Theor Chem Acc (2012) 131:1145
DOI 10.1007/s00214-012-1145-7

REGULAR ARTICLE

Density-functional expansion methods: grand challenges

Timothy J. Giese · Darrin M. York

Received: 13 May 2011 / Accepted: 9 September 2011

Abstract We discuss the source of errors in semiempirical density-functional expansion (VE) methods. In particular, we show that VE methods are capable of well reproducing their standard Kohn-Sham density-functional method counterparts, but suffer from large errors upon using one or more of these approximations: the limited size of the atomic orbital basis, the Slater monopole auxiliary basis description of the response density, and the one- and two-body treatment of the core-Hamiltonian matrix elements. In the process of discussing these approximations and highlighting their symptoms, we introduce a new model that supplements the second-order density-functional tight-binding model with a self-consistent charge-dependent chemical potential equalization correction; we review our recently reported method for generalizing the auxiliary basis description of the atomic orbital response density; and we decompose the first-order potential into a summation of additive atomic components and many-body corrections, and from this examination, we provide new insights and preliminary results that motivate and inspire new approximate treatments of the core-Hamiltonian.

Keywords Tight-binding models · Density-functional theory · Electronic structure

Published as part of the special collection of articles: From quantum mechanics to force fields: new methodologies for the classical simulation of complex systems.

T. J. Giese · D. M. York (✉)
Department of Chemistry and Chemical Biology and BioMaPS Institute for Quantitative Biology, Rutgers University, Piscataway, NJ 08854-8087, USA
e-mail: york@biomaps.rutgers.edu

1 Introduction

A great deal of effort, spanning several decades, has been devoted to the design, implementation, and parametrization of semiempirical quantum models to make them practical, efficient, and accessible to a wide variety of problems [1–22]. As a consequence of the community's success, it can be tempting to dismiss the approximations used in these models as being technical curiosities or to simply underestimate their severity; it is only upon a careful examination of the magnitude of the corrections required to make these methods successful that one gains an appreciation for their ingenuity, and a somewhat worrying sense of amazement that they should even work at all. Although their approximations may be severe, there are only two metrics by which semiempirical models are ultimately judged: they must be *fast* and *accurate*. It is rare for the approximations to be criticized for their severity because the efficiency of the models is founded upon them, and the models can yet be persuaded to reproduce good results. It was for these reasons, after all, that the original developers of the semiempirical models rationalized their use, and it is easy to justify their continued use in the newer models if they are but minor modifications of those original works. This is not to say that the semiempirical development community has stagnated in groupthink, we mean only to highlight the realities that drive the direction of model development. The designer of a next-generation semiempirical model will do so under the influence of one of two prejudices: (1) current models are "fine as they are" and need only minor modifications to their parameters or functional forms; or (2) the parameters of the original semiempirical models have now been sufficiently trained [19], and the parametric freedom of the *ad hoc* functions used to replace the model's missing physics has now been

Published online: 21 February 2012

sufficiently extended [15], that a further significant advance will need to address the fundamental approximations of the method in some inherently new way [23–25]. The former option is attractive because it promises to offer instant gratification and can be more easily demonstrated to be fast and accurate. The latter option is higher risk and potentially requires the development of new mathematical and technical advances to convince people that it is worthwhile.

The above discussion used the words "semiempirical models" and "approximation" vaguely, so it is useful to continue by discussing the effects of a specific approximation used by a specific model. Consider the self-consistent-charge density-functional tight-binding (SCC-DFTB) model [22]: its "core-Hamiltonian" matrix, that is, the matrix of electron kinetic energies and the first-order density-response interactions with the neutral-atom reference, is constructed by an approximation whereby the one-center atomic orbital (AO) blocks presume that the system is but a single atom (itself), and the two-center AO blocks are computed without regard to the presence of any third atom, even if that atom lies directly between the two. Although we feel it is fair to claim that this approximation is severe by using no more than common sense, we have recently examined its use within the framework of Kohn-Sham potential energy expansion (VE) models to quantify its effect [26]. We performed second-order VE calculations with the PBE exchange-correlation (xc) functional [27] and an all-electron 6-31G* AO basis, but which was otherwise free of approximations and parameters, and showed that this reproduces standard PBE/6-31G* results almost exactly; however, when we then applied the above discussed approximation, it caused bond lengths to decrease by about 0.45 Å and angles to have errors in excess of 10°. These poor geometries resulted in dipole moment errors in excess of 50%, and the water dimer was found to be overbound by approximately two orders of magnitude. Although this set of approximations has severe consequences if left unchecked, SCC-DFTB mitigates the adverse effects with surprising success primarily by repelling the atoms with a pairwise additive correction to the energy.

Our goal is not to single-out and criticize SCC-DFTB for the above-mentioned approximation, which has been noted many times in the past [22, 28–30], and other semiempirical models have used conceptually similar approximations. Modified neglect of differential overlap (MNDO) models [11, 13] constructs their core-Hamiltonian matrices using a concept similar to the one described above, even though the theoretical foundations for these methods are quite different [28]. MNDO methods compute the two-center core-Hamiltonian AO blocks from scaled overlap integrals instead of looking up their values from precomputed tables, but like SCC-DFTB, explicit contributions from other atoms are neglected. MNDO's treatment of the one-center core-Hamiltonian AO blocks is also similar to SCC-DFTB in that it considers the other atoms only insofar as it tries to eliminate the influence of neutral atoms in the resulting "Fock" matrix. And like SCC-DFTB, MNDO relies on the use of pairwise repulsive potentials to retain good geometries. Of course, there are important differences between these methods, such as their treatment of electrostatics and AO overlaps, and these differences should lead one to suspect that each method has varying degrees of success at modeling various properties. Indeed, articles that compare SCC-DFTB and MNDO-like methods suggest that this is true [31], but the differences are such that it is difficult to conclude that one is an overall better method than the other. When we read articles that compare these methods [31], and we ponder their similarities and differences [28], we may verily ask: Could it be that MNDO and SCC-DFTB models are of comparable overall accuracy because their core-Hamiltonians employ conceptually similar approximations, and that these approximations are mitigated in both methods by pairwise additive repulsive energy corrections? Is this approximation a *glass ceiling* that will prevent new models from making significant advances in overall accuracy?

Much of our work has focused on the exploration of new methods and the development of mathematical tools which, together, motivate the design of next-generation VE models [26, 32–37]. These techniques are meant to improve current models by reintroducing the physics that are otherwise only implicitly accounted for. In the process of demonstrating that these techniques add missing physics, we isolate and judge the severity of the underlying approximations used within existing models. To this end, we often find it difficult, and sometimes misleading, to judge the severity of specific approximations by interpreting the failures of parametrized models, because multiple approximations can be parametrized to cancel in unpredictable ways. To minimize the obfuscation of our comparisons, we begin with a parameter-free ab initio-like VE model that reproduces density-functional theory (DFT) results very well and then introduce various approximations to identify where and quantify by how much the model breaks down. The insights that are gained from these explorations, in turn, provide the concepts and motivation for new approximations. Using this strategy, we have concluded [26] that a second-order VE method reproduces standard DFT results extremely well if the model is not encumbered by the following approximations: (1) the limited size of the AO basis, (2) the Slater monopole auxiliary basis representation of the response density, and (3) the one- and two-body treatment of the first-order matrix elements (as discussed above). There are other, more benign approximations, but overcoming the most severe approximations with computational efficiency is *the*

grand challenge for Kohn-Sham potential energy expansion models.

Over the years, we have attempted to address these approximations, and we will review and extend some of those works here so as to unify their motivation within the perspective of this grand challenge. The purpose of this work is threefold: (1) We present a new model that supplements second-order SCC-DFTB (DFTB2) with a self-consistent *charge-dependent* dipole polarization correction (DFTB2+CPEQ). This model is based on our earlier work [36, 37] (post-self-consistent field MNDO/d+CPE), but is made self-consistent, and is applied to a new Hamiltonian. In addition, we go beyond the analysis presented in our previous works by making explicit comparison with a *charge-independent* model (DFTB2+CPE0), to ascertain the benefits of the charge dependence. (2) We review our recent work that generalizes the auxiliary basis representation of the response density beyond Slater monopoles [32]. (3) We present new insights into the atomic decomposition of the nonadditive VE reference potential energy and its functional derivatives and discuss how this relates to the modeling of the core-Hamiltonian. Furthermore, we demonstrate how our observations lead to a new mathematical justification for the approximations used within spin-polarized SCC-DFTB models [38].

2 Methods

2.1 Expansion of the Kohn-Sham potential energy

Expansion models are derived from Kohn-Sham DFT through Taylor series expansion [29] of the potential energy [39] $V[\rho, \omega]$ in density response $\delta\rho(\mathbf{r}) = \rho(\mathbf{r}) - \rho_{\text{ref}}(\mathbf{r})$ and spin density $\omega(\mathbf{r}) = \rho^{\alpha}(\mathbf{r}) - \rho^{\beta}(\mathbf{r})$ about the reference density composed from the spin-averaged sum of isolated neutral atoms $\rho_{\text{ref}}(\mathbf{r}) = \sum_a \rho_a(\mathbf{r})$ and $\omega_{\text{ref}}(\mathbf{r}) = 0$. The VE energy $E_{\text{VE}} \equiv E_{\text{VE}}[\rho_{\text{ref}} + \delta\rho, \omega]$ is

$$E_{\text{VE}} = \sum_{ij} P_{ij}T_{ij} + V^{(0)}[\rho_{\text{ref}}] + V^{(1)}[\delta\rho] + W[\delta\rho, \omega], \quad (1)$$

where $\mathbf{P}$ is the AO basis single-particle density matrix; $\mathbf{T}$ is the kinetic energy matrix;

$$V^{(0)}[\rho_{\text{ref}}] = E_{\text{xc}}[\rho_{\text{ref}}, 0] + \sum_{b>a} \frac{Z_a Z_b}{R_{ab}} + J[\rho_{\text{ref}}] - \sum_{ab} \int \frac{Z_a \rho_b(\mathbf{r})}{|\mathbf{r} - \mathbf{R}_a|} d^3r \quad (2)$$

is the reference potential energy; Z_a is a nuclear charge; $\mathbf{R}_a$ is its atomic position;

$$J[\rho] = \frac{1}{2}\int\int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3r d^3r' \quad (3)$$

is the Coulomb functional; $E_{\text{xc}}[\rho, \omega]$ is the xc-functional; $V^{(1)}[\delta\rho] = \sum_{ij} (P_{ij} - P^{(0)}_{ij})\, V^{(1)}_{ij}$, where

$$V^{(1)}_{ij} = \int \left.\frac{\delta V[\rho, \omega]}{\delta\rho(\mathbf{r})}\right|_0 \chi_i(\mathbf{r})\chi_j(\mathbf{r}) d^3r, \quad (4)$$

is the first-order potential energy; $\mathbf{P}^{(0)}$ is the reference density matrix, which reproduces $\rho_{\text{ref}}(\mathbf{r})$ from the AO basis $\{\boldsymbol{\chi}\}$; the "0" subscript outside functional derivatives indicate its evaluation about the spin-averaged reference; and $W[\delta\rho, \omega]$ is a functional that describes all other response interactions. If the Taylor series expansion was carried out to infinite order, then $W[\delta, \omega]$ would be the second- and all higher-order energy corrections; however, we specifically write $W[\delta, \omega]$ as a generic functional to allow us the notational freedom to insert other empirical correction terms that depend on the response density in some complicated way. We will explore the use of empirical correction terms in later sections and continue here by considering the case of a second-order expansion; that is, $W[\delta\rho, \omega] \rightarrow V^{(2)}[\delta\rho, \omega]$, where

$$V^{(2)}[\delta\rho, \omega] = J[\delta\rho] + \frac{1}{2}\int\int \left.\frac{\delta^2 E_{\text{xc}}[\rho, \omega]}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')}\right|_0 \times \delta\rho(\mathbf{r})\delta\rho(\mathbf{r}') d^3r d^3r' + \frac{1}{2}\int\int \left.\frac{\delta^2 E_{\text{xc}}[\rho, \omega]}{\delta\omega(\mathbf{r})\delta\omega(\mathbf{r}')}\right|_0 \times \omega(\mathbf{r})\omega(\mathbf{r}') d^3r d^3r'. \quad (5)$$

If $E_{\text{xc}}[\rho, \omega]$ is a local functional of the density, that is, if it does not require the calculation of nonlocal integrals, then the last two terms in Eq. 5 simplify [40],

$$\left.\frac{\delta^2 E_{\text{xc}}[\rho, \omega]}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')}\right|_0 = \delta(\mathbf{r} - \mathbf{r}') \left.\frac{\delta^2 E_{\text{xc}}[\rho, \omega]}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r})}\right|_0; \quad (6)$$

and the functional derivatives can be expressed as local functions of space

$$E_{\text{VE}} = \sum_{ij} P_{ij}T_{ij} + V^{(0)}[\rho_{\text{ref}}] + \int v^{(1)}_{\text{ref}}(\mathbf{r})\delta\rho(\mathbf{r}) d^3r + J[\delta\rho] + \frac{1}{2}\int v^{(2)}_{\text{ref}}(\mathbf{r})\delta\rho(\mathbf{r})^2 d^3r + \frac{1}{2}\int w^{(2)}_{\text{ref}}(\mathbf{r})\omega(\mathbf{r})^2 d^3r; \quad (7)$$

where

$$v^{(1)}_{\text{ref}}(\mathbf{r}) = -\sum_a \frac{Z_a}{|\mathbf{r} - \mathbf{R}_a|} + \int \frac{\rho_{\text{ref}}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3r' + \left.\frac{\delta E_{\text{PBE}}[\rho, \omega]}{\delta\rho(\mathbf{r})}\right|_0, \quad (8)$$

and $v^{(2)}_{\text{ref}}(\mathbf{r})$ and $w^{(2)}_{\text{ref}}(\mathbf{r})$ are the second-functional derivatives of the local xc-functional. Various models can be

constructed by applying approximations to the above equations, and these models are described in the next sections.

2.2 Second-order SCC-DFTB

The DFTB2 model [22] differs from Eqs. 1–7 in the following ways:

$$V_{ij}^{(1)} \approx \begin{cases} \int v_a^{(1)}(\mathbf{r})\chi_i(\mathbf{r})\chi_j(\mathbf{r})d^3r, & \text{if } i,j \in a \\ \int v_{ab}^{(1)}(\mathbf{r})\chi_i(\mathbf{r})\chi_j(\mathbf{r})d^3r, & i \in a,\ j \in b, \end{cases} \tag{9}$$

where $v_a^{(1)}(\mathbf{r}) = \delta V[\delta\rho,\omega]/\delta\rho|_{\rho(\mathbf{r})=\rho_a(\mathbf{r})}$ is the first-order potential of atom a and $v_{ab}^{(1)}(\mathbf{r})$ is the potential of the ab-dimer; the second-order energy is evaluated using a Slater monopole auxiliary basis

$$\delta\rho(\mathbf{r}) = -\sum_a q^a\varphi^a(\mathbf{r}-\mathbf{R}_a), \tag{10}$$

where

$$q^a = -\sum_i\sum_{j\in a}(P_{ij}-P_{ij}^{(0)})\frac{1}{2}S_{ij} - \sum_{i\in a}\sum_j(P_{ij}-P_{ij}^{(0)})\frac{1}{2}S_{ij} \tag{11}$$

is the Mulliken-partitioned partial electric charge of atom a, and $\mathbf{S}$ is the overlap matrix. The second-order energy $V^{(2)}$ is Coulomb-approximated; the auxiliary basis representation transforms the functional into a function of partial charges and atomic positions $V^{(2)}[\delta\rho,\omega] \rightarrow J(\mathbf{q},\mathbf{R})$, where

$$J(\mathbf{q},\mathbf{R}) = \sum_{ab} q^a q^b \frac{1}{2}\int\int \frac{\varphi^a(\mathbf{r})\varphi^b(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d^3r d^3r'. \tag{12}$$

Although this only explicitly models the electrostatics, the second-order xc-energy is implicitly accounted for by choosing the Slater exponents so that the Coulomb self-energies of the atoms reproduce their experimental chemical hardness [30, 41]. The matrix elements (Eq. 9) are evaluated using the atomic densities and minimal valence AOs resulting from isolated atom calculations in the presence of a radial confinement potential, and the xc-potentials are evaluated with the PBE functional [27]. To compensate for the above approximations, the reference potential energy is replaced by a sum of pairwise additive corrections

$$V^{(0)}[\rho_{\text{ref}}] \rightarrow \sum_{b>a} f_{ab}(R_{ab}), \tag{13}$$

which are parametrized to reproduce molecular geometries and bond energies.

The DFTB2 energy is minimized via a standard self-consistent field (SCF) procedure, and the expression for the Fock matrix $F_{ij} = \delta E/\delta P_{ij}$ is

$$F_{ij} = T_{ij} + V_{ij}^{(1)} - \frac{1}{2}S_{ij}\left(\frac{\partial J(\mathbf{q},\mathbf{R})}{\partial q^a} + \frac{\partial J(\mathbf{q},\mathbf{R})}{\partial q^b}\right), \tag{14}$$

where $i \in a$ and $j \in b$.

2.3 DFTB2 with self-consistent polarization corrections

This section describes a new method that is similar to Refs. [42–44]; in that, it is founded upon our earlier works [36, 37]. The difference between the present method and Refs. [42–44] is: our new method is charge-dependent; but unlike Refs. [36, 37], the present work is self-consistent and applied to a different Hamiltonian. Our motivation for introducing this method here is to address the first grand challenge: The limited size of the AO basis.

Like all minimal valence basis models, DFTB2 lacks the sufficient AO degrees of freedom to adequately respond to external perturbations, and this manifests itself in its underprediction of dipole polarizabilities [45]. Obviously, one could increase the size of the AO basis in an attempt to alleviate the underlying cause of the problem [46]; however, the benefits of doing so would not be fully realized if the electrostatics were limited to monopolar interactions. In lieu of developing an entirely new model, the dipole polarizabilities of DFTB2 can be improved by including a chemical potential equalization (CPE) auxiliary response density [47], represented by atom-centered dipole functions, whose response to external perturbations is not constrained by the limited size of the AO basis.

The CPE response density is the namesake of *the CPE principle*, sometimes referred to as the electronegativity equalization principle [48–51], and is derived from DFT by Taylor expanding the density functional in density response [52, 53]. In this context, the DFTB2 AO density $\rho_{\text{AO}}(\mathbf{r})$ is the reference density in the CPE Taylor expansion, and the CPE density $\delta\widetilde{\rho}(\mathbf{r})$is the response about this reference; that is,

$$\begin{aligned} E[\rho_{\text{AO}}+\delta\widetilde{\rho}] = E_{\text{VE}}[\rho_{\text{AO}}] &+ \int\left.\frac{\delta E[\rho]}{\delta\rho(\mathbf{r})}\right|_{\rho_{\text{AO}}}\delta\widetilde{\rho}(\mathbf{r})d^3r \\ &+ \frac{1}{2}\int\int\left.\frac{\delta^2E[\rho]}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')}\right|_{\rho_{\text{AO}}} \\ &\times\delta\widetilde{\rho}(\mathbf{r})\delta\widetilde{\rho}(\mathbf{r}')d^3rd^3r', \end{aligned} \tag{15}$$

and the CPE correction to the VE energy is

$$E_{\text{CPE}}[\rho_{\text{AO}},\delta\widetilde{\rho}] \equiv E[\rho_{\text{AO}}+\delta\widetilde{\rho}] - E_{\text{VE}}[\rho_{\text{AO}}]. \tag{16}$$

Although they are both Taylor series expansions, Eq. 15 is fundamentally different from Eqs. 1, 15 expands the total density functional, including the electron kinetic energy, as opposed to the Kohn-Sham potential energy functional 39. This distinction is important because it has serious implications on the ability of the model to describe covalent bonding. Equation (15) is not a reasonable model for describing covalent bonds, which are difficult to describe outside of the linear combination of atomic orbitals (LCAO) approach, because the kinetic energy is not well modeled through the density alone [54]. Fortunately, we can rely on DFTB2 for that purpose and use the CPE correction to account for the missing response due to those interactions that are sufficiently separated that they can be accurately modeled by electrostatics only. This is, in fact, what the polarizabilities ultimately *are*: measures of response due to distant electrostatic perturbations. In particular, the isotropic dipole polarizability of a molecule is $\alpha = (\alpha_{xx} + \alpha_{yy} + \alpha_{zz})/3$, where

$$\alpha_{xy} = \frac{\partial^2 E}{\partial E_x \partial E_y} = -\frac{\partial q_x^{(1)}}{\partial E_y}, \tag{17}$$

and

$$q_x^{(1)} = \int x \left[\sum_a \frac{Z_a}{|\mathbf{r} - \mathbf{R}_a|} - \rho_{\mathrm{AO}}(\mathbf{r}) - \delta\widetilde{\rho}(\mathbf{r}) \right] d^3 r \tag{18}$$

and E_x are the electric dipole moment and field strength in the x-direction, respectively.

For the reasons discussed above, we approximate $E[\rho] \approx J[\rho]$ and damp the short-range Coulomb interactions between the CPE density and the DFTB2 AO density, because the short-range interactions of the density are better modeled by the underlying DFTB2 model. Furthermore, being that only the well-separated Coulomb interactions are treated, we ignore the underlying neutral-atom DFTB2 reference; only the DFTB2 partial atomic charges (and any applied electric field) invoke a CPE response $\rho_{\mathrm{AO}}(\mathbf{r}) \approx -\sum_a q^a \varphi^a(\mathbf{r})$. The CPE response is represented by atom-centered primitive Gaussian-dipole functions $\{\widetilde{\varphi}\}$

$$\begin{aligned} \delta\widetilde{\rho}(\mathbf{r}) &= -\sum_a \sum_\mu d_\mu^a \widetilde{\varphi}_\mu^a(\mathbf{r} - \mathbf{R}_a) \\ &= -\sum_a \sum_{\mu=-1}^{1} d_\mu^a C_{1,\mu}(\bigtriangledown_a) \left(\frac{\zeta_a}{\pi}\right)^{3/2} e^{-\zeta_a |\mathbf{r} - \mathbf{R}_a|^2}, \end{aligned} \tag{19}$$

where $C_{l\mu}(\bigtriangledown_a)$ is the spherical tensor gradient operator [55] acting on $\mathbf{R}_a$, and $\mathbf{d}$ is the vector of atomic dipole moments. The basis representation of $\delta\widetilde{\rho}(\mathbf{r})$ allows us to rewrite the CPE correction energy in a simple algebraic form $E_{\mathrm{CPE}}[\rho_{\mathrm{AO}}, \delta\widetilde{\rho}] \rightarrow E_{\mathrm{CPE}}(\mathbf{q}, \mathbf{d}, \mathbf{R})$; that is,

$$E_{\mathrm{CPE}}(\mathbf{q}, \mathbf{d}, \mathbf{R}) = \mathbf{d}^T \cdot \mathbf{B} \cdot \mathbf{q} + \frac{1}{2} \mathbf{d}^T \cdot \mathbf{A} \cdot \mathbf{d}, \tag{20}$$

where

$$B_{kb} = D(R_{ab}) \int \int \frac{\widetilde{\varphi}_k^a(\mathbf{r}) \varphi^b(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3 r d^3 r' \tag{21}$$

is a $3N \times N$ matrix of scaled Coulomb interactions between the CPE Gaussian-dipole and DFTB2 AO Slater monopole auxiliary basis sets;

$$A_{kk'} = \int \int \frac{\widetilde{\varphi}_k^a(\mathbf{r}) \widetilde{\varphi}_{k'}^b(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3 r d^3 r' \tag{22}$$

is a $3N \times 3N$ matrix of Gaussian-dipole Coulomb interactions; and $D(R_{ab})$ is a switching function that removes the short-range Coulomb interactions. This switching function has the form: $D(R_{ab}) = 1$, if $R_{ab} > R_{\mathrm{hi}}$; $D(R_{ab}) = 0$, if $R_{ab} < R_{\mathrm{lo}}$; and

$$D(R_{ab}) = 1 - 10x^3 + 15x^4 - 6x^5, \tag{23}$$

otherwise; where $x = (R_{\mathrm{hi}} - R_{ab}) / (R_{\mathrm{hi}} - R_{\mathrm{lo}})$; and $R_{\mathrm{hi}} = R_{\mathrm{hi},a} + R_{\mathrm{hi},b}$ and $R_{\mathrm{lo}} = R_{\mathrm{lo},a} + R_{\mathrm{lo},b}$ are a sum of atom parameters.

The CPE correction energy is minimized by solving $\partial E_{\mathrm{CPE}}(\mathbf{q}, \mathbf{d}, \mathbf{R}) / \partial d_k^a = 0$, which has an analytic result:

$$\mathbf{d} = -\mathbf{A}^{-1} \cdot \mathbf{B} \cdot \mathbf{q}. \tag{24}$$

In summary, given the short-range damping parameters $R_{\mathrm{lo},a}$ and $R_{\mathrm{hi},a}$, and Gaussian-dipole exponents ζ_a, the matrices $\mathbf{A}$ and $\mathbf{B}$ are computed (Eqs. 21–22), the response is then determined (Eq. 24), and the energy correction is then evaluated (Eq. 20).

The CPE energy correction is a modification of the W-functional, that is, $W[\delta\rho, \omega] \rightarrow J(\mathbf{q}, \mathbf{R}) + E_{\mathrm{CPE}}(\mathbf{q}, \mathbf{d}, \mathbf{R})$, and depends on the density matrix through $\mathbf{q}$. Therefore, its derivatives with respect to $\mathbf{q}$ enter the Fock matrix;

$$\begin{aligned} F_{ij} = T_{ij} + V_{ij}^{(1)} \\ - \frac{1}{2} S_{ij} \left(\frac{\partial J(\mathbf{q}, \mathbf{R})}{\partial q^a} + \frac{\partial E_{\mathrm{CPE}}(\mathbf{q}, \mathbf{d}, \mathbf{R})}{\partial q^a} \right. \\ \left. + \frac{\partial J(\mathbf{q}, \mathbf{R})}{\partial q^b} + \frac{\partial E_{\mathrm{CPE}}(\mathbf{q}, \mathbf{d}, \mathbf{R})}{\partial q^b} \right). \end{aligned} \tag{25}$$

CPE0: The charge-independent model. If the Gaussian-dipole exponents did not depend on charge, then

$$\frac{\partial E_{\mathrm{CPE0}}(\mathbf{q}, \mathbf{d}, \mathbf{R})}{\partial q^a} = \sum_b \sum_{k \in b} d_k^b B_{ka}. \tag{26}$$

CPEQ: The charge-dependent model. In this model, the CPE Gaussian-dipole exponent is made a function of the underlying DFTB2 partial charge

$$\zeta_a(q^a) = \zeta_a(0) e^{b_a q^a}, \tag{27}$$

where $\zeta_a(0)$and b_a are atom parameters.

The purpose for making the model charge-dependent is to introduce some elementary, missing physics. In particular, anions are known to be more polarizable than neutral atoms, which are more polarizable than cations. The atomic polarizabilities of minimal valence basis models tend to have the opposite behavior because, as the AO degrees of freedom are extinguished, there is less opportunity to respond to external perturbations. Equation 27 reproduces the expected behavior when $b_a > 0$: as an atom becomes anionic $q^a < 0, \zeta_a(q^a)$ becomes smaller; the Coulomb self-energy A_{kk} decreases; and therefore, the polarizability increases; that is, the CPE polarizability correction of an isolated atom is inversely proportional to the dipole self-energy, $\Delta\alpha = 1/A_{kk}$.

Notice that the charge dependence takes an exponential-, as opposed to linear form. One can compute the polarizability of atoms and ions using standard ab initio methods to convince themselves of the exponential behavior, but it can also be gleaned intuitively from tables of experimental atomic electron affinities and ionization potentials, which can be used to plot the energy of an atom as a function of charge state. These plots are monotonic and exponential-like and are interpreted to suggest certain truisms: Excess electrons are less bound to the atom and are thus sensitive to external perturbations. The opposite is true when there are a deficient number of electrons. Moreover, the inner-shell electrons are successively bound with ever-increasing strength to the nucleus, making their response to external perturbations exponentially irrelevant.

Having now explained the exponential behavior and appreciating that, by doing so, this behavior now seems obvious; the reader may question why its explanation is even necessary. Our reason for explaining this in detail is because there have been recent works [56, 57] that may cause some readers to believe that the justification for including charge dependence is founded in the expansion of the Taylor series to third order; because, in this way, the second-order energy can be corrected for *linear* changes in the charge. This, of course, is a mathematically sound way for introducing charge dependence; however, it is unfortunate that a reader can easily misconstrue it into suggesting that the problem at hand is a symptom of a premature truncation of the Taylor series expansion. The symptoms, however, result from a poor auxiliary description of the response density and its restricted response in a limited AO basis. Indeed, a VE model with a sufficient AO and auxiliary basis, but in the absence of other severe approximations, reproduces closed-shell standard DFT results very well even when the VE Taylor expansion of the xc-functional is truncated at first order!

By making the ζ_a's a function of the DFTB2 partial charges, the **A** and **B** matrix elements now have an explicit **q**-dependence, and this complicates the Fock matrix correction

$$\frac{\partial E_{\mathrm{CPEQ}}(\mathbf{q},\mathbf{d},\mathbf{R})}{\partial q^a} = \sum_b \sum_{k\in b} d_k^b B_{ka} + \sum_{k\in a}\sum_b d_k^a \frac{\partial B_{kb}}{\partial q^a} q^b + \sum_{k\in a}\sum_b \sum_{k'\in b} d_k^a \frac{1}{2}\frac{\partial A_{kk'}}{\partial q^a} d_{k'}^b. \quad (28)$$

If the only dependence of the matrix elements are through the CPE Gaussian-dipole exponent, as is the case for DFTB2, then their required derivatives can be expressed solely through a chain rule of those exponents; for example,

$$\frac{\partial B_{kb}}{\partial q^a} = \left.\frac{\partial B_{kb}}{\partial \zeta_a}\right|_{\zeta_a=\zeta_a(q^a)} \frac{\partial \zeta_a(q^a)}{\partial q^a}, \quad \text{for } k \in a; \quad (29)$$

however, if one were to apply the CPE correction to a third-order SCC-DFTB model (DFTB3) [56–58], whose Slater monopole exponents also have a **q**-dependence, then an additional derivative of **B** would be necessary.

2.4 PBE/6-31G*-based VE models

As discussed above, DFTB2 is limited in both the completeness of its AO basis and the electrostatic description of its response density. In the previous sections, we described how one can try to mask these limitations by making a modest change or extension to the model. In this section, we describe a series of models that approach the problem from a different perspective: we start with a base model that reproduces standard PBE/6-31G* very well and then apply approximations that worsen the model. This allows us to better identify and quantify the limitations of the model, motivate and test new corrections, and convincingly demonstrate that those corrections reintroduce missing physics.

To describe these models, it is useful to discuss how they differ from a base model, which we call VE. As the name suggests, it is a direct application of the VE method (Eqs. 1–8), and it is our best attempt at avoiding all other approximations. The VE model is evaluated with the AOs and reference density resulting from standard PBE/6-31G* calculations, uses the PBE xc-functional within the zeroth-order reference energy and first-order potential, and uses the SPW92 xc-functional [59] within the second-order terms; that is, the second-order potentials, $v_{\mathrm{ref}}^{(2)}(\mathbf{r})$ and $w_{\mathrm{ref}}^{(2)}(\mathbf{r})$, are the second functional derivatives of SPW92. Had we used the PBE xc-functional for the second functional derivative and carried out the Taylor series expansion to infinite order, we would expect VE to exactly

reproduce standard PBE/6-31G*; therefore, we will compare all of our VE models to standard PBE/6-31G*.

The other models that we describe employ various approximations; some of which are more severe than others. The benign approximations are relevant in our examinations insofar as they are commonly used in existing semiempirical models and yet provide certain computational advantages. Therefore, we will also introduce more severe approximations upon those, and in these cases, we will continue to compute their errors relative to standard PBE/6-31G*, but with the understanding that, even if we remedied the severe approximations with a new technique, we'd still expect the model to suffer from the errors attributed to the more benign approximations.

The remaining, closely related models are summarized below. Some of the approximations used in these models require more explanation than what is described here, and we defer those explanations to the ensuing sections that follow.

VE: The base model described by Eqs. 1–8, where the AO basis and reference densities result from 6-31G* atom calculations, and $E_{\mathrm{xc}}[\rho, \omega] = E_{\mathrm{PBE}}[\rho, \omega]$, except for the second functional derivatives, which use $E_{\mathrm{xc}}[\rho, \omega] = E_{\mathrm{SPW92}}[\rho, \omega]$. This model differs from standard PBE/6-31G* because of our truncation of the Taylor expansion to second order, and our uniform gas xc-approximation for the second functional derivatives.

VE0: The VE model, where the reference potential energy is approximated by a cluster expansion, that is,

$$V[\rho_{\mathrm{ref}}, 0] \approx \sum_a V_a^{(0)} + \sum_{b>a} \Delta V_{ab}^{(0)}, \tag{30}$$

where

$$\Delta V_{ab}^{(0)} = V_{ab}^{(0)} - V_a^{(0)} - V_b^{(0)}, \tag{31}$$

$V_{ab}^{(0)} \equiv V[\rho_a + \rho_b, 0]$, and $V_a^{(0)} \equiv V[\rho_a, 0]$. If the cluster expansion was exact, then we would expect VE0 reproduce the VE results

VE1: The VE0 model, where the first-order potential energy is evaluated using a one- and two-body approximation (Eq. 9). If Eq. 9 was exact, we would expect VE1 to reproduce the VE0 results.

VE1S: The VE0 model, where the first-order potential (Eq. 8) is approximated by a sum of isolated atom potentials; that is, $v_{\mathrm{ref}}^{(1)}(\mathbf{r}) \approx \sum_a v_a^{(1)}(\mathbf{r})$. If this approximation was exact, we would expect VE1S to reproduce the VE0 results.

VE1W: The VE0 model, where the first-order potential is approximated by a weighted cluster expansion (Eq. 48), described later in this text. If Eq. 48 was exact, we would expect VE1W to reproduce the VE0 results.

VEJ: The VE model, where the second-order xc-potentials are completely ignored; that is, $v_{\mathrm{ref}}^{(2)}(\mathbf{r}) = w_{\mathrm{ref}}^{(2)}(\mathbf{r}) = 0$. VEJ cannot describe spin polarization at all.

VEJ/1S(M): The VEJ model, where the AO products are represented in an auxiliary basis of atom-centered Slater monopole functions, whose exponents are chosen to reproduce the experimental hardness of the atoms. This auxiliary basis is analogous to that used in DFTB2; like DFTB2, the atomic charges are chosen by Mulliken partitioning. If the auxiliary basis exactly reproduced the AO products, then we'd expect VEJ/1S(M) to reproduce the VEJ results.

VEJ/2P,3D,4F: The VEJ model, where the AO products are represented with an auxiliary basis of Gaussian-multipole expansions (GME)

$$\varphi_{lm}^a(\mathbf{r}) = \frac{C_{lm}(\bigtriangledown_a)}{(2l-1)!!} \times \sum_{\gamma=1}^{3} c_\gamma \left(\frac{\zeta_\gamma(\xi)}{\pi}\right)^{3/2} e^{-\zeta_\gamma(\xi)|\mathbf{r}-\mathbf{R}_a|^2}, \tag{32}$$

whose primitive contraction coefficients c_γ, and exponents $\zeta_\gamma(\xi)$, have been chosen to mimic a Slater multipole-like expansion with exponent ξ. The notation 2P,3D,4F indicates that period 1 elements use two GMEs of $l_{\max} = 1$, each corresponding to one of two Slater exponents; period 2 elements use three GMEs of $l_{\max} = 2$; and period 3 elements use four GMEs of $l_{\max} = 3$. The relationship between this generalized auxiliary basis and the AO products is discussed in the next section and described completely in Ref. [32]. If the auxiliary basis exactly reproduced the AO products, then we'd expect it to match VEJ's results.

2.5 Generalization of the auxiliary basis

This section reviews a method that has only recently been reported in Ref. [32]. Our purpose for reviewing that work here is to unify its motivation with the second grand challenge: The limited auxiliary basis description of the response density.

DFTB2 explicitly treats the electrostatic interactions of the response density and is therefore a major advance beyond the traditional first-order tight-binding models whence it is based. Not only does the reintroduction of this missing physics improve the description of interatomic interactions, but it makes DFTB2 a more transferable and accurate predictor of atomic charges. The DFTB2 response density is represented in a basis of Slater monopoles, whose exponents are chosen to reproduce the atom's chemical hardness [60, 61]. The response density is made to interact with itself electrostatically, and thus, the DFTB2

response density self-energy is a quadratic function of atomic partial charges.

The electron affinity and ionization potentials of an atom suggest that the energy of an atom is not a pure-quadratic function of charge, but is instead asymmetrical. As electrons are removed from the atom, the chemical hardness increases; because the next electron to be removed is closer to the nucleus, its interaction with the nucleus is stronger and thus requires even greater energy to eject. This asymmetry might motivate one to try and improve DFTB2 by making the Slater exponent a function of charge, so that the Slater self-energies mimic an exponential-like or asymmetrical behavior. Invoking charge dependence in this way is an ad hoc correction. By this, we mean that the charge dependence is being used to alleviate a particular symptom without regard to the underlying problem that gives rise to that symptom. As such, there is no limitation on the functional form of the charge dependence; that is, one could make the Slater exponents a linear function of charge, or an exponential function of charge, or a sixth-order spline polynomial of charge, etc.

DFTB3 improves upon DFTB2 by making the Slater exponents a linear function of charge and then justifies this mathematical form by extending the Taylor series expansion of the Kohn-Sham potential energy to third order [56–58]. When reading those works [56–58], a reader may misconstrue the use of a third-order expansion to mean that the underlying problem giving rise to the quadratic behavior of the hardness is a premature truncation of the Taylor series to second order. Our previous works suggest that this is not the case [26, 32]. The pure-quadratic response self-energy of an isolated atom, as modeled by DFTB2, results only from having limited the auxiliary basis so completely that both the diffuse and tightly bound electron densities are represented by the same spatial function and thus contribute equally to the self-energy. One would instead intuitively expect the excess electrons to be more diffuse and therefore contribute less to the atom's self-energy than the more tightly bound electrons.

We attribute the source of the error to the incomplete AO and auxiliary basis being used and *not* the truncation of the expansion. We've come to this conclusion by repeatedly testing the following elementary logic [26, 32]: (1) If the source of error was the lack of a third-order term, then the error should continue to manifest itself when all other approximations of the model are removed. (2) If a second-order model is used, and the error ceases to manifest itself when some approximation (other than the expansion order) is removed, then the source of the error is not the second-order truncation. As a simple demonstration, Fig. 1 plots the hardness of oxygen as a function of charge using a variety of approximations. The VE model lacks third- and higher-order expansion terms, but it agrees extremely well

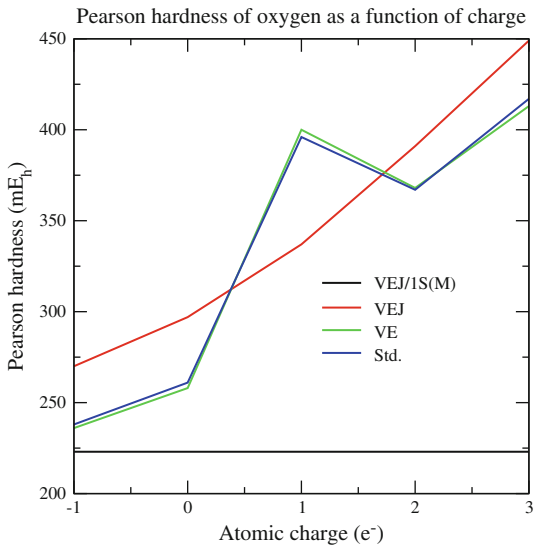

Fig. 1 The Pearson hardness of oxygen atom and ions for various models. Let $E(q)$ represent the energy of oxygen with an electric charge of q. We then compute the hardness as $[E(q-1) - E(q+1)]/2$. When the auxiliary basis is limited to a single, charge-independent Slater function [VEJ/1S(M)], the hardness is constant. The hardness becomes variable when more auxiliary functions are used (VEJ is equivalent to having used a complete auxiliary basis). Furthermore, upon including second-order LDA-xc (VE), the second-order model reproduces standard PBE/6-31G* (Std.) extremely well. By having removed the other approximations in the model and showing a strong agreement between the second-order Kohn-Sham expansion with its standard DFT counterpart, we conclude that the second-order expansion is not a premature truncation of the Kohn-Sham potential energy. If the lack of a third-order expansion term introduced errors of any consequence, then we'd expect those errors to continue to manifest themselves when the other approximations of the model are removed

with the standard PBE/6-31G* results; therefore, the lack of third- and higher-order expansion terms is not the source of error preventing an asymmetrical charge dependence.

From the above discussion, it should now be clear that an avenue for improving DFTB2, as an alternative to explicitly modeling a charge dependence or extending the order of the Taylor series expansion, would be to simply include more Slater functions: those which model the diffuse electrons, and those that model the inner-electrons. Furthermore, one could include higher-order multipoles, as this would: allow for an improved description of dipole polarizabilities, provided that the AO degrees of freedom were available; improve the strength and angular dependence of hydrogen bonds; and improve the nonbonded interactions, in general [62].

Of course, this is all rather easy to say, but how would you actually do it and still have it be practical? Although auxiliary basis sets have been used in traditional ab initio methods, the techniques they use typically require three-center integrals [63–69], whereas the efficiency of SCC-DFTB's technique is that it requires only two-center

integrals and thus can be precomputed and splined as a function of atom separation. Our approach was chosen to preserve this property, and we call it a generalization of the auxiliary basis because, like SCC-DFTB, we will use the auxiliary basis to directly model the AO products.

Some readers might insist that SCC-DFTB is modeling the atomic response densities, *not the individual AO products!* So, to convince those readers, and to more clearly demonstrate how our method is a generalization, let us insert Eq. 11 into (10)

$$\delta\rho(\mathbf{r}) = \sum_a \sum_{ij\in a} (P_{ij} - P_{ij}^{(0)}) \left[S_{ij}\varphi^a(\mathbf{r} - \mathbf{R}_a) \right] + \sum_{a\neq b} \sum_{i\in a} \sum_{j\in b} (P_{ij} - P_{ij}^{(0)}) \times \left[\frac{1}{2} S_{ij}\varphi^a(\mathbf{r} - \mathbf{R}_a) + \frac{1}{2} S_{ij}\varphi^b(\mathbf{r} - \mathbf{R}_b) \right]; \quad (33)$$

and now, upon comparing this to

$$\delta\rho(\mathbf{r}) = \sum_{ij} (P_{ij} - P_{ij}^{(0)}) \chi_i(\mathbf{r})\chi_j(\mathbf{r}); \quad (34)$$

we see that $\chi_i(\mathbf{r})\chi_j(\mathbf{r}) \approx S_{ij}\varphi^a(\mathbf{r} - \mathbf{R}_a)$, for one-center AO products; and

$$\chi_i(\mathbf{r})\chi_j(\mathbf{r}) \approx \frac{1}{2} S_{ij}\varphi^a(\mathbf{r} - \mathbf{R}_a) + \frac{1}{2} S_{ij}\varphi^b(\mathbf{r} - \mathbf{R}_b), \quad (35)$$

for two-center AO products. The factors of S_{ij} and $S_{ij}/2$ are the coefficients of a simple partitioning that "map" the AO product onto the atoms' auxiliary basis.

Our generalization is to simply use more auxiliary functions; that is,

$$\chi_i(\mathbf{r})\chi_j(\mathbf{r}) \approx \sum_{k\in a} M_{ijk}^a \varphi_k^a(\mathbf{r} - \mathbf{R}_a), \quad (36)$$

for one-center AO products; and

$$\chi_i(\mathbf{r})\chi_j(\mathbf{r}) \approx \sum_{k\in a} M_{ijk}^a \varphi_k^a(\mathbf{r} - \mathbf{R}_a) + \sum_{k\in b} M_{ijk}^b \varphi_k^b(\mathbf{r} - \mathbf{R}_b), \quad (37)$$

for two-center AO products, where $\mathbf{M}^a$ and $\mathbf{M}^b$ are matrices of mapping coefficients, which depend on the orientation and separation of the two atoms.

When the two atoms are aligned along the z-axis, and the angular dependence of the AO and auxiliary functions are described by spherical harmonics, many of the mapping coefficients vanish due to symmetry [20]. Therefore, one can spline the nonzero values of these matrices as a function of separation along the z-axis, interpolate their values for a given separation, and then use Wigner-D matrices to rotate those interpolated matrices into the molecular orientation [70].

The values of the nonzero matrix elements in the z-axis orientation are a *choice*. For example, we've described DFTB2 as using a Mulliken partitioning, but some have chosen to use CM3 charge mappings in their post-SCF analysis [71]. We have chosen to use is a constrained electrostatic fitting procedure [63, 67, 72–76]. The mathematical details of this procedure are fully presented in Ref. [32]; and one may choose differently; so we discuss this only briefly here. An electrostatic fitting procedure chooses the mapping coefficients in an attempt to reproduce the electrostatic potential of the AO product everywhere in space. It is a variational procedure that can be understood as minimizing the vector-norm squared difference between the AO product electric field with its auxiliary basis representation everywhere. By saying that we use a *constrained* electrostatic fitting procedure, we mean that Lagrange multipliers have been included within the variational solution to preserve the AO product multipole moments.

Just as DFTB2's Mulliken partitioning yielded partial atomic charges, the generalized auxiliary basis yields vectors of atomic charge moments, $\mathbf{m}^a : \delta\rho(\mathbf{r}) = -\sum_a \sum_{k\in a} m_k^a \varphi_k^a(\mathbf{r})$, where

$$m_k^a = -\sum_{ij\in a} (P_{ij} - P_{ij}^{(0)}) M_{ijk}^a - \sum_{i\in a} \sum_{j\notin a} (P_{ij} - P_{ij}^{(0)}) M_{ijk}^a - \sum_{i\notin a} \sum_{j\in a} (P_{ij} - P_{ij}^{(0)}) M_{ijk}^a. \quad (38)$$

The spin density can be represented in an analogous way: $\omega(\mathbf{r}) = -\sum_a \sum_{k\in a} s_k^a \varphi_k^a(\mathbf{r})$, where the auxiliary spin-charge moment of atom a is

$$s_k^a = -\sum_{ij\in a} (P_{ij}^\alpha - P_{ij}^\beta) M_{ijk}^a - \sum_{i\in a} \sum_{j\notin a} (P_{ij}^\alpha - P_{ij}^\beta) M_{ijk}^a - \sum_{i\notin a} \sum_{j\in a} (P_{ij}^\alpha - P_{ij}^\beta) M_{ijk}^a. \quad (39)$$

These auxiliary representations of the densities transform the $W[\delta\rho, \omega]$ functional into a function of auxiliary charge moments, spin-charge moments, and atomic positions; $W[\delta\rho, \omega] \rightarrow W(\mathbf{m}, \mathbf{s}, \mathbf{R})$; and the σ-spin Fock matrix generalizes to

$$F_{ij}^\sigma = T_{ij} + V_{ij}^{(1)} - \sum_a \sum_{k\in a} M_{ijk}^a \phi_k^{a,\sigma}, \quad (40)$$

for one-center products; and

$$F_{ij}^{\sigma} = T_{ij} + V_{ij}^{(1)} - \sum_{a}\sum_{k\in a} M_{ijk}^{a}\phi_{k}^{a,\sigma} - \sum_{b}\sum_{k\in b} M_{ijk}^{b}\phi_{k}^{b,\sigma}, \quad (41)$$

for two-center products; where

$$\phi_{k}^{a,\sigma} = \frac{\partial W(\mathbf{m},\mathbf{s},\mathbf{R})}{\partial m_{k}^{a}} + (-1)^{\delta_{\sigma,\beta}}\frac{\partial W(\mathbf{m},\mathbf{s},\mathbf{R})}{\partial s_{k}^{a}}, \quad (42)$$

is a spin-resolved auxiliary charge potential.

The VEJ/2P,3D,4F model computes the mapping coefficients as described above; its second-order energy is a Coulomb approximation; and the generalized auxiliary basis transforms Eq. 12 into

$$J(\mathbf{q},\mathbf{R}) = \frac{1}{2}\sum_{ab}\sum_{k\in a}\sum_{k'\in b} m_{k}^{a}m_{k'}^{b} \times \int\int \frac{\varphi_{k}^{a}(\mathbf{r})\varphi_{k'}^{b}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d^{3}r d^{3}r'. \quad (43)$$

2.6 Decomposition of the nonadditive xc-potentials

This section introduces new, preliminary methods that we've been using to gain insight into the decomposition of the xc-potentials. Our purpose for describing them here is to motivate and inspire the development of new approximations that address the third grand challenge: Moving beyond the one- and two-body treatment of the first-order matrix elements.

A topic on which we have yet to publish, and which is a topic that few people have ever attempted to address, is new approximations that go beyond the one- and two-center treatment of $\mathbf{V}^{(1)}$ and approximations that consider the many-body character of the second-order VE energy [77]. We believe that this is because it is not immediately apparent how these nonadditive potentials decompose into one-centered functions and two-centered corrections, or whether they can be adequately expressed in that manner at all. This brings us to the purpose of this section: we will determine how the nonadditivity of the first- and second-order potentials decompose into atomic contributions. Our motivation is to gain the insight necessary to propose new VE integral approximations that explicitly consider all the atoms in the system and to develop the new mathematical techniques that those proposals would require.

What new mathematical techniques potentially need to be developed, and how would that be aided by the present exploration? (1) If the first-order potential was an additive function, or approximated as such, then the contribution of a third-atom requires the potential of only that atom. (2) If one needs to explicitly treat the three-body nonadditivity, then one ultimately requires the potential of atom triplets. (3) If the first-order potential can be expressed as a sum of one-body potentials, *or a sum of one-body potentials and two-body corrections*, then there is some possibility of reducing the calculation of $\mathbf{V}^{(1)}$ to one that requires no more than two-center integrals, even if it requires a loop over third atoms. There are various mathematical tricks that would could try to use for this purpose; for example, using auxiliary basis representations of the AOs and/or first-order potential, or using resolution-of-the-identity techniques [69]. But the problem we are trying to emphasize here is a general lack of insight as to whether it is even possible to avoid many-body corrections [as in case (2)] or not; and if so, how?

Previous descriptions of DFTB2 [30] have attempted to rationalize the use of Eq. 9 by explaining how it differs from a superposition of atom-centered potentials

$$v_{\mathrm{ref}}^{(1)}(\mathbf{r}) \approx \sum_{a} v_{a}^{(1)}(\mathbf{r}). \quad (44)$$

Although Eq. 44 ignores the nonadditivity of the xc-potential, it at least recognizes that the system is composed of more than one or two atoms. Being that the only nonadditive component of $v_{\mathrm{ref}}^{(1)}(\mathbf{r})$ is the xc-potential, it begs to question whether or not the error introduced by ignoring nonadditivity in Eq. 44 is acceptable. It is for this reason that we construct the VE1S model and include it in our comparisons within Table 1.

The potential resulting from Eq. 44 is too negative relative to the exact potential; that is, $v_{\mathrm{ref}}^{(1)}(\mathbf{r}) - \sum_{a} v_{a}^{(1)}(\mathbf{r}) > 0$; however, a nonadditive multicenter function can be *cluster expanded* into additive single-center components and many-body corrections; for example,

$$v_{\mathrm{ref}}^{(1)}(\mathbf{r}) = \sum_{a} v_{a}^{(1)}(\mathbf{r}) + \sum_{b>a}\Delta v_{ab}^{(1)}(\mathbf{r}) + \sum_{c>b>a}\Delta v_{abc}^{(1)}(\mathbf{r}) + \cdots \quad (45)$$

where $v_{\mathrm{ref}}^{(1)}(\mathbf{r}) - \sum_{a} v_{a}^{(1)}(\mathbf{r})$ is the nonadditive behavior of $v_{\mathrm{ref}}^{(1)}(\mathbf{r})$, and

$$\Delta v_{ab}^{(1)}(\mathbf{r}) = v_{ab}^{(1)}(\mathbf{r}) - v_{a}^{(1)}(\mathbf{r}) - v_{b}^{(1)}(\mathbf{r}), \quad (46)$$

$$\Delta v_{abc}^{(1)}(\mathbf{r}) = v_{abc}^{(1)}(\mathbf{r}) - \Delta v_{ab}^{(1)}(\mathbf{r}) - \Delta v_{ac}^{(1)}(\mathbf{r}) - \Delta v_{bc}^{(1)}(\mathbf{r}) - v_{a}^{(1)}(\mathbf{r}) - v_{b}^{(1)}(\mathbf{r}) - v_{c}^{(1)}(\mathbf{r}), \quad (74)$$

are two- and three-body corrections, respectively. The two-body and larger-body corrections owe their existence solely to the nonadditivity of the xc-functional; therefore, in subsequent discussion, we refer to the nonadditivity of the first-order potential or the xc-potential interchangeably, depending on context.

Although the one-body sum (Eq. 44) is too negative (see e.g., Fig. 2), our experience is that the two-body

Table 1 Bond, angle, and dipole moment errors relative to standard PBE/6-31G* results for 52 molecules upon geometry optimization

	Bond		Angle		Dipole		ΔE_{rxn}	
N	114		98		52		6	
Avg.	1.322 Å		112.546°		0.496 a.u.		158.637 kcal/mol	
Model	μ_{SE}	μ_{UE}	μ_{SE}	μ_{UE}	μ_{SE}	μ_{UE}	μ_{SE}	μ_{UE}
VE	0.002	0.002	−0.006	0.177	−0.007	0.012	0.076	0.221
VE0	−0.013	0.014	−0.629	0.858	−0.005	0.015	−8.408	11.708
VE1	−0.473	0.473	0.785	10.666	0.120	0.373	369.505	511.706
VE1S	−0.128	>0.130	−1.497	>6.070	0.030	>0.185	−79.629	>130.563
VE1W	−0.008	0.011	−0.454	0.827	−0.006	0.013	−0.166	11.621
VEJ	0.012	0.013	−0.198	0.564	−0.045	0.052	−1.161	1.615
VEJ/1S(M)	−0.045	0.046	0.384	1.148	−0.207	0.257	−15.165	15.165
VEJ/2P,3D,4F	0.013	0.014	−0.258	0.577	−0.039	0.053	0.045	1.337

The statistics under the heading $\Delta E_{rxn}$are adiabatic SCF energy differences corresponding to the homolytic bond dissociation reactions: $C_2H_6 + H_2 \rightarrow 2CH_4$, $C_2H_4 + 2H_2 \rightarrow 2CH_4$, $C_2H_2 + 3H_2 \rightarrow 2CH_4$, $N_2H_4 + H_2 \rightarrow 2NH_3$, $Si_2H_6 + H_2 \rightarrow 2SiH_4$, and $H_2O_2 + H_2 \rightarrow 2H_2O$. μ_{SE} and μ_{UE} are the signed and unsigned errors of the property, respectively, and whose units are shown in the row "Avg.", which displays the average reference value. "N" is the amount of data included in the statistics. The VE1S statistics exclude any contributions from Si_2H_6, which geometry optimizes into a hopelessly nonsensical configuration where all 6 hydrogens are clustered between the two silicons, at which point we can no longer coax SCF convergence. Thus, we expect the true VE1S errors to be larger than those reported here

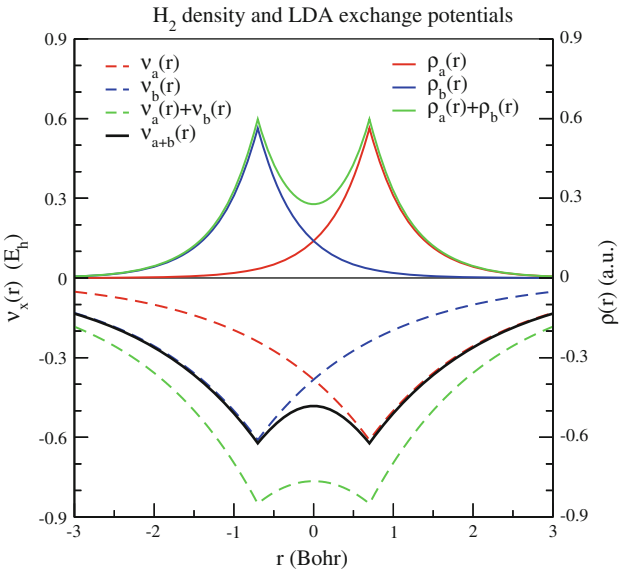

Fig. 2 Example of exchange-potential nonadditivity using the LDA exchange functional in molecular hydrogen. The hydrogen atom density is the quantum mechanically exact density, and the atoms are separated by 1.4011 Bohr. The x axis is the position along the internuclear axis. The sum of the isolated atom exchange potentials (*dashed green line*) is more negative than the molecular exchange potential (*solid black line*). At the bond midpoint ($r = 0$), the ratio of the summed potentials to the molecular potential is exactly $2{:}2^{1/3}$

corrections over-correct it (see e.g., Fig. 3), and the resulting potential is arguably worse than the simple sum; however, we have empirically found that the third- and higher-order many-body corrections act to cancel some of the two-body corrections. Specifically, the many-body corrections cancel those two-body corrections for the pairs of atoms whose contribution to the reference density in the region of their overlap is dwarfed by a larger or intervening third atom. In other words, we have found that one can truncate the cluster expansion and weight the two-body corrections by some factor $0 < W_{ab} < 1$ that accounts for the many-body corrections in an effective way. We call this the *weighted cluster expansion*

$$v_{\text{ref}}^{(1)}(\mathbf{r}) \approx \sum_a v_a^{(1)}(\mathbf{r}) + \sum_{b>a} W_{ab}\Delta v_{ab}^{(1)}(\mathbf{r}), \tag{48}$$

and this is the primary approximation in the VE1W model.

We have explored many forms of W_{ab}; the simplest of these is $W_{ab} = S_f(X_{ab} + X_{ba})$, where $S_f(x) = x$, if $x < x_0$; $S_f(x) = 1$, if $x > 1$;

$$S_f(x) = x_0 + (1 - x_0)\sum_{i=1}^{5} c_i \left(\frac{x - x_0}{1 - x_0}\right)^i, \tag{49}$$

if $x_0 < x < 1$; $x_0 = 19/20$; $c_1 = 1$; $c_2 = 0$; $c_3 = 4$; $c_4 = -7$; $c_5 = 3$; $X_{ab} = S_{ab}^{\text{ref}}/\sum_c S_{cb}^{\text{ref}}$; $S_{aa}^{\text{ref}} = 0$; and $S_{ab}^{\text{ref}} = \int \rho_a(\mathbf{r})\rho_b(\mathbf{r})d^3r$. We don't want the reader to become overwhelmed by technical details that are presented for the sole purpose of ensuring the reproducibility of our results. Our discussion can be understood by interpreting the math as being a model that results in $W_{ab} \approx 1$ for those pairs of atoms that one would intuitively consider to be covalently bonded, and $W_{ab} \approx 0$ for those that are not.

The reader may question why the xc-potential is nonadditive, and why the simple sum of potentials (Eq. 44) leads to a net potential that is too negative, and why the two-body corrections over-correct the one-body sum. In brief, the xc-functional's nonadditivity results from a

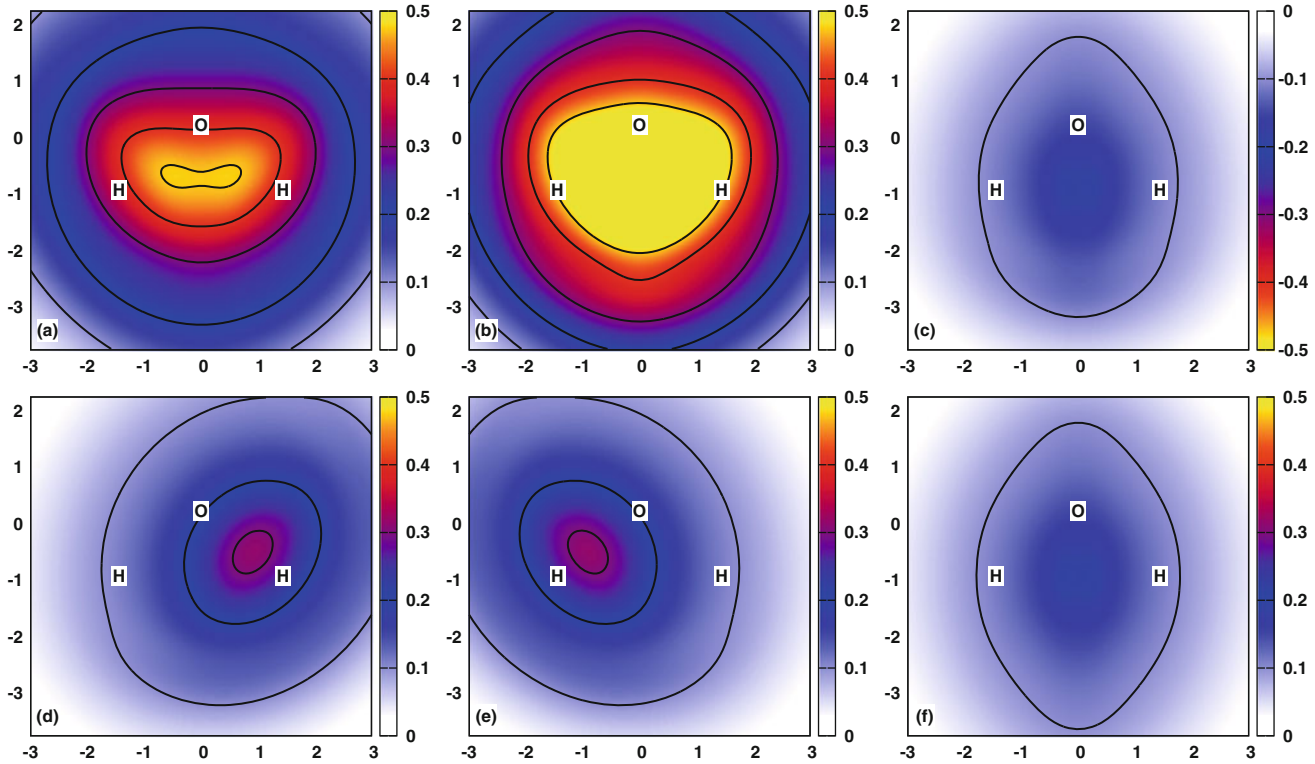

Fig. 3 Decomposition of water's PBE first-order potential $v_{\mathrm{ref}}^{(1)}(\mathbf{r})$. The nonadditive potential is depicted in **a**, which is the sum of the two- and three-body corrections shown in **b** and **c**, respectively. The two-body correction (**b**) is the sum of individual pair-corrections $\Delta v_{ab}^{(2)}(\mathbf{r})$depicted in **d**, **e**, and **f**. The colors represent the value of the potential along the plane formed by the atoms. The x and y axis are expressed in Bohr, and the *color scale* is in E_h. Note that potential in **c** is negative, whereas all other potentials are positive. The *color scheme* emphasizes that the magnitude of the three-body correction is so similar to the H–H pair-correction that these potentials cancel. As a result, the nonadditivity shown in **a** is well reproduced by the sum of **d** and **e**. The solid lines are isocontour values of 0.1, 0.2, 0.3, 0.4, and 0.47 E_h

dependence on the density to some *fractional order*, and the reason why the one-body sum is too negative is because that fractional order is *between one and two*. Let us clarify through a brief demonstration. Consider the uniform electron gas exchange functional $E_x[\rho] = -C_x \int \rho(\mathbf{r})^{4/3} d^3r$, where $C_x = \frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3}$. The first-order exchange potential is

$$v_x(\mathbf{r}) = -\frac{4C_x}{3}\rho(\mathbf{r})^{1/3}. \tag{50}$$

Let us evaluate $v_x(\mathbf{r})$ with a homonuclear dimer $\rho(\mathbf{r}) = \rho_a(\mathbf{r} - \mathbf{R}) + \rho_b(\mathbf{r} + \mathbf{R})$. At the bond midpoint $\mathbf{r} = 0$, it holds that $\rho(0) = \rho_a(-\mathbf{R}) + \rho_b(\mathbf{R}) = 2\rho_a(\mathbf{R})$ [since $\rho_a(\mathbf{r}) = \rho_b(\mathbf{r})$ and each are spherically symmetric], and the exchange potential is

$$v_{x,a+b}(0) = -2^{1/3}\left[\frac{4C_x}{3}\rho_a(\mathbf{R})^{1/3}\right], \tag{51}$$

whereas the sum of atom potentials is

$$v_{x,a}(0) + v_{x,b}(0) = -2\left[\frac{4C_x}{3}\rho_a(\mathbf{R})^{1/3}\right]. \tag{52}$$

We see that the sum of atom potentials is more negative by a ratio of $2{:}2^{1/3}$. With this insight, it is easy to explain why the two-body corrections (Eq. 46) over-correct the one-body sum (Eq. 44). Just as the sum of one-body potentials (Eq. 44) results in a net potential that is too negative, the subtraction of one-body potentials in Eq. 46 results in a two-body correction that is too positive.

At this point, we have completed our description of the approximations as they relate to the calculation of the core-Hamiltonian and to those models discussed in the Results section; however, we have a few comments regarding the second-order potentials. The observations that we describe here are of interest to the community, as they can be used to argue the validity of the approximations employed in spin-polarized SCC-DFTB models [38, 78], and also suggest ways that those approximations can be improved.

Unlike the first-order potential, the second-order potentials, $v_{\mathrm{ref}}^{(2)}(\mathbf{r})$ and $w_{\mathrm{ref}}^{(2)}(\mathbf{r})$, are not well behaved; they are strictly negative functions that approach $-\infty$ as $\rho_{\mathrm{ref}}(\mathbf{r}) \to 0$. Thus, $w_{\mathrm{ref}}^{(2)}(\mathbf{r})$ is poorly modeled by a simple sum of one-body $w_a^{(2)}(\mathbf{r})$'s. Furthermore, the many-body

cluster corrections introduce as many large errors as they remove, and the expansion does not converge until it has been totally exhausted; however, we have found the following approximation to be almost exactly correct:

$$w^{(2)}_{\mathrm{ref}}(\mathbf{r}) \approx \sum_a L_a(\mathbf{r}) w^{(2)}_a(\mathbf{r}) + \sum_{b>a} W_{ab}\Big\{[L_a(\mathbf{r}) + L_b(\mathbf{r})] w^{(2)}_{ab}(\mathbf{r}) - L_a(\mathbf{r}) w^{(2)}_a(\mathbf{r}) - L_b(\mathbf{r}) w^{(2)}_b(\mathbf{r})\Big\}, \quad (53)$$

and similarly for $v^{(2)}_{\mathrm{ref}}(\mathbf{r})$, where $L_a(\mathbf{r}) = \rho_a(\mathbf{r})^2 \,/\, \sum_b \rho_b(\mathbf{r})^2$ is a spatial partition function, which one can interpret as being a "fuzzy" Voronoi-like cell. It holds the properties: $L_a(\mathbf{r}) \approx 1$ in the region of a; $L_a(\mathbf{r}) \approx 0$ near the region of any other atom; $\sum_a L_a(\mathbf{r}) = 1$ for all $\mathbf{r}$; and as a consequence of all the above, $L_a(\mathbf{r}) L_{b \neq a}(\mathbf{r}) \approx 0$. The sum of partitioned one-body potentials in the *partitioned cluster expansion* is a much-improved representation of the second-order potentials (see Fig. 4), and the weighted two-body partitioned corrections are relatively small.

If one approximates $w^{(2)}_{\mathrm{ref}}(\mathbf{r})$ by ignoring the two-body corrections, then the potential is described by the isolated atom potentials within their respective Voronoi cells, and the spin-polarization energy $E[\omega]$ becomes

$$\begin{aligned} E[\omega] &= \frac{1}{2}\int \omega(\mathbf{r})^2 w^{(2)}_{\mathrm{ref}}(\mathbf{r}) d^3r \\ &= \sum_a \frac{1}{2}\int \omega(\mathbf{r})^2 \Big[L_a(\mathbf{r}) w^{(2)}_a(\mathbf{r})\Big] d^3r \\ &= \sum_a \frac{1}{2}\int \omega(\mathbf{r}) \omega_a(\mathbf{r}) w^{(2)}_a(\mathbf{r}) d^3r \end{aligned} \quad (54)$$

where $\omega_a(\mathbf{r}) = L_a(\mathbf{r})\omega(\mathbf{r})$ is the atom-partitioned spin density about a. If $\omega(\mathbf{r})$ was represented by an atom-centered auxiliary basis and was chosen to partition the density in a manner analogous to $L_a(\mathbf{r})$, then the spin-polarization energy reduces to one- and two-center overlap-like integrals between auxiliary functions.

A different result is obtained if one employs the exact decomposition $\omega(\mathbf{r}) = \sum_b \omega_b(\mathbf{r}) = \sum_b L_b(\mathbf{r})\omega(\mathbf{r})$, and then approximates their overlap $L_a(\mathbf{r}) L_{b \neq a}(\mathbf{r}) \approx 0$; that is,

$$\begin{aligned} E[\omega] &= \sum_{ab} \frac{1}{2}\int \omega(\mathbf{r})^2 L_b(\mathbf{r}) \Big[L_a(\mathbf{r}) w^{(2)}_a(\mathbf{r})\Big] d^3r \\ &\approx \sum_a \frac{1}{2}\int \omega_a(\mathbf{r})^2 w^{(2)}_a(\mathbf{r}) d^3r \\ &\approx \sum_a E[\omega_a]. \end{aligned} \quad (55)$$

If one presumed that the $L_a(\mathbf{r})$'s are spherical, then Eq. 55 is no different than what is used in spin-polarized SCC-DFTB [38]; however, we feel that we've shown that the arguments needed to reach this result are more justified than what one might infer from reading Refs. [38] and [79]; and it also suggests ways that one can try to improve upon their work.

2.7 Computational details

2.7.1 *DFTB2 with charge-dependent polarizability corrections*

Figure 5 compares the polarizabilities of DFTB2, DFTB2+CPEQ, and DFTB2+CPE0 to standard B3LYP/6-31++G** results. The comparisons include 33 anionic, 142 neutral, and 35 cationic molecules taken from the QCRNA database [80]. The molecules were selected for having contained some combination of H, C, N, and O, but no other elements.

The H, C, N, and O DFTB2 parameters are described in Ref. [22], and these parameters remain unmodified in our CPE-corrected variations, which we henceforth refer to without the "DFTB2+" prefix. The CPEQ and CPE0 models have four additional atomic parameters: $\zeta_a(0), b_a, R_{\mathrm{lo},a}$, and $R_{\mathrm{hi},a}$. Although note that the CPE0 model's b_a parameter is zero for all atoms, these parameters were optimized for each model to B3LYP/6-311++G** dipole moments and B3LYP/6-31++G** isotropic polarizabilities, and the merit function used to judge the goodness of the fit is

$$\chi^2 = \sum_{i=1}^{142} 80|\Delta\mathbf{q}^{(1)}|_i + \sum_{i=1}^{210} |\Delta\alpha|_i; \quad (56)$$

where $|\Delta\mathbf{q}^{(1)}|$is the norm of the difference between the model and reference dipole moment vectors; $|\Delta\alpha|$is the absolute difference between the model and reference isotropic polarizabilities; and the summation over dipole moment errors includes only neutral molecules. Both the reference and model calculations are evaluated at the B3LYP/6-31++G** geometry-optimized structures.

2.7.2 *Generalization of the auxiliary basis*

Table 1 compares the geometry-optimized bond, angle, dipole moment, and relative energy errors of various PBE/6-31G*-based models to standard PBE/6-31G*. The statistics include results from 52 molecules [81] taken from the G2/97 neutral small molecule test set [82], which were also used in our previous VE studies [26, 32]. For the purpose of discussing the generalization of the auxiliary basis, the reader should compare VEJ/1S(M) and VEJ/2P,3D,4F to VEJ, and note how well VEJ compares to VE. The parametrization of the VEJ/2P,3D,4F auxiliary basis Slater exponents is described elsewhere [32].

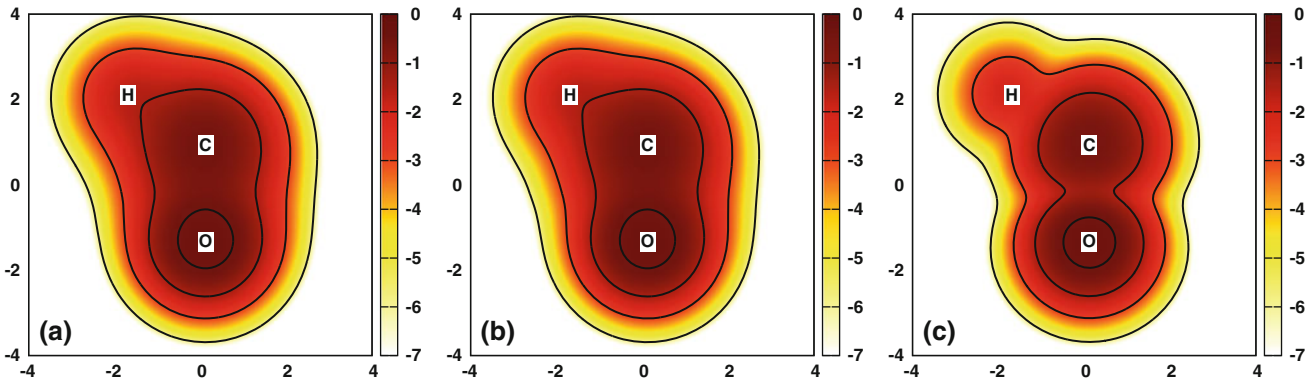

Fig. 4 Decomposition of the SPW92 second-order spin-polarization potential $w_{\mathrm{ref}}^{(2)}(\mathbf{r})$ for the HCO molecule. The panels depict: **a** the potential, without approximation; **b** the one- and two-body partitioned cluster approximation (Eq. 53); and **c** the one-body partitioned cluster approximation; that is, $W_{ab} = 0$ for all pairs. The *colors* represent the value of these potentials in the plane formed by the atoms. The x and y axis are in units of Bohr, and the *color scale* is E_h. The *solid lines* are isocontour values of -0.7, -1.5, -3, and -6 E_h

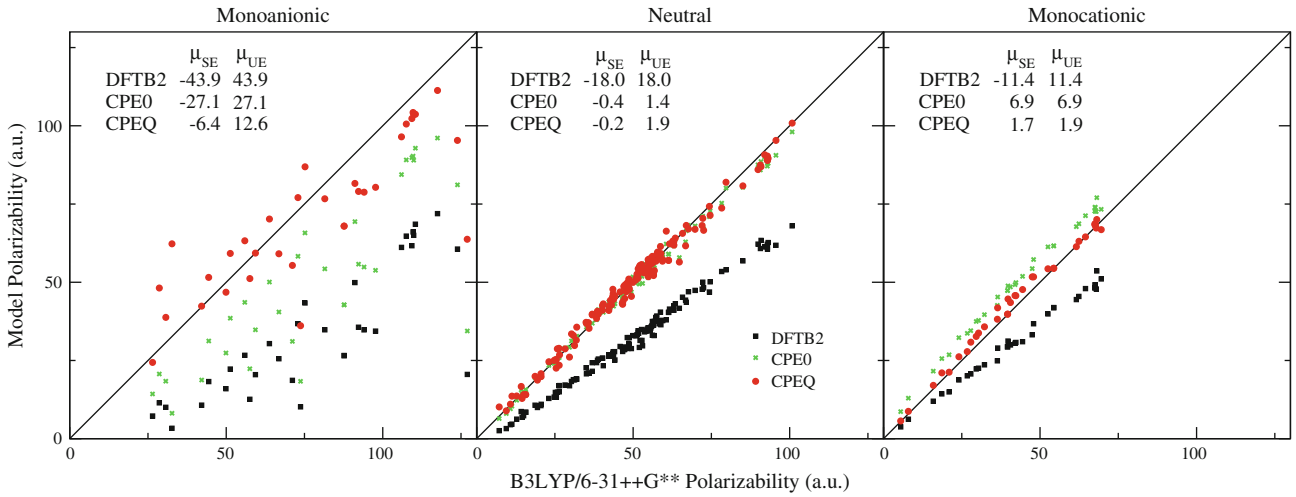

Fig. 5 Comparison of DFTB2, DFTB2+CPE0, and DFTB2+CPEQ polarizabilities to standard B3LYP/6-31++G** results. μ_{SE} and μ_{UE} are the mean signed and unsigned errors (a.u.), respectively

2.7.3 Decomposition and nonadditivity of the xc-potentials

Table 1 is used to compare the errors of the VE1, VE1S, and VE1W models, all of which are variations of the VE0 model.

Figure 3 displays the nonadditivity of the PBE first-order potential of water and its decomposition into two- and three-body corrections and the decomposition of the two-body correction into the individual pair-corrections.

3 Results

3.1 DFTB2 with charge-dependent polarizability corrections

Figure 5 shows that DFTB2 underpredicts the B3LYP/6-31++G** polarizabilities; the extent to which depends on the ionic character of molecules. The mean unsigned relative percent errors of the anionic, neutral, and cationic molecules are 59, 38, and 26%, respectively. In other words, the anionic molecules need a larger polarizability correction than the neutrals, and neutrals need a larger correction than the cations.

To emphasize this further, we examine the results of a charge-independent model, CPE0. It greatly improves the neutral molecule polarizabilities, but the resulting cations are too polarizable, and the anions are still not polarizable enough.

However, the errors are largely corrected with the charge-dependent model, CPEQ. It reduces the neutral and cationic polarizability errors to 5%; and the anionic polarizability errors are two and three times smaller than the CPE0 and DFTB2 errors, respectively.

The DFTB2 mean unsigned dipole moment errors relative to the B3LYP reference is 0.175 a.u., and the CPE0

and CPEQ models reduce this to 0.167 and 0.166 a.u., respectively. As expected, the CPE0 and CPEQ models do not significantly improve the DFTB2 dipole moments, because the models have been chosen to screen the short-range interactions. The CPE method lacks the proper physics to describe covalent bonds and thus cannot be relied upon to improve molecular dipole moments. For this purpose, it would be more fruitful to improve the underlying DFTB2 model, for example, by increasing the number of AOs.

We note that there is very little difference between the DFTB2+CPEQ method of the present work and the MNDO/d+CPE method presented in Ref. [37]. The main difference is that we have included Fock matrix corrections (Eq. 28) in the present work. There are some more trivial differences as well, such as our use of Slater dipole CPE response functions in DFTB2+CPEQ, as opposed to the Gaussian-dipole functions used in MNDO/d+CPE. The only other difference is: the reference density about which the DFTB2+CPEQ correction is expanded is a collection of Slater charges, whereas the MNDO/d+CPE correction is expanded about a collection of point charges. The beauty of the CPE correction is that it is expanded about a reference density, and it doesn't matter what underlying method gave rise to that reference density. However, just as the CPE correction is an ad hoc correction to MNDO/d, so too is it an ad hoc correction to DFTB2.

3.2 Generalization of the auxiliary basis

Table 1 demonstrates that a Taylor series truncation of the xc-functional to first order (VEJ) has a small effect on geometries and a modest effect on dipole moments. Furthermore, it is capable of closely reproducing closed-shell molecule relative energies. Some new models have extended the Taylor series expansion to third order as a mechanism for including charge dependence [56–58], but the VEJ results presented here should clarify that one should take care to not misinterpret those models as having corrected a symptom originating from a premature truncation of the Taylor series expansion.

Approximating the second-order energy with a SCC-DFTB-like Slater monopole auxiliary basis [VEJ/1S(M)] significantly increases the errors, but these errors are alleviated when the auxiliary basis is extended to better reproduce the AO products (VEJ/2P,3D,4F).

Our previous work [32] showed that VEJ/1S(M) underpredicts the hydrogen bond strength of the water dimer by approximately 4 mE_h relative to both VEJ and standard PBE/6-31G*, and that this error was corrected when using a more sophisticated auxiliary basis.

3.3 Decomposition and nonadditivity of the xc-potentials

The introduction to this manuscript emphasized the severity of the VE1 approximation, and we show its results in Table 1. Its severity may seem intuitively obvious because it neglects all atoms in the system other than those coincident with the AO product, and so one may try to improve it by including the potentials caused by the other atoms. The VE1S model attempts to overcome this severe approximation by (incorrectly) assuming that the neutral-atom potentials are additive functions; however, this model is arguably just as bad as VE1. One can try to construe the numerical values in a manner to suggest that VE1S is a minor improvement, but both models optimize the molecular geometries into utterly nonphysical configurations.

The reader may notice that one of the molecules (Si_2H_6) fails to SCF converge once VE1S optimizes it into an unrealistic geometry, but it is best not to over-interpret this. The geometries of VE1 and VE1S are both extremely poor, and the fact that one has SCF converge difficulties when evaluated at a nonsensical configuration, whereas the other does not, does not indicate that its approximations are somehow better. Had we expanded the test set to include even larger molecules, we'd expect the VE1 model to eventually fail as well. We have tried a number of approximations, most of which we do not describe here; our experience suggests that larger molecules are more likely to optimize into strange configurations and then fail to SCF converge, because the errors introduced by the approximations increase with the system size. It is because of this that we feel our limited test set is sufficient to identify which approximations are severe without reaching that extreme limit of failure.

The VE1S first-order potential is too negative because it incorrectly assumed that the neutral-atom xc-potentials were additive, and as the size of the system increases, the magnitude of this error is compounded. To account for this nonadditivity, one might include two-body corrections. Alas, the resulting potential is then not negative enough, and the magnitude of its error increases with the size of the system. Like VE1S, it readily geometry optimizes to nonsensical configurations.

The prospect of explicitly including three-body or larger-body corrections is sufficiently daunting that it may be tempting to just give up and use VE1 or the VE1-like approximation that most semiempirical models use. To make an advance beyond that, we need to have some way to account for the many-body effects in some new way. To this end, Fig. 3 decomposes the xc-potential of water into its two- and three-body components, and we see that the three-body nonadditivity cancels the two-body correction

between the H's. One can perform similar decompositions for other molecules, and from these, it seems that this is true for all pairs of atoms whose overlap is dwarfed by the densities of the other atoms. This general idea can inspire many different forms of the weighted cluster approximation, and Table 1 displays the statistics for a very simple model (VE1W). This model reduces the errors to those comparable to VE0, the model on which it is based.

4 Conclusion

Kohn-Sham expansion methods can be made to be as accurate and transferable as the density-functional methods on which they are based. The primary sources of errors in a second-order VE model are *not* a consequence of a premature truncation of the Kohn-Sham functional. Instead, the source of errors is much more obvious; in that, just as one would expect standard DFT models to suffer from approximations, so too would one expect the analogous VE models. The most severe VE approximations that we've encountered are: the limited size of the AO basis; the poor auxiliary basis representation of the AO products; and the one- and two-body treatment of the first-order interactions.

As a mechanism for discussing these approximations: we've introduced a new self-consistent *charge-dependent* dipole polarization correction model (DFTB2+CPEQ) that improves DFTB2 polarizabilities; we've reviewed our recently published work that generalizes the auxiliary basis; and we're examined the decomposition of the first-order potential into many-body components, and used these observations to gain the insight and motivation for the development of new approximations.

From these discussions, we hope to have communicated the following:

(1) The errors introduced from having truncated the Taylor series expansion to second order are small relative to the other errors that we've discussed.
(2) The beneficial effects of third-order models arise from introducing charge-dependent corrections, and these corrections mask the limited size of the AO basis.
(3) Like all minimal valence AO basis models, DFTB2 underpredicts dipole polarizabilities; however, this underprediction will not be alleviated by increasing the size of the AO basis alone, because it would still lack atomic dipole response. In lieu of generating an entirely new DFTB model, one can improve the polarizabilities with an auxiliary CPE response. If the CPE response is not made charge-dependent, then the anionic and cationic molecular polarizabilities will be under- and over-predicted, respectively; but the polarizability of various charge states can be improved by making the CPE response functions charge-dependent.
(4) The VE response density can be modeled more accurately by increasing the radial and angular completeness of the auxiliary basis description of the AO products, and this can be done in a way that requires only two-center integrals, which can be precomputed and splined as a function of internuclear separation.
(5) The one- and two-body treatment of the first-order matrix elements is a severe approximation; however, it is not improved by representing the first-order potential as a sum of isolated atom potentials. Nor is it improved by modeling it with a sum of one-body potentials and two-body corrections. We have presented some initial evidence that the many-body nonadditivity of the xc-potential can be modeled by damping some of the two-body corrections.
(6) The second-order potentials are not well described by *any* truncated cluster approximation, but is well modeled by a partitioned cluster approximation. By inserting this approximate form into the expression for the spin-polarization energy, one can derive its model form used in spin-polarized DFTB2 methods by making some additional approximations.

It is the hope that the insights provided in this work will be instrumental in directing further research effort toward development of the next-generation of semiempirical quantum models for broad application.

Acknowledgments The authors are grateful for financial support provided by the National Institutes of Health (GM084149). Computational resources from the Minnesota Supercomputing Institute for Advanced Computational Research (MSI) were utilized in this work. This research was supported in part by the National Science Foundation through TeraGrid resources provided by the National Center for Supercomputing Applications and the Texas Advanced Computing Center under grant TG-CHE100072.

References

1. Hückel E (1931) Z Phys 70:204
2. Hückel E (1931) Z Phys 72:310
3. Hückel E (1932) Z Phys 76:628
4. Hückel E (1933) Z Phys 83:632
5. Pariser R, Parr RG (1953) J Chem Phys 21:767
6. Hoffmann R (1963) J Chem Phys 39:1497
7. Pople JA, Segal GA (1966) J Chem Phys 44:3289
8. Pople JA, Beveridge DL, Dobosh PA (1967) J Chem Phys 47:2026
9. Baird NC, Dewar MJS (1969) J Chem Phys 50:1262
10. Bingham RC, Dewar MJS, Lo DH (1975) J Am Chem Soc 97:1285
11. Dewar MJS, Thiel W (1977) Theor Chim Acta 46:89
12. Thiel W, Voityuk AA (1996) Theor Chim Acta 93:315
13. Dewar MJS, Zoebisch E, Healy EF, Stewart JJP (1985) J Am Chem Soc 107:3902

14. Stewart JJP (1989) J Comput Chem 10:221
15. Stewart JJP (2007) J Mol Model 13:1173
16. Clark T (2000) J Mol Struct (Theochem) 530:1
17. Winget P, Selçuki C, Horn A, Martin B, Clark T (2003) Theor Chem Acc 110:254
18. Winget P, Clark T (2005) J Mol Model 11:439
19. Rocha GB, Freire RO, Simas AM, P Stewart JJ (2006) J Comput Chem 27:1101
20. Slater JC, Koster GF (1954) Phys Rev 94:1498
21. Porezag D, Frauenheim T, Köhler T, Seifert G, Kaschner R (1995) Phys Rev B 51:12947
22. Elstner M, Porezag D, Jungnickel G, Elsner J, Haugk M, Frauenheim T, Suhai S, Seifert G (1998) Phys Rev B 58:7260
23. Tuttle T, Thiel W (2008) Phys Chem Chem Phys 10:2125
24. Kolb M, Thiel W (1993) J Comput Chem 14:775
25. Weber W, Thiel W (2000) Theor Chem Acc 103:495
26. Giese TJ, York DM (2010) J Chem Phys 133:244107
27. Perdew JP, Burke K, Ernzerhof M (1996) Phys Rev Lett 77:3865
28. Elstner M (2007) J Phys Chem A 111:5614
29. Seifert G (2007) J Phys Chem A 111:5609
30. Frauenheim T, Seifert G, Elstner M, Hajnal Z, Jungnickel G, Porezag D, Suhai S, Scholz R (2000) Phys Status Solidi B 217:41
31. Otte N, Scholten M, Thiel W (2007) J Phys Chem A 111:5751
32. Giese TJ, York DM (2011) J Chem Phys 134:194103
33. Giese TJ, York DM (2008) J Chem Phys 128:064104
34. Giese TJ, York DM (2008) J Chem Phys 129:016102
35. Giese TJ, York DM (2008) J Comput Chem 29:1895
36. Giese TJ, York DM (2007) J Chem Phys 127:194101
37. Giese TJ, York DM (2005) J Chem Phys 123:164108
38. Köhler C, Seifert G, Gerstmann U, Elstner M, Overhof H, Frauenheim T (2001) Phys Chem Chem Phys 3:5109
39. Kohn W, Sham L (1965) Phys Rev A 140:A1133
40. Frauenheim T, Seifert G, Elstner M, Niehaus T, Köhler C, Amkreutz M, Sternberg M, Hajnal Z, Di Carlo A, Suhai S (2002) J Phys Condens Matter 14:3015
41. Elstner M, Frauenheim T, Kaxiras E, Seifert G, Suhai S (2000) Phys Status Solidi B 217:357
42. Murdachaew G, Mundy CJ, Schenter GK (2010) J Chem Phys 132:164102
43. Maerzke KA, Murdachaew G, Mundy CJ, Schenter GK, Siepmann JI (2009) J Phys Chem A 113:2075
44. Chang DT, Schenter GK, Garrett BC (2008) J Chem Phys 128:164111
45. Matsuzawa N, Dixon DA (1992) J Phys Chem 96:6232
46. Fiedler L, Gao J, Truhlar DG (2011) J Chem Theory Comput 7:852
47. York DM, Yang W (1996) J Chem Phys 104:159
48. Nalewajski RF (1984) J Am Chem Soc 106:944
49. Mortier WJ, Van Genechten K, Gasteiger J (1985) J Am Chem Soc 107:829
50. Mortier WJ, Ghosh SK, Shankar S (1986) J Am Chem Soc 108:4315
51. Morales J, Martínez TJ (2001) J Phys Chem A 105:2842
52. Itskowitz P, Berkowitz ML (1997) J Phys Chem A 101:5687
53. York DM (1995) Int J Quantum Chem 56:385
54. Zhou B, Ligneres VL, Carter EA (2005) J Chem Phys 122:044103
55. Hobson EW (1892) Proc Lond Math Soc 24:55
56. Yang Y, Yu H, York DM, Cui Q, Elstner M (2007) J Phys Chem A 111:10861
57. Gaus M, Cui Q, Elstner M (2011) J Chem Theory Comput 7:931
58. Yang Y, Yu H, York D, Elstner M, Qiang C (2008) J Chem Theory Comput 4:2067
59. Perdew JP, Wang Y (1992) Phys Rev B 45:13244
60. Parr RG, Yang W (1984) J Am Chem Soc 106:4049
61. Pearson RG (1988) J Am Chem Soc 110:7684
62. Paxton AT, Kohanoff JJ (2011) J Chem Phys 134:044130
63. Dunlap BI (2000) J Mol Struct (Theochem) 529:37
64. Glaesemann KR, Gordon MS (2000) J Chem Phys 112:10728
65. Hamel S, Casida ME, Salahub DR (2001) J Chem Phys 114:7342
66. Ahlrichs R (2004) Phys Chem Chem Phys 6:5119
67. Sodt A, Subotnik JE, Head-Gordon M (2006) J Chem Phys 125:194109
68. Pedersen TB, Aquilante F, Lindh R (2009) Theor Chem Acc 124:1
69. Hohenstein EG, Sherrill CD (2010) J Chem Phys 132:184111. doi:10.1063/1.3426316
70. Choi CH, Ivanic J, Gordon MS, Ruedenberg K (1999) J Chem Phys 111:8825
71. Kalinowski JA, Lesyng B, Thompson JD, Cramer CJ, Truhlar DG (2004) J Phys Chem A 108:2545
72. Dunlap BI, Rösch N, Trickey SB (2010) Mol Phys 108:3167
73. Jung Y, Sodt A, Gill PW, Head-Gordon M (2005) Proc Natl Acad Sci 102:6692
74. Piquemal J, Cisneros G, Reinhardt P, Gresh N, Darden TA (2006) J Chem Phys 124:104101
75. Cisneros GA, Piquemal J, Darden TA (2006) J Chem Phys 125:184101
76. Elking DM, Cisneros GA, Piquemal J, Darden TA, Pedersen LG (2010) J Chem Theory Comput 6:190
77. Tu Y, Jacobsson SP, Laaksonen A (2006) Phys Rev B 74:205104
78. Zheng G, Witek HA, Bobadova-Parvanova P, Irle S, Musaev DG, Prabhakar R, Morokuma K (2007) J Chem Theory Comput 3:1349
79. Frauenheim T, Seifert G, Elstner M, Hajnal Z, Jungnickel G, Porezag D, Suhai S, Scholz R (2000) Phys Stat Sol 217:41
80. Giese TJ, Gregersen BA, Liu Y, Nam K, Mayaan E, Moser A, Range K, Nieto Faza O, Silva Lopez C, Rodriguezde Lera A, Schaftenaar G, Lopez X, Lee T, Karypis G, York DM (2006) J Mol Graph Model 25:423
81. The molecules in the test set are: BeH, C_2H_2, C_2H_4, C_2H_6, CH_2, CH_3, CH_4, CH_4O, CH_4S, CH_3Cl, CN, CS, HC, HCN, HCO, HF, HCl, LiH, NH, HO, H_2O, H_2O_2, OCl, NO, OS, O_2, CO, SiO, CO_2, SO_2, F_2, Cl_2, FCl, Li_2, LiF, Na_2, NaCl, N_2, NH_2, NH_3, N_2H_4, HOCl, H_2CO, P_2, PH_2, PH_3, S_2, H_2S, SiH_2, SiH_3, SiH_4, and Si_2H_6
82. Curtiss LA, Raghavachari K, Redfern PC, Pople JA (1997) J Chem Phys 106:1063