 Uppsala, Sweden

Note: This paper is part of the 2024 JCP Emerging Investigators Special Collection.

^{a)}Electronic mail: j.romero@alumni.fct.unl.pt

^{b)}Author to whom correspondence should be addressed: michael.probst@uibk.ac.at

ABSTRACT

High-accuracy molecular force field models suited for hot gases and plasmas are not as abundant as those geared toward ambient pressure and temperature conditions. Here, we present an improved version of our previous electron-density based force field model that can now account for polarization effects by adjusting the atomic valence electron contributions to match *ab initio* calculated Mulliken partial charges. Using a slightly modified version of the Hohenberg–Kohn theorem, we also include an improved theoretical formulation of our model when applied to systems with degenerate ground states. We present two variants of our polarizable model, fitted from *ab initio* reference data calculated at CCSD(T)/cc-pVTZ and CCSD(T)/CEP-31G levels of theory, that both accurately model water dimer interaction energies. Further improvements include the additional interaction components with fictitious non-spherically symmetric, yet atom-centered, electron densities and fitting the exchange and correlation coefficients against analytical expressions. The latter removes all unphysical oscillations that are observed in the previous non-polarizable variant of our force field.

© 2024 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC) license (<https://creativecommons.org/licenses/by-nc/4.0/>). <https://doi.org/10.1063/5.0210949>

I. INTRODUCTION

In computational chemistry and physics, the rationale behind employing force fields stems from the ability to “integrate out” the electronic degrees of freedom. One can thereby avoid solving the high-dimensional partial differential Schrödinger equation in favor of a tractable many-body problem of coupled second-order ordinary differential equations (ODEs) embodied in Newton’s laws of motion.

Force fields are typically formulated as sums of inverse power laws, with the Lennard-Jones 12-6 potential form standing out as a prominent example. Some examples of force field models that are often used in Molecular Dynamics (MD) simulations and that rely on such inverse power laws include Jorgensen’s Transferable Intermolecular Potential (TIP) family of water models (e.g., TIP3P and TIP4P),^{1–4} Optimized Potential for Liquid Simulations (OPLS),^{5–7} and Assisted Model Building and Energy Refinement (AMBER)^{8,9} family of force field models. Many other examples of

different empirical formulas can be found in the literature, such as Halgren's 14-7 buffered potential,¹⁰ featured in the Merck Molecular Force Field (MMFF).¹¹⁻¹⁵ A standard issue associated with force fields is the lack of transferability of optimized parameters between models. In contrast to *ab initio* calculations, where the option exists to combine different levels of theory for different molecules and/or for parts of a same molecule(s) of a quantum mechanical setup,¹⁶⁻²² such flexibility is not always feasible for force fields. Furthermore, unless readily available, the force field parameterization for a set of molecules must be fitted against reliable experimental data and/or accurate *ab initio* calculations. This task is often very time-consuming and can be discouraging for scientists with limited experience in parameterizing molecular force field models.

Efforts have been made in the literature to connect quantum chemical calculations and force field expressions on a more fundamental level. For example, Bond Order Potentials (BOPs) can be viewed as expressions closely related to the realm of density functional (or wave function based) calculations.^{23,24} In addition to providing physical clarity, these approaches avoid a proliferation of fitting parameters. However, despite their appealing nature, these approaches are not widely used, except for the Tight Binding (TB) method, which essentially also falls in this category. Much more widely used at present are machine learning approaches. These methods typically assume that the total energy of a system is the sum of the energy of all atoms in their respective environments. As a first step, the Cartesian coordinates of this environment are converted into symmetry-adapted descriptors. In the two most popular approaches, these descriptors are inputs to a neural network²⁵ or to a Gaussian approximation potential.²⁶ The parameters of these models are obtained by fitting to the results of Density Functional Theory (DFT) calculations. The main advantage of the machine-learning approach is the flexibility of the potential energy expression and its learning-on-the-fly potential.

Another strategy for constructing force fields that has gained popularity involves approximating electronic densities inside integral expressions that resemble Coulomb interaction integrals. Such an approach offers several advantages. It overcomes the problems arising from interaction energy formulas directly based on internuclear distances by instead using electron density distributions, which are more physical in nature. Such an approach also exploits the experience gained from density functional calculations over the years, and, consequently, requires fewer fitting parameters. This kind of approach often relies on Gaussian and/or Slater types of basis-functions to approximate the electron clouds. Exact expressions for these integrals exist when such basis functions are used. Furthermore, compared to machine learned potentials, empirical intuition can be built in more easily since their resulting expressions strongly resemble the solution(s) of simple classical and/or quantum mechanical problems. Notable examples include the Gaussian Electrostatic Model (GEM),²⁷⁻²⁹ the Monomer Electron Density Force Field (MEDFF),³⁰ and our own force field model,³¹ which we intend to improve here.

One unique aspect of our model is how it calculates interaction energies within a system. We achieve this by repeatedly applying the Laplace operator to the non-interacting electron densities of the system's molecules. This process results in a finite sum of fictitious non-interacting electron densities, for which exact expressions can be derived. In the same fashion as the orbitals

calculated with Kohn-Sham DFT (KS-DFT), these electron densities are described as fictitious as they solely are convenient mathematical abstractions that lack a rigorous physical interpretation. Using Hohenberg and Kohn's theorem,³² it is demonstrated that for non-degenerate cases, mapping the non-interacting electron densities of the molecules to the energy of the interacting system is possible. This argument extends to systems with degenerate ground states in their non-interacting components through Levy's constrained search formulation.³³ The main advantage of this approach is that besides requiring the usual molecular specification needed by any, full atom, force field model, e.g., the Z-matrix, it only needs partial atomic charges, e.g., Mulliken or Löwdin charges, of every molecule in the system (when isolated) once a set of exchange and correlation coefficients are fitted against experimental or accurate *ab initio* reference data. However, in our previous paper, these partial charges were treated as constants, which, therefore, made the previous version of our model non-polarizable, and potentially too inaccurate for some systems.

In Sec. II, we highlight three main limitations in our previous model. Those limitations are mostly a consequence of the non-polarizability of the model, and we provide theoretical means to overcome them. In particular, as a theoretical foundation in this improved version of our model, we use a slightly modified, yet equivalent, version of a proof that deals with quantum mechanical systems with degenerate ground states³⁴ and was previously derived by Katriel *et al.* We favor our version of this proof as it is consistent with the style of notation adopted in the theoretical derivation of our previous model. Here in Sec. II, we also provide a detailed description of a minimization problem that allows us to calculate atomic contributions of the atomic electron density contributions that take their surroundings into account. This minimization problem, like similar others found in the literature,^{35,36} is interpreted as a self-energy minimization of the electron clouds. Moreover, like many of those other methods, the minimization problem we propose simply involves a quadratic function, which translates to having to solve a system of linear equations, therefore eliminating any convergence related issues. Conservation of charge is guaranteed with a Lagrange multiplier.

In Sec. III describes the systems used as reference and validation datasets and the software used to calculate both datasets. We also describe all the other well-established methods and the respective software that we use to benchmark the accuracy of our model.

In Sec. IV presents preliminary findings for both calibration and validation datasets, employing our model with an order ten approximation alongside several other established force field models.

Finally, in Sec. V, we summarize our work, add final remarks, and mention potential avenues for future developments of our model.

II. THEORY AND PROBLEM FORMULATION

This section is divided into three parts. Each of these parts focuses on improving different limitations of our previous force field model, limitations that, despite the model's merits, remained either unresolved or incompletely answered. In summary, in the present work, we target the following weaknesses of our previous model:

1. Its theoretical foundation could only establish unique maps between non-interacting and interacting electron densities if all systems (interacting and non-interacting) are non-degenerate.
2. All the contributions for the fictitious electron density inside the integrals of the exchange and correlation functionals were limited to atom-centered functions that have spherical symmetry.
3. The net atomic contributions to the electron densities were scaled to match the Mulliken charges of the isolated monomers regardless of their surroundings in the system (hence non-polarizable).

In the continuation of this section, we will address, point-by-point, each of those weaknesses by providing explicit formulae and/or proofs wherever needed and/or applicable. Thus, subsection A below refers to point 1 above, B to point 2, and C to point 3.

A. Hohenberg-Kohn theorem and degenerate ground states

Using the Hohenberg-Kohn theorem, one can prove that there exists a map between the non-interacting electron densities of two molecules to the energy of the interacting system. With this knowledge, in our prior model, we aimed to, empirically, find an expression for this map in the form of a sum of fictitious Coulomb-like integral expressions. However, this argument fails in justifying the existence of a map between the non-interacting electron densities and the interacting electron density due to ambiguity issues should the interacting system allow degenerate ground states. Such kind of map is useful to serve as a theoretical foundation of a polarizable model, that is, a model whose interacting electron density, fictitious or not, depends on the surroundings of the molecules of a system. In this subsection, we provide theoretical means to overcome this issue.

Let us begin by considering a molecule with N electrons in its neutral state. The Hamiltonian operator governing this molecular is denoted as $\hat{H} : L^2(\mathbb{R}^{3N}) \rightarrow L^2(\mathbb{R}^{3N})$, which we assume to be self-adjoint. The ground state energy, represented by the smallest eigenvalue of \hat{H} , is denoted as $\varepsilon_0 \in \mathbb{R}$. The degeneracy of the ground state is denoted as K , and let the wavefunctions $\psi_1, \dots, \psi_K \in L^2(\mathbb{R}^{3N})$ collectively form an orthonormal basis for the lowest energy eigenspace of \hat{H} . We define the eigendensity, denoted as $\rho_{\varepsilon_0} \in L^1(\mathbb{R}^3)$, for this eigenspace as the following integral:

$$\rho_{\varepsilon_0}(r) = \int_{\mathbb{R}^3} \dots \int_{\mathbb{R}^3} \sum_{k=1}^K \frac{|\psi_k(r, r_2, \dots, r_N)|^2}{K/N} dr_2 \dots dr_N. \quad (1)$$

From Lemma 1, we establish that for every eigenspace of any Hamiltonian operator, there is a unique eigendensity associated with it. This eigendensity can be easily computed when an orthonormal basis of the corresponding eigenspace is known. Consequently, the existence of the map $f := \hat{H} \mapsto \rho_{\varepsilon_0}$ is affirmed. In addition, the reciprocal mappings of any Hamiltonian operator to its external potential and vice versa are facilitated by the following functions:

$$g(\hat{H}) = \hat{H} - \hat{T} - \hat{U}_{EE}, \quad (2a)$$

$$h(\hat{U}_{Ext}) = \hat{T} + \hat{U}_{EE} + \hat{U}_{Ext}, \quad (2b)$$

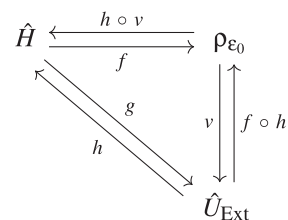


FIG. 1. Commutative diagram depicting the maps we have shown to exist of any molecular system between its Hamiltonian, its external potential, and its ground state eigendensity.

and finally, from Theorem 1, we know that the map $v := \rho_{\varepsilon_0} \mapsto \hat{U}_{Ext}$ exists. All maps just described are summarized in the diagram shown in Fig. 1.

We assume a pair of molecules A and B, with respective ground state energies ε_0^A and ε_0^B and respective non-interacting ground state eigendensities $\rho_{\varepsilon_0^A}$ and $\rho_{\varepsilon_0^B}$. Each eigendensity can be mapped to its respective non-interacting Hamiltonian (\hat{H}_A and \hat{H}_B), and, therefore, the map

$$\rho_{\varepsilon_0^A} \otimes \rho_{\varepsilon_0^B} \mapsto \hat{H}_A \otimes \hat{H}_B \quad (3)$$

must also exist. Since the Hamiltonian of each molecule can be mapped to the position and atomic number of all atoms, the non-interacting Hamiltonians can be mapped to the corresponding interacting Hamiltonian, that is, the map

$$\rho_{\varepsilon_0^A} \otimes \rho_{\varepsilon_0^B} \mapsto \hat{H}_{AB} \quad (4)$$

must exist; therefore, the non-interacting ground state eigendensities of molecules A and B can be mapped to the eigendensity of the interacting system; in other words, the map

$$\rho_{\varepsilon_0^A} \otimes \rho_{\varepsilon_0^B} \mapsto \rho_{\varepsilon_0^{AB}} \quad (5)$$

must also exist. This argument can be used on a per atom basis, which serves to justify separating the entire interaction energy on a per atom pair sum of contributions. Mapping the non-interacting ground state eigendensities of each atom in our system to the ground state eigendensity of the whole is advantageous since it allows us to exploit the spherical symmetry of their ground state eigendensities. However, just like when Hohenberg and Kohn's original proof is used instead, this generalization only holds for the ground state eigendensities of both interacting and non-interacting systems.

B. Atom-centered non-spherical electron density contributions

Going back to our prior study, we introduced a straightforward model that, through the adjustment of a finite set of parameters, aspired to serve as a universal model for the interaction energies between two interacting molecules. Although those aspirations were not exactly realized, our previous model exhibited promising results,

particularly for neutral systems.³¹ However, it lacks any atom-wise contributions in the fictitious electron densities, refer to Eqs. (8a) and (8b), with no spherical symmetry, e.g., dipole-like contributions. The interaction energy between two molecules, A and B, as defined by our simple model, is given by

$$\Delta E_{AB} = \Delta E_{AB}^{EE} + \Delta E_{AB}^{EN} + \Delta E_{AB}^{NN} + \sum_{k=0}^K \left(a_k \text{XC}_{AB}^{EE(k)} + b_k \text{XC}_{AB}^{EN(k)} \right), \quad (6)$$

in which the naïve model contributions are

$$\Delta E_{AB}^{EE} = \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \frac{\rho_{\epsilon_0^A}(r) \rho_{\epsilon_0^B}(r')}{\|r - r'\|} dr' dr, \quad (7a)$$

$$\Delta E_{AB}^{EN} = - \int_{\mathbb{R}^3} \left(\sum_{m=1}^{M_A} \frac{Z_m^A \rho_{\epsilon_0^B}(r)}{\|r - R_m^A\|} + \sum_{m=1}^{M_B} \frac{Z_m^B \rho_{\epsilon_0^A}(r)}{\|r - R_m^B\|} \right) dr, \quad (7b)$$

$$\Delta E_{AB}^{NN} = \sum_{m=1}^{M_A} \sum_{m'=1}^{M_B} \frac{Z_m^A Z_{m'}^B}{\|R_m^A - R_{m'}^B\|}, \quad (7c)$$

and the exchange and correlation contributions are

$$\text{XC}_{AB}^{EE(k)} = \frac{1}{2} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \left(\frac{\rho_{\epsilon_0^A}(r) \nabla_r^{2k} \rho_{\epsilon_0^B}(r')}{\|r - r'\|} + \frac{\rho_{\epsilon_0^B}(r) \nabla_r^{2k} \rho_{\epsilon_0^A}(r')}{\|r - r'\|} \right) dr' dr, \quad (8a)$$

$$\text{XC}_{AB}^{EN(k)} = \int_{\mathbb{R}^3} \left(\sum_{m=1}^{M_A} \frac{Z_m^A \nabla_r^{2k} \rho_{\epsilon_0^B}(r)}{\|r - R_m^A\|} + \sum_{m=1}^{M_B} \frac{Z_m^B \nabla_r^{2k} \rho_{\epsilon_0^A}(r)}{\|r - R_m^B\|} \right) dr, \quad (8b)$$

where R_m^A and Z_m^A label the coordinates and atomic number of the m th atom of A, R_m^B and Z_m^B label the m th atom of B, and the operator ∇^{2k} is shorthand notation for k consecutive applications of the Laplace operator. Despite its simplicity, in our previous study, we found this model to reproduce water dimer interactions to a fair degree of accuracy, albeit with a rather high approximation order (six or higher).

One can easily see from the exchange and correlation functionals above that the heuristic philosophy driving this model is the introduction of fictitious electron clouds by means of consecutive applications of the Laplace operator on the non-interacting atomic density contributions, thus making each of these contributions spherically symmetric by construction. This is likely a limiting factor in our model, since capturing asymmetries in the interacting densities is likely to be more successfully achieved by the inclusion of atom-centered contributions that have no spherical symmetry in the fictitious interacting densities. An example where this plays a significant role is a water molecule with its lone electron pair, which can clearly not be described by spherical densities residing on the oxygen atom. This may also have contributed to the requirement of having to use rather high order expansions (six or higher) to accurately model water dimer interactions in our previous model.

To address this, we add to our model the following pair of family of functionals of the non-interacting ground state eigendensities of A and B:

$$\text{XC}_{AB}^{\text{ED}(k)} : W^{2k+1,1}(\mathbb{R}^3) \times W^{2k+1,1}(\mathbb{R}^3) \rightarrow \mathbb{R}, \quad (9a)$$

$$\text{XC}_{AB}^{\text{ND}(k)} : W^{2k+1,1}(\mathbb{R}^3) \times W^{2k+1,1}(\mathbb{R}^3) \rightarrow \mathbb{R}, \quad (9b)$$

wherein the electron clouds and nuclei are made to interact with fictitious atom-centered electron density contributions with cylindrical symmetry, which are also conveniently parameterized by the ground state eigendensities of A and B, whose exact expressions we assume to be

$$\text{XC}_{AB}^{\text{ED}(k)} = \frac{1}{2} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \left(\frac{\rho_{\epsilon_0^A}(r) \bar{r}_{AB}}{\|\bar{r}_{AB}\|} \cdot \frac{\nabla_{r'}^{2k+1} \rho_{\epsilon_0^B}(r')}{\|r - r'\|} + \frac{\rho_{\epsilon_0^B}(r) \bar{r}_{BA}}{\|\bar{r}_{BA}\|} \cdot \frac{\nabla_{r'}^{2k+1} \rho_{\epsilon_0^A}(r')}{\|r - r'\|} \right) dr' dr, \quad (10a)$$

$$\text{XC}_{AB}^{\text{ND}(k)} = \int_{\mathbb{R}^3} \left(\sum_{m=1}^{M_A} \frac{Z_m^A \bar{r}_{AB}}{\|\bar{r}_{AB}\|} \cdot \frac{\nabla_r^{2k+1} \rho_{\epsilon_0^B}(r)}{\|r - R_m^A\|} + \sum_{m=1}^{M_B} \frac{Z_m^B \bar{r}_{BA}}{\|\bar{r}_{BA}\|} \cdot \frac{\nabla_r^{2k+1} \rho_{\epsilon_0^A}(r)}{\|r - R_m^B\|} \right) dr, \quad (10b)$$

in which ∇^{2k+1} is shorthand notation for one and k consecutive applications of the gradient and Laplacian operators, respectively, and the distance vectors $\bar{r}_{BA}, \bar{r}_{AB} \in \mathbb{R}^3$ are

$$\bar{r}_{AB} = \int_{\mathbb{R}^3} \frac{r \rho_{\epsilon_0^B}(r)}{N_B} dr - \int_{\mathbb{R}^3} \frac{r \rho_{\epsilon_0^A}(r)}{N_A} dr, \quad (11a)$$

$$\bar{r}_{BA} = \int_{\mathbb{R}^3} \frac{r \rho_{\epsilon_0^A}(r)}{N_A} dr - \int_{\mathbb{R}^3} \frac{r \rho_{\epsilon_0^B}(r)}{N_B} dr = -\bar{r}_{AB}. \quad (11b)$$

One can, therefore, see that for the zeroth-order contributions of Eqs. (10a) and (10b), that is, when $k = 0$, the fictitious electron densities obtained through differentiation are atom-centered dipoles, which resemble p-type atomic orbitals, oriented along the distance vectors between molecules A and B. Finally, the interaction energy between A and B in our improved model is

$$\Delta E_{AB} = \Delta E_{AB}^{EE} + \Delta E_{AB}^{EN} + \Delta E_{AB}^{NN} + \sum_{k=0}^K \left(a_k \text{XC}_{AB}^{EE(k)} + b_k \text{XC}_{AB}^{EN(k)} + c_k \text{XC}_{AB}^{\text{ED}(k)} + d_k \text{XC}_{AB}^{\text{ND}(k)} \right), \quad (12)$$

where $a_k, b_k, c_k,$ and d_k denote the multiplying coefficients, $4(K + 1)$ in total, which need to be fitted against reliable *ab initio* and/or experimental data.

C. Modeling Mulliken partial charges

The final limitation in our previous model that we intend to correct is its inability to polarize the molecules in a form that allows for charge redistribution. With this, we mean that the atomic contributions to the total electron density can shift depending on the environment of the atoms. It is noted that our previous model accounts for dipole moments of the molecules to an extent, since the atomic density contributions are multiplied by scaling a scaling factor that ensures that these contributions agree with the Mulliken partial charges obtained from high accuracy *ab initio* calculations. This has the disadvantage of requiring *ab initio* calculations on every molecular species in the system that one aims to model.

In this subsection, we improve our model by lifting this requirement by not only modeling interaction energies, but also the Mulliken charges. Even though Eqs. (10a) and (10b) can account for asymmetries in the interacting ground state eigendensities, they do not allow the net atomic contributions to the density to change, akin to how in our previous version of our model the atomic density contributions are simply scaling factors that match the Mulliken charges of an isolated monomer *ab initio* calculation.

Let us begin by considering a simple system with two molecules A and B. We approximate the interacting ground state eigendensities of these molecules as

$$\tilde{\rho}_{e_0^A}(r) = \sum_{m=1}^{M_A} \sum_n \frac{\chi_m^A c_{m,n}^A}{\exp(\lambda_{m,n}^A \|r - R_m^A\|^2)} \left(\frac{\lambda_{m,n}^A}{\pi} \right)^{3/2}, \quad (13a)$$

$$\tilde{\rho}_{e_0^B}(r) = \sum_{m=1}^{M_B} \sum_n \frac{\chi_m^B c_{m,n}^B}{\exp(\lambda_{m,n}^B \|r - R_m^B\|^2)} \left(\frac{\lambda_{m,n}^B}{\pi} \right)^{3/2}, \quad (13b)$$

where $\chi_m^A, \chi_m^B \in \mathbb{R}$ serve as labels for unknown scalar variables that allow the net atomic contributions to the eigendensity of the molecules to match the ideal *ab initio* calculated Mulliken partial charges, whereas the constants labeled by $c_{m,n}^A, c_{m,n}^B, \lambda_{m,n}^A$, and $\lambda_{m,n}^B$ are taken from our previously tabulated atom-specific values.³¹ A new set of atom-specific coefficients, for the full electron variant of our model, is calculated and presented in Table S1 of the [supplementary material](#), for all elements in the first three rows of the Periodic Table. This new set of coefficients are optimized so that the last three Gaussian contributions are the same as their Effective Core Potential (ECP) counterparts, thus allowing us to separate between core and valence electrons. We obtain the polarization variables from the following minimization problem:

$$\begin{cases} \text{minimize} & U_{AB}^{\text{Pol}}(\chi_1^A, \dots, \chi_{M_A}^A, \chi_1^B, \dots, \chi_{M_B}^B) \\ \text{subject to} & \sum_{m=1}^{M_A} \sum_n \chi_m^A c_{m,n}^A + \sum_{m=1}^{M_B} \sum_n \chi_m^B c_{m,n}^B = N_A + N_B, \end{cases} \quad (14)$$

where N_A and N_B are the number of electrons of molecules A and B and $U_{AB}^{\text{Pol}}: \mathbb{R}^{M_A+M_B} \rightarrow \mathbb{R}$ is a function, that is yet to be defined, of the polarization scalars associated with every atom in the system, which we treat as optimization variables. Naturally, the equality constraint in Eq. (14) is imposed by using a Lagrange multiplier. For anionic systems, $N_A + N_B$, in the equality constraint of Eq. (14), is replaced with $N_A + N_B + 1$. In addition, for the full electron variant of our model, only the valence electron contributions are allowed to

change, that is, the inner core shell electron density contributions of the atoms of each molecule are treated as frozen electron shells. Mathematically, this translates to their, the core shell contributions, associated polarization scalars being constrained to always being equal to one regardless of what surrounds them; hence, we do not treat these as optimization variables but rather as mere constants.

Contrary to our previous model, where computing all intramolecular interactions can be avoided since they cancel out when calculating interaction energies, this is no longer the case if we allow the molecules in our system to become polarized. This is a direct consequence of the intramolecular interaction energy components being dependent on what surrounds every molecule in our system. Therefore, the energies of each molecule when isolated need to be calculated, only once, and then subtracted from the total energy, that is, intra- and intermolecular interaction energies where the latter is disregarded in our previous model due to it canceling as a consequence of its, inaccurate, non-polarizability.

While it can be proved that there exists a functional that maps the non-interacting neutral ground state eigendensity of all the atoms in a polyatomic system to its interacting, not necessarily neutral, ground state eigendensity, which can be done using the modified Hohenberg–Kohn theorem, it does not provide an explicit formula for the map. As an ansatz, we model U_{AB}^{Pol} as being a functional whose mathematical expression is identical to our exchange and correlation corrected interaction energy [Eq. (12)], and whose polarization coefficients minimize it, akin to a self-energy minimization. However, the physical interpretation of this functional is dubious; thus, a set of different exchange and correlation coefficients are used to model the Mulliken charges. This separate set of exchange and correlation coefficients are obtained from least squares minimization against the Mulliken charges of the same dataset used to calibrate the interaction energy model.

This method extends beyond simply modeling the Mulliken charges of diatomic systems (Fig. 2), which is the typical focus in studies using molecular dynamics simulations. These simulations often involve thousands of molecules.

III. METHODS

The current, improved version of our model is calibrated with the same reference data that were used to calibrate the previous non-polarizable version.³¹ In our previous study, the valence electron densities of all the atoms, in the first three rows of the Periodic Table, are fitted, and tabulated, via least-squares minimization against the electronic density obtained from Density Functional Theory (DFT) calculations using the B3LYP³⁷ hybrid functional with Grimme's empirical dispersion formula GD3.³⁸ The Effective Core Potential (ECP) basis-set CEP-31G^{39–41} were used for all atoms except hydrogen and helium where the full-electron basis-set cc-pVDZ^{42–45} is used instead. The water dimer interaction calculations used as reference data consist of a 25×200 uniform grid of calculations in the $O \cdots O$ distance coordinate and a 25×200 uniform grid of calculations in the $H \cdots O$ distance coordinate. This means that the interatomic distances are sampled with 200 points, whereas the angular component at which one of the molecules of water is rotated in an axis perpendicular to the plane of the molecule is sampled with 25 points, using CCSD(T)^{46–50} level of theory with the basis-set cc-pVTZ.⁵¹

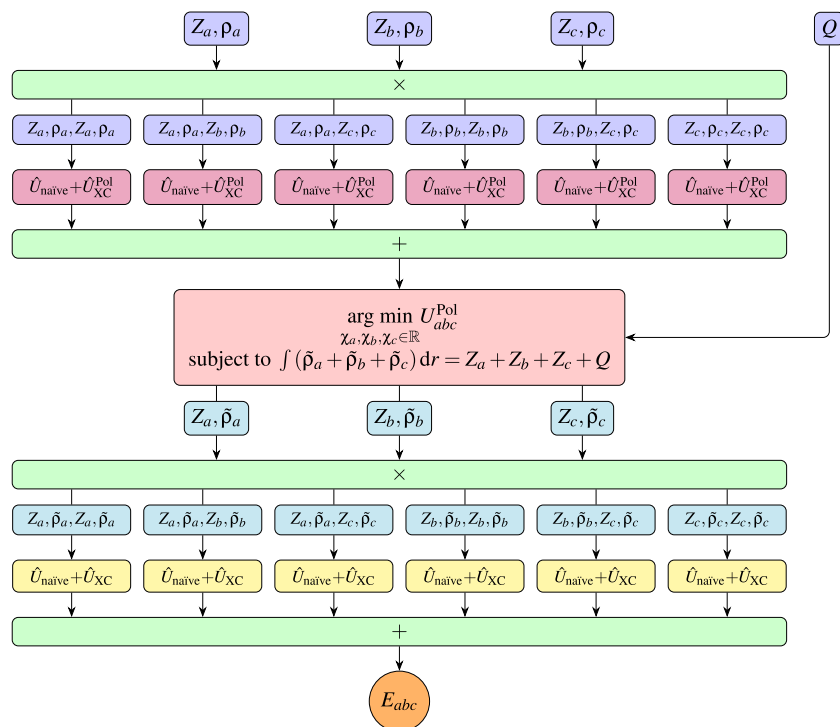


FIG. 2. Diagram illustrating the energy calculation of a hypothetical system comprising three atoms, denoted as a , b , and c , whose respective non-interacting neutral atomic ground state eigendensities are ρ_a , ρ_b , and ρ_c . Contrasting these, the atomic contributions to the modeled interacting eigendensity for the system are represented as $\tilde{\rho}_a$, $\tilde{\rho}_b$, and $\tilde{\rho}_c$. The overall charge of the system is denoted by Q , where for a neutral system, $Q = 0$. See text for explanation of the labels.

To validate our model, comparisons between different energies predicted by our model and accurate *ab initio* calculations are made on a set of different complexes with small molecules, namely H_2 , N_2 , O_2 , and CO_2 . These comparisons include the binding energies between two atoms, for the diatomic molecules mentioned above. To determine its ability to predict covalent bonds. Comparisons between Vertical Electron Affinities (VEAs) of four small molecules and interaction energies between dimer complexes of two identical molecules in their neutral and anion states are also provided. The comparisons of the dimer complexes consist in displacing one of the molecules by a distance $\Delta L \in [1, 5]$, in Å, and rotating it by an angle $\theta \in \{0, \pi/4, \pi/2\}$.

The bond lengths and angles used for all the small molecules considered for this study, for both the calibration and validation, are summarized in Table I.

All the calculations used to validate our model are obtained from ROCCSD(T) and CCSD(T) calculations (for non-singlet and singlet states, respectively), which use the basis-set cc-pVTZ. Moreover, we also compare the results in the validation dataset with the predicted values from the already established force field models: General AMBER Force Field (GAFF),⁵² Merck Molecular Force Field (MMFF94S),^{11–15} both of them using the open-source software Open Babel,⁵³ Effective Fragment Potential (EFP)^{54–56} method using the open-source software LIBEFP,⁵⁷ and MB-nrg^{58,59} using the open-source software MBX.⁶⁰ All *ab initio* calculations that are used and presented for this study are obtained from the Gaussian

TABLE I. Bond lengths and angles used for all the small molecules studied in this article. This list of molecules includes all the molecules in the calibration and validation datasets.

Molecule	Type	Length (Å)	Angle (°)
H_2O	O–H	0.983	108.9
	H–O–H		
CO_2	C–O	1.163	180.0
	O–C–O		
H_2	H–H	0.74	
N_2	N–N	1.09	
O_2	O–O	1.16	

16 software⁶¹ using the LEO High Performance Computing (HPC) facilities in the University of Innsbruck.

The fitting of the exchange and correlation coefficients used to polarize the molecules as described in Eq. (14) is performed with the following combination of methods: An initial optimization using the Nelder–Mead simplex algorithm^{62–64} using only 220 randomly picked water dimer configurations of our calibration dataset, followed by another minimization using a low memory Broyden–Fletcher–Goldfarb–Shanno (LBFGS)^{65–71} algorithm, followed by a final minimization using the same LBFGS algorithm against all water dimer configurations in our calibration dataset.

Exact values for the gradients for the minimization problem to calibrate, and just the calibration, the XC coefficients for the polarization model are obtained using automatic (forward mode) differentiation in Julia using the open-source package ForwardDiff.⁷² For the full electron variant of our method, only the Nelder–Mead algorithm is employed for the fitting of its parameters, due to numerical noise issues stemming from huge numbers that arise when using this variant of our molecular force field model.

Unlike our non-polarizable model, where the XC coefficients are obtained through an unconstrained least squares minimization problem, in this study, we constrain them to follow exact alternating expressions to avoid both overfitting and unphysical oscillations when extrapolating to shorter interatomic distances than those existing in our calibration dataset. The exact expressions used for every XC coefficient, that is not zeroth order, are

$$a_k = (-1)^k \frac{A}{(2k)!}, \quad (15a)$$

$$b_k = (-1)^k \frac{B}{(2k)!}, \quad (15b)$$

$$c_k = (-1)^k \frac{C}{(1+2k)!}, \quad (15c)$$

$$d_k = (-1)^k \frac{D}{(1+2k)!}, \quad (15d)$$

where $A, B, C, D \in \mathbb{R}$ are optimized parameters (distinct for the polarization and interaction energy models, and, hence, a total of eight fitting parameters are used regardless of the order of the model) $\forall n \in \mathbb{N}^+ \setminus \{0\}$. All zeroth-order coefficients are set to zero. Equations (15a)–(15d) are chosen due to these bearing a

strong resemblance the coefficients previously obtained in a numerical experiment, whose values are not presented in this study, in which we perform least squares minimization against our calibration dataset with every coefficient being treated as an unconstrained optimization variable.

IV. RESULTS AND DISCUSSION

Allowing any of the zeroth-order XC coefficients a_0 , b_0 , c_0 , and d_0 [refer to Eq. (12)] to be different than zero significantly compromises the accuracy of our model, especially for low-energy interaction configurations. However, setting these zeroth-order coefficients to zero ensures that our model closely resembles the naïve model for large interatomic distances. This is a desirable feature since, at considerable distances between interacting molecules, the naïve model approximates the exact answer due to the negligible correlation between the electrons of each molecule. Moreover, by constraining the other XC coefficients to exact expression as shown in Eq. (15a) through Eq. (15d), and then minimizing the coefficients of these expressions with respect to our reference data, rather than having each XC coefficients as a minimization variable, eliminates undesired oscillations at short interatomic distances.

Despite this, the ECP variant of our model noticeably underestimates the interaction energies, compared to accurate *ab initio* calculations of a water dimer in the $\text{H} \cdots \text{O}$ distance coordinate shown in the left panels of Fig. 3. This is likely due to the core electron contributions of the oxygen atom to the electron density not being modeled. In contrast, the full electron variant of our model accurately predicts these interaction energies; however, it displays discernible numerical noise, despite 64-bit floating-point precision being used throughout our entire code. Similar numerical instability

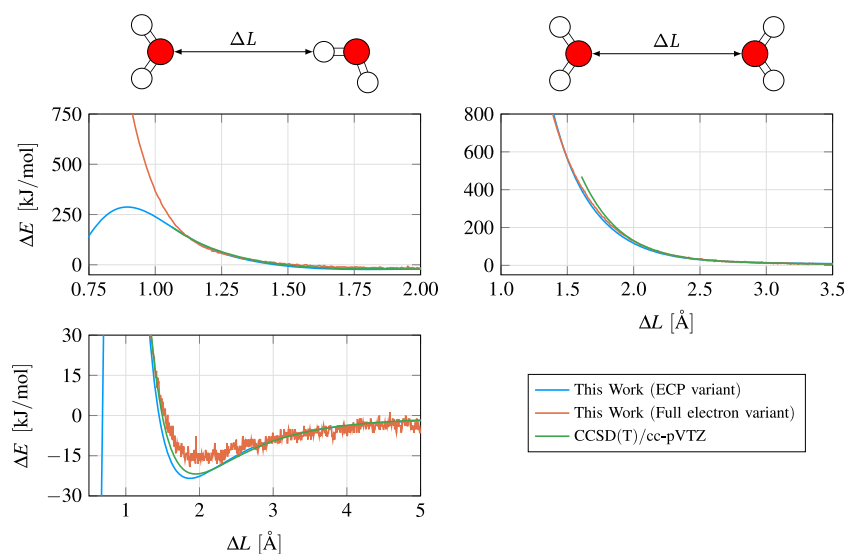


FIG. 3. Comparison between the interaction energies of full electron and ECP-fitted models, both at order ten, and the reference *ab initio* energies for a water dimer in hydrogen bonding ($\text{H} \cdots \text{O}$; left panels) and oxygen facing ($\text{O} \cdots \text{O}$; right panel) conformations. In the left panels, ΔL is the distance between O and H atoms, and in the right panel, ΔL is the distance between the O atoms.

issues have also been reported in other studies that use comparable polarizable force field models.³⁶ This numerical noise is due to the high-order powers of the extinction/decay coefficients of the Gaussian s-type functions, which are up to 20 for an order ten approximation.

A close inspection of the water dimer displaced along the O··O distance coordinate shown in the right panel of Fig. 3 shows that both the full electron and ECP variants of our model reproduce its respective interaction energies with a fair degree of accuracy. Neither of the two variants of our model display unphysical oscillations through this distance coordinate. Like the H··O distance coordinate, the full electron variant of our model displays the same numerical noise in the O··O distance coordinate. Interestingly, for interatomic distances between 1.5 and 2.0 Å, compared to the reference *ab initio* reference data, our full electron fitted model underestimates energies. Nonetheless, for interatomic distances shorter than those illustrated in Fig. 3, the accuracy of our ECP fitted model may diminish. This discrepancy arises from its inaccurate depiction of inner shell electrons, rendering it unsuited for extremely high-energy molecular dynamics simulations where such configurations might be more prevalent. To enhance the accuracy of the full electron variant of our model, for high-energy interactions, the incorporation of similar polarization coefficients into the core shell contributions is needed, that is, no longer treating the inner shell contributions as frozen electron clouds. Incorporating these inner shell coefficients would require significant changes to the fitting of the parameters used to model partial charges since relying solely on Mulliken charges as reference would be insufficient. Instead, this would require the consideration of shell-specific contributions to the electron density for each configuration within our calibration dataset. This more nuanced approach is beyond the scope

of this study. Moreover, the inclusion of post-Born–Oppenheimer corrections for high-energy configurations is another factor to consider.

We analyzed the atomic partial charges by comparing the ones obtained from our model and the reference *ab initio* dataset. Looking at the hydrogen bonded dimer configuration (left panels of Fig. 4), a noticeable decline in accuracy is always observed when the H··O distances become shorter. All errors of the partial charges in the all-electron variant are smaller than 0.15 e and only the error of hydrogen bond acceptor oxygen number 4 is larger than 0.1 e , where e is the elementary charge. The partial charge errors for the ECP variant are even much smaller, for every atom less than 0.07 e . The right panels of Fig. 4 provide an identical examination for the “repulsive” water dimer conformation. Notwithstanding the completely different water–water orientation with the oxygen atoms facing each other, the same conclusion concerning the superiority of the ECP basis set variant and the overall accuracy is drawn.

In a broader analysis of the predicted partial charges from both our ECP and full electron fitted variants of our model, compared to their respective Mulliken charges from the SCF electron densities computed at CCSD(T)/CEP-31G and CCSD(T)/cc-pVTZ levels of theory, a consistent trend emerges. Both versions of our fitted model showcase acceptable accuracy in estimating Mulliken charges, particularly for configurations with lower energy (see Fig. 5). Notably, the ECP variant slightly outperforms in terms of accuracy, boasting a slightly larger coefficient of determination against its reference *ab initio* data. This marginal improvement in accuracy is likely also influenced by the inherent numerical error noise introduced due to the use of limited floating-point precision representation and the necessity to handle exponentiation of relatively large numbers, as evident in the decaying coefficients of the Gaussian basis functions

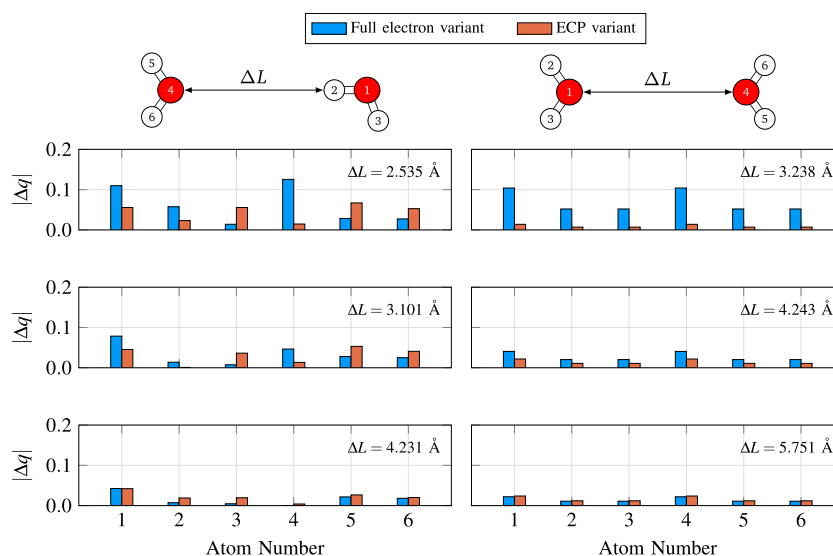


FIG. 4. Comparison of the errors in the partial charges from the two variants of our model (full-electron and ECP) for different water dimer geometries. The bars indicate the absolute value of the differences with respect to Mulliken charges from *ab initio* calculations on the CCSD(T)/cc-pVTZ and CCSD(T)/CEP-31G levels of theory, respectively. The two water dimer configurations used and the numbering of the atoms are shown in the sketches on top of the error bar plots.

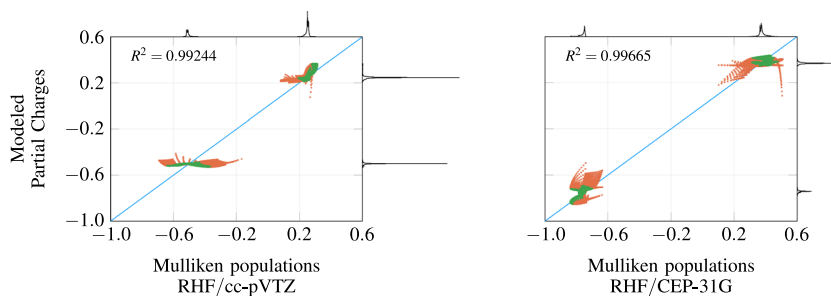


FIG. 5. Modeled atomic partial charges plotted against reference Mulliken charges, distinguished by the interaction energy of the respective geometry. Left panel: All-electron variant; right panel: ECP variant. The orange and green data points refer to charges from configurations with interaction energies above or below +50 kJ/mol, respectively. For clarity, the distribution of the data points along each axis is plotted above and to the right of the 2D plot panels.

used to approximate atomic electron contributions (refer to Table S1 of the [supplementary material](#)), namely for fitted full electron fitted densities.

While the accuracy of our results is qualitatively comparable to findings reported using other similar and well-established methods found in the literature, such as Atom-Bond Electronegativity Equalization Method (ABEEM)⁷³ and Electronegativity Equalization Method (EEM),⁷⁴ it is noted that these results are confined to our calibration dataset. As such, validation against configurations outside this dataset is crucial to affirm the broader applicability of our model. Given the numerical instabilities inherent in our full electron model, arising from the challenges of handling exponentiation of large numbers, our validation is solely directed toward the ECP variant of our model. Specifically, we validate both the interaction energy and polarization models against dimer configurations of molecules that are small enough to allow CCSD(T)/CEP-31G level of theory to be used.

A. Limit toward an infinite order expansion

Expressing our exchange and correlation contributions with coefficients following the exact expressions in Eq. (15a) through Eq. (15d) establishes the legitimacy of treating this expansion as a series. This is evident in Fig. 6, where a normalized sum of electron–nuclei contributions is expanded up to orders 10 and 100 for a hypothetical system comprising an atom with an atomic number of one.

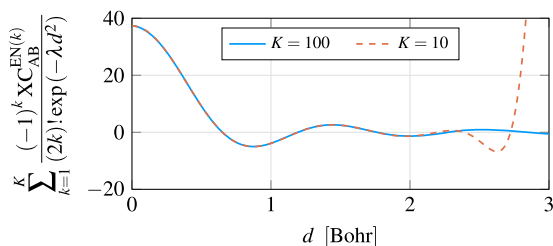


FIG. 6. Comparison between normalized sums of electron–nuclei exchange and correlation functionals with 10 and 100 contributions of a fictitious system of just one atom and an electron cloud modeled by a single s-type Gaussian basis-function with unit amplitude and decay coefficient $\lambda = 3$.

In this fictitious scenario, a single electron cloud, characterized by an amplitude equal to one and with decay coefficient $\lambda = 3$, its center is separated from the atom by a distance d . Figure 6 illustrates the convergence of the series, demonstrating that as the order increases, the sum converges to a defined limit. While we do not provide exact, nor approximate, expressions for the limit when $K \rightarrow \infty$ in this study, this topic still warrants to be mentioned given that obtaining such an expression would contribute not only to reduce the overall computational cost, but also to alleviate numerical noise, like it is observed in Fig. 3. Floating-point precision errors are prevented in Fig. 6 by obtaining exact values using the open-source software Maxima,⁷⁵ using its default arbitrary-precision arithmetic, and only casting the result of each sum to a floating-point precision representation after a factorization of the exact result.

B. Validation against neutral complexes

To validate the order ten ECP variant of our model, for high interaction energies, a set of different dimer complexes of two identical molecules, none of them being part of the calibration dataset, is analyzed (Fig. 7). Specifically, the complexes $\text{H}_2 + \text{H}_2$, $\text{N}_2 + \text{N}_2$, $\text{O}_2 + \text{O}_2$, and $\text{CO}_2 + \text{CO}_2$ are considered, due to these being small enough to allow high accuracy *ab initio* calculations to be made, and since the molecules that are part of these complexes are of considerable interest in various other scientific fields, such as biology, atmospheric sciences, and materials science. Refer to Fig. S2 through Fig. S5 of the [supplementary material](#) for diagrams illustrating the geometric configuration of the dimer complexes used in Fig. 7. For the *ab initio* calculations of the interacting dimer complexes, all the systems are assumed to be in their singlet states, except for all the $\text{O}_2 + \text{O}_2$ complexes where the systems are assumed to be quintets. While these results are not shown in this article, identical results are obtained for the same geometric configurations of $\text{O}_2 + \text{O}_2$ complexes in their triplet states if the interatomic separation between the dimers is kept at not too long distances. For larger separations, the *ab initio* calculations fail to converge with equal or less than 200 SCF iterations, possibly due to basis-set limitations.

The ECP order ten variant of our model replicated the *ab initio* calculated energies with a fair degree of accuracy, except for the $\text{N}_2 + \text{N}_2$ complexes. For the $\text{H}_2 + \text{H}_2$ complexes, the EFP results displayed higher accuracy compared to our model, even at higher

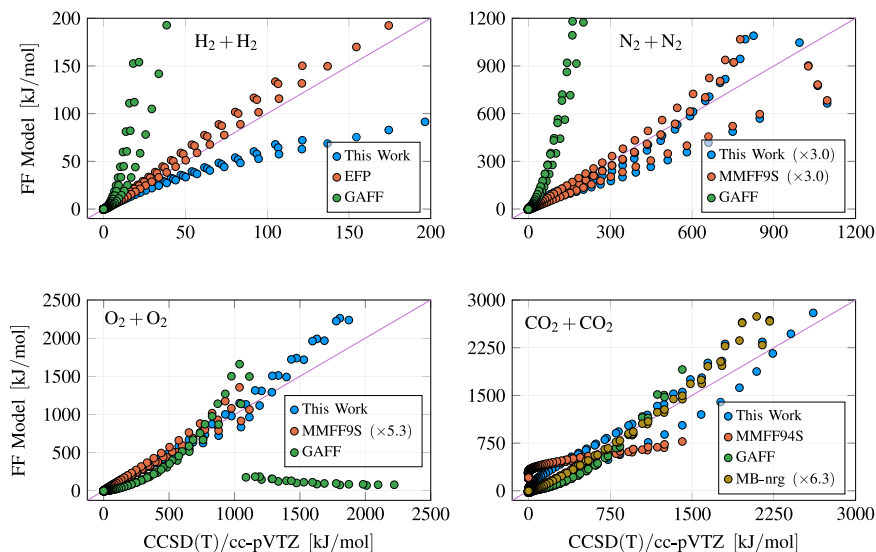


FIG. 7. Comparison between the interaction energies of various neutral complexes obtained with an order ten ECP variant of our model against its ideal *ab initio* results. For reference, comparisons with other established molecule force field models often found in the literature are also included. None of the dimer configurations shown in these comparisons are part of the calibration dataset.

energy regimes. Nevertheless, the ECP variant of our model displays a fair degree of accuracy for this complex for energies roughly up to 50 kJ/mol. For the N₂ + N₂ complexes, both our model and MMFF94S displayed similar levels of accuracy, both differing to the ideal *ab initio* result by a factor of ~ 3.0 (Fig. 7). For the O₂ + O₂ and CO₂ + CO₂ complexes, our model displayed high level accuracy outperforming all the other molecular force fields considered in the high energy regions. Notably, for the CO₂ + CO₂ complexes, MB-nrg displayed a very close to linear relation against the *ab initio* calculated data; however, a constant multiplicative factor of roughly 6.3 between the *ab initio* and MB-nrg calculated results is observed as shown in the bottom right panel of Fig. 7. This factor is in agreement with the scaling factor of 0.15 for the two-body energies reported to be necessary for accurately simulating liquid and vapor phases.⁵⁹ All the comparisons considered against other established force field models are constrained to instances wherein the molecules of the dimer complexes are encompassed within their default fragment libraries, and/or those cases where the models do not fail to run.

While our model demonstrates commendable accuracy in estimating interaction energies for high-energy neutral dimer systems, its performance diminishes notably for lower energies (2 kJ/mol or less), as illustrated in Fig. S2 through Fig. S5 of the [supplementary material](#). In these instances, our model struggles to predict the formation of any weak bonds among neutral gases. Consequently, for simulations involving low-energy scenarios, such as liquid-phase simulations of H₂, N₂, O₂, and/or CO₂, this version of our force field model still proves to be inadequate and is likely to yield inaccurate predictions for such systems.

Despite these improvements, our model, as depicted in Fig. S5 of the [supplementary material](#), still struggles to predict any covalent bonds, similar to its shortcomings in low energy interaction

configurations. Consequently, it proves inadequate for intramolecular geometry optimization. Therefore, to ensure accurate molecular geometries, reliance on either a dependable database or theoretical calculations capable of optimizing molecular geometry is still needed.

Nonetheless, for systems comprising multiple molecules that are treated as rigid objects, conducting geometry optimizations remains feasible. Although explicit calculation or comparison of forces is lacking, in the dimer complexes, our data align with a unit slope in most cases. Thus, no scaling factor for the energies is required and it seems likely that also the forces derived from our model are in good agreement with those obtained at the CCSD(T)/cc-pVTZ level theory, provided that the molecules are treated as rigid objects.

C. Validation against anionic complexes

For the same dimer complexes in their anion states, we find results that somewhat resemble a linear trend between the interaction energies calculated with our order ten ECP variant of our model and the equivalent total energies obtained from ROCCSD(T)/cc-pVTZ level of theory as shown in Fig. 8. However, unlike the interaction energies of Fig. 7, the linear regressions obtained for all systems show large intersects. No clear trend is observed between molecular size nor atomic numbers, and the numerical value of these intersects. Furthermore, the slope of these linear regressions varied roughly between 0.4 and up to 2.0 depending on the system of dimers. As it has already been mentioned, two likely reasons for the discrepancies between our polarizable model and their ideal *ab initio* calculation counterparts are the narrowness of our calibration dataset, and our model's inability to discern between atomic-shell specific contributions to the electron density.

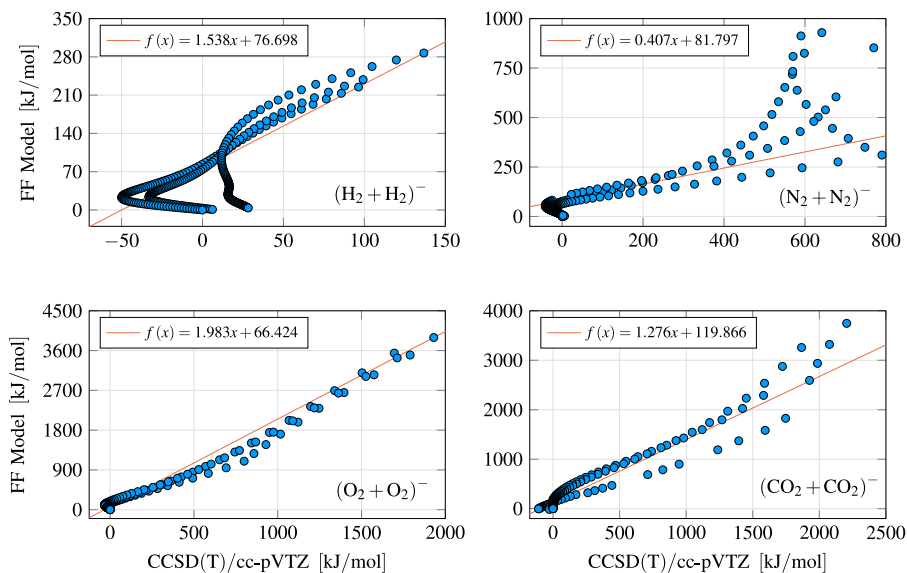


FIG. 8. Comparison between the total energies of various anion complexes obtained with an order ten ECP variant of our model against its ideal *ab initio* results. None of the dimer configurations shown in these comparisons are part of the calibration dataset.

All the anion dimer complexes used in Fig. 8 are assumed to be in their doublet states (one unpaired electron), except for the $(\text{O}_2 + \text{O}_2)^-$ complexes that are assumed to be in their quartet states (three unpaired electrons). This makes sense as one of the two interacting O_2 monomers will be in its triplet state for very large interatomic separations, while the other monomer is, in its ground state, a doublet. While not presented, doublet state calculations for the $(\text{O}_2 + \text{O}_2)^-$ complexes yield similar results, albeit with results deviating slightly more from a linear regression. The geometric configurations used for the anion complexes are the same as the ones used for the neutral complexes. For every panel of Fig. 8, the respective collinear arrangement ($\theta = 0$ for the dimer configurations shown in Fig. S2 through Fig. S5 of the [supplementary material](#)) where the interatomic separation is at its maximum (5 Å) is used as the zeroth energy reference point for the interaction energy calculations.

V. CONCLUSIONS

We present an upgraded version of our electron density-based molecular force field, designed to replicate electrostatic polarization effects. This improvement hinges on minimizing an energy functional derived from the non-interacting ground state eigendensities of all atoms within the system. Notably, this energy functional mirrors the proposed interaction energy, employing distinct optimized parameters fitted to predict the Mulliken charges of a given system. To validate the existence of such a functional, we furnish a proof using a modified version of the Hohenberg–Kohn theorem, establishing a robust theoretical foundation for our polarizable molecular force field.

In this enhanced version, we have implemented several refinements. These include incorporating electron density contributions

devoid of spherical symmetry, acquired from directional derivatives concerning the distance vector between pairs of atoms. In addition, we have allowed the net atomic contributions to electron density to vary based on the molecular surroundings. Analogous to electron–electron and electron–nuclei exchange and correlation contributions, these non-spherical contributions can be analytically resolved. Crucially, by constraining the fitted exchange and correlation coefficients to exact expressions instead of treating them as free parameters, we have effectively eliminated unphysical oscillations in predicted interaction energies in our numerical experiments. This enhancement is evident across both our full electron and ECP cases, particularly when extrapolating to exceedingly small interatomic distances.

Our upgraded model exhibits heightened flexibility and versatility since it can capture polarization effects. The flow of electron density allows us to model atomic partial charges. Parameters for the elements from the first three rows of the Periodic Table are supplied. This already allows for application to a wide range of molecular systems. Furthermore, adapting the model for different ionic states is straightforward, simply involving the adjustment of the equality constraint of a Lagrange multiplier.

Notwithstanding the convenience and broad applicability, we found that the accuracy of the presented model may occasionally deviate from ideal standards. This arises from various factors, including the inherent complexities of molecular interactions and the specificity of the calibration dataset.

Despite significant progress, certain aspects warrant further investigation. For instance, exploring the incorporation of polarization effects to inner core electronic shells could be pivotal, particularly in high-energy molecular dynamics simulations. In addition, efforts could be directed toward deriving exact or approximate expressions when employing infinite expansions to mitigate numer-

ical noise associated with low-energy interactions. Nonetheless, our ECP calibrated variant demonstrates negligible numerical noise for up to a tenth-order truncated series expansion across various dimer systems explored, yielding reasonably accurate predictions for both interaction energies and partial charges in neutral low-energy configurations.

SUPPLEMENTARY MATERIAL

In the [supplementary material](#), a table with fitting parameters for atomic electron densities in the full-electron variant of our model can be found. The valence electron contribution is represented by the last three Gaussian contributions for each element. The mathematical expressions for exchange and correlation functionals are also given, together with the fitted coefficients. In addition, comparisons of the energies predicted by our model with various small molecules are provided.

ACKNOWLEDGMENTS

J.R. acknowledges the support from the Portuguese national funding agency Fundação para a Ciência e a Tecnologia (FCT) for the PhD scholarship Grant No. PD/BD/142846/2018 and together with P.L.V. the research Grant No. CEFITEC UIDB/00068/2020 (DOI: 10.54499/UIDB/00068/2020). This work was also supported by the Radiation Biology and Biophysics Doctoral Training Program (RaBBiT, Grant No. PD/00193/2012; UCIBIO Grant No. UIDB/04378/2020). K.H. acknowledges the support from the Swedish Research Council (Vetenskapsrådet), the Swedish National Infrastructure for Computing (SNIC/NAISS), and the Swedish national strategic e-Science program eSSSENCE. The results presented here have been achieved, in part, using the LEO High Performance Computing (HPC) infrastructure of the University of Innsbruck. This work was partially carried out within the framework of the EUROfusion Consortium and received funding from the Euratom Research and Training programme under Grant Agreement No. 101052200-EUROfusion. The views and opinions expressed herein do not necessarily reflect those of the European Commission.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

José Romero: Investigation (equal); Methodology (equal); Validation (equal); Writing – original draft (equal). **Paulo Limão-Vieira:** Formal analysis (equal); Supervision (equal); Validation (equal); Writing – review & editing (equal). **Thana Maihom:** Formal analysis (equal); Methodology (equal); Validation (equal); Writing – review & editing (equal). **Kersti Hermansson:** Formal analysis (equal); Investigation (equal); Methodology (equal); Writing – original draft (equal); Writing – review & editing (equal). **Michael**

Probst: Investigation (equal); Validation (equal); Writing – review & editing (equal).

DATA AVAILABILITY

All the data and programming scripts used to make this manuscript are freely available under a GPL-3.0 license and can be accessed from the GitHub repository: <https://github.com/JoseRodriguezRomero/PolElectronDensityForceModel>.

A software implementation intended for general use in the C++ programming language of this model for calculating point energies and performing geometry optimization of molecular systems is also freely available under the GPL-3.0 license on GitHub and Zenodo,⁷⁶ along with documentation. The software relies on the following C/C++ libraries: Eigen⁷⁷ for efficient linear algebra operations, libLBFGS⁷⁸ an implementation of J. Nocedal's low-memory BFGS algorithm^{69–71} in the C programming language, used in all geometry optimizations, and autodiff⁷⁹ for forward mode automatic differentiation, guaranteeing the calculation of accurate gradients of the energy.

APPENDIX A: HOHENBERG-KOHN THEOREM AND GROUND STATE ELECTRON EIGENDENSITIES

Lemma 1. The eigendensity of any eigenspace associated with a Hamiltonian operator remains invariant under any rotations applied to the orthonormal basis being used.

Proof. Let the wavefunctions ψ_1, \dots, ψ_K form an orthonormal basis of the eigenspace of \hat{H} with eigenvalue ϵ , and let the wavefunctions ψ'_1, \dots, ψ'_K be a distinct orthonormal basis of the same eigenspace. Since both sets of functions serve as a basis for the same functional space, there exists an irreducible unitary representation (norm-preserving map) $\mathcal{R} : G_\epsilon \rightarrow U(K)$ that maps any element in the group G_ϵ composed of all wavefunctions that are eigenfunctions of \hat{H} , with eigenvalue ϵ , to a $K \times K$ unitary matrix that operates on a K -dimensional Hilbert space $V(K)$, due to \hat{H} being self-adjoint, such that

$$\mathcal{R} \begin{bmatrix} \psi_1 \\ \vdots \\ \psi_K \end{bmatrix} = \begin{bmatrix} \psi'_1 \\ \vdots \\ \psi'_K \end{bmatrix}, \quad (\text{A1})$$

in which taking the norm of both sides of the equation concludes the proof. \square

Lemma 2. An eigenspace of a Hamiltonian operator cannot be a smaller subset of an eigenspace of another distinct (more than just a constant) Hamiltonian.

Proof. Let $\psi_1, \dots, \psi_K \in L^2(\mathbb{R}^{3N})$ be an orthonormal basis of the eigenspace of a N electron Hamiltonian $\hat{H} = \hat{T} + \hat{U}_{\text{EE}} + \hat{U}_{\text{Ext}}$ of eigenvalue ϵ and $\psi_1, \dots, \psi_{K'} \in L^2(\mathbb{R}^{3N})$ an orthonormal basis of the eigenspace of a distinct Hamiltonian $\hat{H}' = \hat{T} + \hat{U}_{\text{EE}} + \hat{U}'_{\text{Ext}}$, in which $K' \leq K$. It follows from integration over any $\Omega \subseteq \mathbb{R}^{3N}$ that it must be true that

$$\begin{cases} \varepsilon \sum_{n=1}^{K'} \langle \psi_n \rangle_{\Omega} = \sum_{n=1}^{K'} \langle \psi_n | \hat{T} + \hat{U}_{EE} + \hat{U}_{Ext} | \psi_n \rangle_{\Omega}, \\ \varepsilon' \sum_{n=1}^{K'} \langle \psi_n \rangle_{\Omega} = \sum_{n=1}^{K'} \langle \psi_n | \hat{T} + \hat{U}_{EE} + \hat{U}'_{Ext} | \psi_n \rangle_{\Omega}, \end{cases} \quad (\text{A2a})$$

$$\Rightarrow (\varepsilon - \varepsilon') \sum_{n=1}^{K'} \langle \psi_n \rangle_{\Omega} = \sum_{n=1}^{K'} \langle \psi_n | \hat{U}_{Ext} - \hat{U}'_{Ext} | \psi_n \rangle_{\Omega}, \quad (\text{A2b})$$

and therefore, $\hat{U}_{Ext} - \hat{U}'_{Ext}$ must be simply a constant, implying that $K' = K$. Furthermore, if we assume the vacuum level to be the zeroth energy level, then $\hat{U}_{Ext} = \hat{U}'_{Ext}$. \square

Theorem 1. *There is a unique functional of the ground state eigendensity of a Hamiltonian that maps it to its external potential.*

Proof. Let \hat{H} and \hat{H}' be two distinct N electron Hamiltonian operators whose respective lowest eigenvalues are ε_0 and ε'_0 . Furthermore, let the degeneracies of their respective ground state eigenspaces be K and K' ; then, their respective ground state eigendensities are

$$\rho_{\varepsilon_0}(r) = \int_{\mathbb{R}^3} \dots \int_{\mathbb{R}^3} \sum_{k=1}^K \frac{|\psi_k(r, r_2, \dots, r_N)|^2}{K/N} dr_2 \dots dr_N, \quad (\text{A3a})$$

$$\rho_{\varepsilon'_0}(r) = \int_{\mathbb{R}^3} \dots \int_{\mathbb{R}^3} \sum_{k=1}^{K'} \frac{|\psi'_k(r, r_2, \dots, r_N)|^2}{K'/N} dr_2 \dots dr_N, \quad (\text{A3b})$$

and thus, by hypothesis,

$$\rho_{\varepsilon_0}(r) = \rho_{\varepsilon'_0}(r) = \rho_0(r) \quad \forall r \in \mathbb{R}^3. \quad (\text{A4})$$

Without loss of generality, we get from Lemma 2 and by integrating that

$$\begin{cases} \varepsilon_0 = \sum_{n=1}^K \sum_{n'=1}^{K'} \frac{\langle \psi_n | \hat{T} + \hat{U}_{EE} + \hat{U}_{Ext} | \psi_n \rangle}{KK'}, \\ \varepsilon'_0 = \sum_{n=1}^K \sum_{n'=1}^{K'} \frac{\langle \psi'_{n'} | \hat{T} + \hat{U}_{EE} + \hat{U}'_{Ext} | \psi'_{n'} \rangle}{KK'}, \end{cases} \quad (\text{A5a})$$

$$\Rightarrow \begin{cases} \varepsilon'_0 < \varepsilon_0 + \sum_{n=1}^K \sum_{n'=1}^{K'} \frac{\langle \psi_n | \hat{T} + \hat{U}_{EE} + \hat{U}'_{Ext} | \psi_n \rangle}{KK'}, \\ \varepsilon_0 < \sum_{n=1}^K \sum_{n'=1}^{K'} \frac{\langle \psi'_{n'} | \hat{T} + \hat{U}_{EE} + \hat{U}_{Ext} | \psi'_{n'} \rangle}{KK'}, \end{cases} \quad (\text{A5b})$$

$$\Rightarrow \begin{cases} \varepsilon'_0 < \varepsilon_0 + \sum_{n=1}^K \sum_{n'=1}^{K'} \frac{\langle \psi_n | \hat{U}'_{Ext} - \hat{U}_{Ext} | \psi_n \rangle}{KK'}, \\ \varepsilon_0 < \varepsilon'_0 + \sum_{n=1}^K \sum_{n'=1}^{K'} \frac{\langle \psi'_{n'} | \hat{U}_{Ext} - \hat{U}'_{Ext} | \psi'_{n'} \rangle}{KK'}, \end{cases} \quad (\text{A5c})$$

$$\Rightarrow \begin{cases} \varepsilon'_0 < \varepsilon_0 + \sum_{n=1}^K \int_{\mathbb{R}^3} \frac{\rho_0(r)(u'(r) - u(r))}{NK'} dr, \\ \varepsilon_0 < \varepsilon'_0 + \sum_{n=1}^K \int_{\mathbb{R}^3} \frac{\rho_0(r)(u(r) - u'(r))}{NK} dr, \end{cases} \quad (\text{A5d})$$

which is a contradiction identical to the one found in Hohenberg and Kohn's original proof.³² \square

TABLE II. Comparison of the Vertical Electron Affinity (VEA) values of several few small molecules obtained from high accuracy *ab initio* calculations, CCSD(T)/cc-pVTZ level of theory, and our order ten full electron and ECP variants of our model. All the values reported are expressed in eV.

	CCSD(T)	ECP variant	Full elec. variant
H ₂	-4.217	-3.676	3.878 × 10 ¹
N ₂	-3.444	2.966	4.772 × 10 ³
O ₂	-1.433	-52.189	6.828 × 10 ⁴
CO ₂	-4.489	43.980	3.304 × 10 ¹

APPENDIX B: COMPARING VEA OF SMALL MOLECULES

As one would expect, neither the full electron nor ECP order ten variants of our model provide accurate estimations of the vertical electron affinities of the various small molecules as summarized in Table II. Specifically, the full electron variant shows a much lesser accuracy, in terms of magnitude, of the two when compared to the equivalent *ab initio* calculated result with discrepancies with two or more orders of magnitude higher. These results come to no surprise given that our model is unable to model chemical bonds (refer to Fig. S6 of the supplementary material). Nevertheless, a detailed exploration and reporting of the strengths and weaknesses of our model is essential for those aiming to conduct accurate MD simulations.

REFERENCES

- W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *J. Chem. Phys.* **79**, 926–935 (1983).
- M. A. González and J. L. F. Abascal, "A flexible model for water based on TIP4P/2005," *J. Chem. Phys.* **135**, 224516 (2011).
- M. M. Ghahremanpour, J. Tirado-Rives, and W. L. Jorgensen, "Refinement of the optimized potentials for liquid simulations force field for thermodynamics and dynamics of liquid alkanes," *J. Phys. Chem. B* **126**, 5896–5907 (2022).
- M. W. Mahoney and W. L. Jorgensen, "A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions," *J. Chem. Phys.* **112**, 8910–8922 (2000).
- W. L. Jorgensen and J. Tirado-Rives, "Potential energy functions for atomic-level simulations of water and organic and biomolecular systems," *Proc. Natl. Acad. Sci. U. S. A.* **102**, 6665–6670 (2005).
- L. S. Dodda, J. Z. Vilseck, J. Tirado-Rives, and W. L. Jorgensen, "1.14*CM1A-LBCC: Localized bond-charge corrected CM1A charges for condensed-phase simulations," *J. Phys. Chem. B* **121**, 3864–3870 (2017).
- L. S. Dodda, I. Cabeza de Vaca, J. Tirado-Rives, and W. L. Jorgensen, "LigParGen web server: An automatic OPLS-AA parameter generator for organic ligands," *Nucleic Acids Res.* **45**, W331–W336 (2017).
- D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods, "The Amber biomolecular simulation programs," *J. Comput. Chem.* **26**, 1668–1688 (2005).
- D. Case, H. Aktulga, K. Belfon, I. Ben-Shalom, J. Berryman, S. Brozell, D. Cerutti, T. Cheatham III, G. C. Andrés Cisneros, V. Cruzeiro, T. Darden, R. Duke, G. Giambasu, M. Gilson, H. Gohlke, A. Goetz, R. Harris, S. Izadi, S. Izmailov, K. Kasavajhala, M. Kaymak, E. King, A. Kovalenko, T. Kurtzman, T. Lee,

- S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K. O'Hearn, A. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, A. Shajan, J. Shen, C. Simmerling, N. Skrynnikov, J. Smith, J. Swails, R. Walker, J. Wang, J. Wang, H. Wei, R. Wolf, X. Wu, Y. Xiong, Y. Xue, D. York, S. Zhao, and P. Kollman, *AMBER* 2022, 2022.
- ¹⁰T. A. Halgren, "The representation of van der Waals (vdW) interactions in molecular mechanics force fields: potential form, combination rules, and vdW parameters," *J. Am. Chem. Soc.* **114**, 7827–7843 (1992).
- ¹¹T. A. Halgren, "Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94," *J. Comput. Chem.* **17**, 490–519 (1996).
- ¹²T. A. Halgren, "Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions," *J. Comput. Chem.* **17**, 520–552 (1996).
- ¹³T. A. Halgren, "Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94," *J. Comput. Chem.* **17**, 553–586 (1996).
- ¹⁴T. A. Halgren and R. B. Nachbar, "Merck molecular force field. IV. Conformational energies and geometries for MMFF94," *J. Comput. Chem.* **17**, 587–615 (1996).
- ¹⁵T. A. Halgren, "Merck molecular force field. V. Extension of MMFF94 using experimental data, additional computational data, and empirical rules," *J. Comput. Chem.* **17**, 616–641 (1996).
- ¹⁶F. Maseras and K. Morokuma, "IMOMM: A new integrated *ab initio* + molecular mechanics geometry optimization scheme of equilibrium structures and transition states," *J. Comput. Chem.* **16**, 1170–1179 (1995).
- ¹⁷S. Dapprich, I. Komáromi, K. Byun, K. Morokuma, and M. J. Frisch, "A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives," *J. Mol. Struct.: THEOCHEM* **461–462**, 1–21 (1999).
- ¹⁸T. Vreven and K. Morokuma, "On the application of the IMOMO (integrated molecular orbital + molecular orbital) method," *J. Comput. Chem.* **21**, 1419–1432 (2000).
- ¹⁹M. Svensson, S. Humbel, and K. Morokuma, "Energetics using the single point IMOMO (integrated molecular orbital + molecular orbital) calculations: Choices of computational levels and model system," *J. Chem. Phys.* **105**, 3654–3661 (1996).
- ²⁰M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber, and K. Morokuma, "ONIOM: A multilayered integrated MO + MM method for geometry optimizations and single point energy predictions. A test for Diels-Alder reactions and Pt(P(*t*-Bu)₃)₂ + H₂ oxidative addition," *J. Phys. Chem.* **100**, 19357–19363 (1996).
- ²¹T. Matsubara, S. Sieber, and K. Morokuma, "A test of the new 'integrated MO + MM' (IMOMM) method for the conformational energy of ethane and *n*-butane," *Int. J. Quantum Chem.* **60**, 1101–1109 (1996).
- ²²S. Humbel, S. Sieber, and K. Morokuma, "The IMOMO method: Integration of different levels of molecular orbital approximations for geometry optimization of large systems: Test for *n*-butane conformation and S_N2 reaction: RCl+Cl⁻," *J. Chem. Phys.* **105**, 1959–1967 (1996).
- ²³D. G. Pettifor and I. I. Oleinik, "Analytic bond-order potentials beyond Tersoff-Brenner. I. Theory," *Phys. Rev. B* **59**, 8487–8499 (1999).
- ²⁴D. G. Pettifor, M. Finnis, D. Nguyen-Manh, D. Murdick, X. Zhou, and H. Wadley, "Analytic bond-order potentials for multicomponent systems," *Mater. Sci. Eng.: A* **365**, 2–13 (2004).
- ²⁵J. Behler and M. Parrinello, "Generalized neural-network representation of high-dimensional potential-energy surfaces," *Phys. Rev. Lett.* **98**, 146401 (2007).
- ²⁶N. Bernstein, G. Csányi, and V. L. Deringer, "De novo exploration and self-guided learning of potential-energy surfaces," *npj Comput. Mater.* **5**, 99 (2019).
- ²⁷G. A. Cisneros, K. T. Wikfeldt, L. Ojamäe, J. Lu, Y. Xu, H. Torabifard, A. P. Bartók, G. Csányi, V. Molinero, and F. Paesani, "Modeling molecular interactions in water: From pairwise to many-body potential energy functions," *Chem. Rev.* **116**, 7501–7528 (2016).
- ²⁸H. Gökcen, E. Kratz, T. A. Darden, J.-P. Piquemal, and G. A. Cisneros, "QM/MM simulations with the Gaussian electrostatic model: A density-based polarizable potential," *J. Phys. Chem. Lett.* **9**, 3062–3067 (2018).
- ²⁹S. Naseem-Khan, J.-P. Piquemal, and G. A. Cisneros, "Improvement of the Gaussian electrostatic model by separate fitting of Coulomb and exchange-repulsion densities and implementation of a new dispersion term," *J. Chem. Phys.* **155**, 194103 (2021).
- ³⁰S. Vandenbrande, M. Waroquier, V. V. Speybroeck, and T. Verstraelen, "The monomer electron density force field (MEDFF): A physically inspired model for noncovalent interactions," *J. Chem. Theory Comput.* **13**, 161–179 (2017).
- ³¹J. Romero, P. Limão-Vieira, K. Hermansson, and M. Probst, "A simple electron-density based force field model for high-energy interactions between atoms and molecules," *J. Phys. Chem. A* **128**, 1163–1172 (2024).
- ³²P. Hohenberg and W. Kohn, "Inhomogeneous electron gas," *Phys. Rev.* **136**, B864–B871 (1964).
- ³³M. Levy and J. P. Perdew, "The constrained search formulation of density functional theory," in *Density Functional Methods in Physics*, edited by R. M. Dreizler and J. da Providência (Springer US, Boston, MA, 1985), pp. 11–30.
- ³⁴J. Katriel, F. Zahariev, and K. Burke, "Symmetry and degeneracy in density functional theory," *Int. J. Quantum Chem.* **85**, 432–435 (2001).
- ³⁵Y. Jung, A. Sodt, P. M. W. Gill, and M. Head-Gordon, "Auxiliary basis expansions for large-scale electronic structure calculations," *Proc. Natl. Acad. Sci. U. S. A.* **102**, 6692–6697 (2005).
- ³⁶G. A. Cisneros, "Application of Gaussian electrostatic model (GEM) distributed multipoles in the AMOEBA force field," *J. Chem. Theory Comput.* **8**, 5072–5080 (2012).
- ³⁷A. D. Becke, "Density-functional thermochemistry. III. The role of exact exchange," *J. Chem. Phys.* **98**, 5648–5652 (1993).
- ³⁸S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, "A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu," *J. Chem. Phys.* **132**, 154104 (2010).
- ³⁹W. J. Stevens, H. Basch, and M. Krauss, "Compact effective potentials and efficient shared-exponent basis sets for the first- and second-row atoms," *J. Chem. Phys.* **81**, 6026–6033 (1984).
- ⁴⁰W. J. Stevens, M. Krauss, H. Basch, and P. G. Jasien, "Relativistic compact effective potentials and efficient, shared-exponent basis sets for the third-fourth-and fifth-row atoms," *Can. J. Chem.* **70**, 612–630 (1992).
- ⁴¹T. R. Cundari and W. J. Stevens, "Effective core potential methods for the lanthanides," *J. Chem. Phys.* **98**, 5555–5565 (1993).
- ⁴²T. H. Dunning, "Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen," *J. Chem. Phys.* **90**, 1007–1023 (1989).
- ⁴³D. E. Woon and T. H. Dunning, "Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon," *J. Chem. Phys.* **98**, 1358–1371 (1993).
- ⁴⁴D. E. Woon and T. H. Dunning, "Gaussian basis sets for use in correlated molecular calculations. V. Core-valence basis sets for boron through neon," *J. Chem. Phys.* **103**, 4572–4585 (1995).
- ⁴⁵A. K. Wilson, T. van Mourik, and T. H. Dunning, "Gaussian basis sets for use in correlated molecular calculations. VI. Sextuple zeta correlation consistent basis sets for boron through neon," *J. Mol. Struct.: THEOCHEM* **388**, 339–349 (1996).
- ⁴⁶J. Cížek, in *Advances in Chemical Physics*, edited by R. LeFebvre and C. Moser (John Wiley & Sons, Inc., 1969).
- ⁴⁷G. D. Purvis and R. J. Bartlett, "A full coupled-cluster singles and doubles model: The inclusion of disconnected triples," *J. Chem. Phys.* **76**, 1910–1918 (1982).
- ⁴⁸J. A. Pople, M. Head-Gordon, and K. Raghavachari, "Quadratic configuration interaction. A general technique for determining electron correlation energies," *J. Chem. Phys.* **87**, 5968–5975 (1987).
- ⁴⁹G. E. Scuseria, C. L. Janssen, and H. F. Schaefer, "An efficient reformulation of the closed-shell coupled cluster single and double excitation (CCSD) equations," *J. Chem. Phys.* **89**, 7382–7387 (1988).
- ⁵⁰G. E. Scuseria and H. F. Schaefer, "Is coupled cluster singles and doubles (CCSD) more computationally intensive than quadratic configuration interaction (QCISD)?," *J. Chem. Phys.* **90**, 3700–3703 (1989).
- ⁵¹R. A. Kendall, T. H. Dunning, and R. J. Harrison, "Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions," *J. Chem. Phys.* **96**, 6796–6806 (1992).

- ⁵²J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general amber force field," *J. Comput. Chem.* **25**, 1157–1174 (2004).
- ⁵³N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison, "Open Babel: An open chemical toolbox," *J. Cheminf.* **3**, 33 (2011).
- ⁵⁴M. S. Gordon, M. A. Freitag, P. Bandyopadhyay, J. H. Jensen, V. Kairys, and W. J. Stevens, "The effective fragment potential method: A QM-based mm approach to modeling environmental effects in chemistry," *J. Phys. Chem. A* **105**, 293–307 (2001).
- ⁵⁵M. S. Gordon, L. Slipchenko, H. Li, and J. H. Jensen, "Chapter 10. The effective fragment potential: A general method for predicting intermolecular interactions," *Annu. Rep. Comput. Chem.* **3**, 177–193 (2007).
- ⁵⁶L. V. Slipchenko and M. S. Gordon, "Damping functions in the effective fragment potential method," *Mol. Phys.* **107**, 999–1016 (2009).
- ⁵⁷I. A. Kaliman and L. V. Slipchenko, "LIBEFP: A new parallel implementation of the effective fragment potential method as a portable software library," *J. Comput. Chem.* **34**, 2284–2292 (2013).
- ⁵⁸M. Riera, E. P. Yeh, and F. Paesani, "Data-driven many-body models for molecular fluids: CO₂/H₂O mixtures as a case study," *J. Chem. Theory Comput.* **16**, 2246–2257 (2020).
- ⁵⁹S. Yue, M. Riera, R. Ghosh, A. Z. Panagiotopoulos, and F. Paesani, "Transferability of data-driven, many-body models for CO₂ simulations in the vapor and liquid phases," *J. Chem. Phys.* **156**, 104503 (2022).
- ⁶⁰M. Riera, C. Knight, E. F. Bull-Vulpe, X. Zhu, H. Agnew, D. G. A. Smith, A. C. Simmonett, and F. Paesani, "MBX: A many-body energy and force calculator for data-driven many-body simulations," *J. Chem. Phys.* **159**, 054802 (2023).
- ⁶¹M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, *Gaussian 16 Revision C.01*, Gaussian Inc., Wallingford CT, 2016.
- ⁶²J. A. Nelder and R. Mead, "A simplex method for function minimization," *Comput. J.* **7**, 308–313 (1965).
- ⁶³J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the Nelder–Mead simplex method in low dimensions," *SIAM J. Optim.* **9**, 112–147 (1998).
- ⁶⁴F. Gao and L. Han, "Implementing the Nelder–Mead simplex algorithm with adaptive parameters," *Comput. Optim. Appl.* **51**, 259–277 (2012).
- ⁶⁵C. G. Broyden, "The convergence of a class of double-rank minimization algorithms 1. General considerations," *IMA J. Appl. Math.* **6**, 76–90 (1970).
- ⁶⁶R. Fletcher, "A new approach to variable metric algorithms," *Comput. J.* **13**, 317–322 (1970).
- ⁶⁷D. Goldfarb, "A family of variable-metric methods derived by variational means," *Math. Comput.* **24**, 23–26 (1970).
- ⁶⁸D. F. Shanno, "Conditioning of quasi-Newton methods for function minimization," *Mathematics of Computation* **24**, 647–656 (1970).
- ⁶⁹J. Nocedal, "Updating quasi-Newton matrices with limited storage," *Math. Comput.* **35**, 773–782 (1980).
- ⁷⁰D. C. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," *Math. Program.* **45**, 503–528 (1989).
- ⁷¹J. Nocedal and J. W. Stephen, in *Numerical Optimization*, 2nd ed. (Springer, New York, 2006).
- ⁷²J. Revels, M. Lubin, and T. Papamarkou, "Forward-mode automatic differentiation in Julia," [arXiv:1607.07892](https://arxiv.org/abs/1607.07892) [cs.MS] (2016).
- ⁷³Z.-Z. Yang and C.-S. Wang, "Atom–bond electronegativity equalization method. 1. Calculation of the charge distribution in large molecules," *J. Phys. Chem. A* **101**, 6315–6321 (1997).
- ⁷⁴P. Bultinck, W. Langenaeker, P. Lahorte, F. De Proft, P. Geerlings, M. Waroquier, and J. P. Tollenaere, "The electronegativity equalization method I: Parametrization and validation for atomic charge calculations," *J. Phys. Chem. A* **106**, 7887–7894 (2002).
- ⁷⁵Maxima, Maxima, a computer algebra system. version 5.47.0, 2023.
- ⁷⁶J. Romero, MolFFSim - A C++ implementation of an OFDFT based molecular force field model, 2024.
- ⁷⁷G. Guennebaud, B. Jacob *et al.*, Eigen v3, 2010.
- ⁷⁸N. Okazaki, libLBFGS: L-BFGS library written in C, 2014.
- ⁷⁹A. M. M. Leal, autodiff, a modern, fast and expressive C++ library for automatic differentiation, 2018.