Generalization of the Gaussian electrostatic model: Extension to arbitrary angular momentum, distributed multipoles, and speedup with reciprocal space methods

G. Andrés Cisneros
Laboratory of Structural Biology, National Institute of Environmental Health Sciences, Research Triangle Park (RTP), North Carolina, 27709

Jean-Philip Piquemal
Laboratory of Structural Biology, National Institute of Environmental Health Sciences, Research Triangle Park (RTP), North Carolina, 27709 and Laboratoire de Chimie Théorique, Université Pierre et Marie Curie, Case 137, 4 Place Jussieu, 75252 Paris Cedex 05, France

Thomas A. Darden[a)]
Laboratory of Structural Biology, National Institute of Environmental Health Sciences, Research Triangle Park (RTP), North Carolina, 27709

The simulation of biological systems by means of current empirical force fields presents shortcomings due to their lack of accuracy, especially in the description of the nonbonded terms. We have previously introduced a force field based on density fitting termed the Gaussian electrostatic model-0 (GEM-0) J.-P. Piquemal *et al.* [J. Chem. Phys. **124**, 104101 (2006)] that improves the description of the nonbonded interactions. GEM-0 relies on density fitting methodology to reproduce each contribution of the constrained space orbital variation (CSOV) energy decomposition scheme, by expanding the electronic density of the molecule in *s*-type Gaussian functions centered at specific sites. In the present contribution we extend the Coulomb and exchange components of the force field to auxiliary basis sets of arbitrary angular momentum. Since the basis functions with higher angular momentum have directionality, a reference molecular frame (local frame) formalism is employed for the rotation of the fitted expansion coefficients. In all cases the intermolecular interaction energies are calculated by means of Hermite Gaussian functions using the McMurchie-Davidson [J. Comput. Phys. **26**, 218 (1978)] recursion to calculate all the required integrals. Furthermore, the use of Hermite Gaussian functions allows a point multipole decomposition determination at each expansion site. Additionally, the issue of computational speed is investigated by reciprocal space based formalisms which include the particle mesh Ewald (PME) and fast Fourier-Poisson (FFP) methods. Frozen-core (Coulomb and exchange-repulsion) intermolecular interaction results for ten stationary points on the water dimer potential-energy surface, as well as a one-dimensional surface scan for the canonical water dimer, formamide, stacked benzene, and benzene water dimers, are presented. All results show reasonable agreement with the corresponding CSOV calculated reference contributions, around 0.1 and 0.15 kcal/mol error for Coulomb and exchange, respectively. Timing results for single Coulomb energy-force calculations for $(H_2O)_n$, $n=64$, 128, 256, 512, and 1024, in periodic boundary conditions with PME and FFP at two different rms force tolerances are also presented. For the small and intermediate auxiliaries, PME shows faster times than FFP at both accuracies and the advantage of PME widens at higher accuracy, while for the largest auxiliary, the opposite occurs. © *2006 American Institute of Physics*. [DOI: 10.1063/1.2363374]

I. INTRODUCTION

The simulation of biological systems at present relies mainly on the use of classical empirical force fields[1–5] due to their speed. In general these methods represent the nonbonded contributions by a collection of point charges for the Coulomb interaction and a 6-12 Van der Waals term that attempts to reproduce the exchange and dispersion contributions.[6] A disadvantage of using simple isotropic atom-atom potentials is that this results in a loss of accuracy.[7]

More recently, methods based on distributed multipole expansions[7–10] have been proposed in order to improve the description of the electrostatic interaction.[11–13] However, the multipolar expansion fails to account for the charge penetration effects present at short range. This disadvantage has been addressed to an extent by the use of damping functions that correct the interaction energies at close intermolecular distances.[14–16]

a)Electronic mail: cisnero1@niehs.nih.gov

It has been shown that another possibility for the accurate determination of the intermolecular Coulomb contribution is by interacting the unperturbed frozen densities of the two molecules either numerically[17] or analytically.[18] The latter method is based on the density fitting (DF) formalism (Coulomb fitting),[19–21] where the electron density is expanded using Gaussian auxiliary basis sets (ABSs) centered on specific sites on the molecule. By employing the DF methodology, the intermolecular electrostatic contribution can be determined with any level of theory from which a one-electron relaxed density matrix can be obtained.

This formalism has been the foundation for the development of the Gaussian electrostatic model-0 (GEM-0).[22] GEM-0 provides a total energy expression that separately reproduces each of the components of the interaction energy, namely, Coulomb, exchange repulsion, polarization, and charge transfer. It has been parametrized in order to reproduce reference constrained space orbital variation[23,24] (CSOV) values by using an ABS restricted to s-type functions ($1=0$). The exchange-repulsion contribution is computed by means of the overlap model proposed by Wheatley and Price,[25] and Domene et $al.$[26] while the polarization and charge transfer terms are calculated in the spirit of the sum of interactions between fragments ab initio[12,16] (SIBFA) scheme, using the electrostatic potentials and electric fields calculated with the fitted densities.

The choice of s-type Gaussian functions for the fitting procedure facilitates the rotation of the fitted densities of the rigid fragments, in addition to allowing a fast evaluation of the required integrals for the energy components. However, this requires the use of several extra fitting sites in order to achieve a good fit of the initial ab $initio$ density. It was shown that the calculated expansion coefficients of the fitted density for a rigid molecule could be transferred to any fragment with the same internal geometry in a different orientation. Thus, the fitting procedure had to be performed only once for a single fragment of interest, thereby allowing the calculation of intermolecular interactions for any system once the fitting coefficients have been determined.

In order to demonstrate the applicability of the method, a water model with nine expansion sites was used to compute the total interaction energy and its individual components[22] for several systems. Ten stationary points in the water dimer potential-energy surface (PES),[27,28] three water clusters, as well as two metal-water dimers were used as examples. Errors below thermal noise at room temperature ($k_B T$) were obtained for each component of the interaction energy.

In this contribution we present an extension of GEM-0 in order to include Gaussian ABSs of arbitrary angular momentum. We term this extension the Gaussian electrostatic model (GEM). The use of fitting basis with angular momentum greater than zero presents two issues that need to be addressed: numerical instabilities (noise) in the fitting procedure and rotation of the fitting coefficients. Noise in the fitting procedure is controlled by using a modified Tikhonov regularization method.[29] Additionally, a damped Coulomb kernel[30] for the fitting integrals is also employed for certain systems.

We have chosen to use Hermite Gaussian functions[31] for

the ABSs (Ref. [32]) to take advantage of the McMurchie-Davidson (MD) recursion for the Coulombic[33] and overlap[34] integrals. Hermite Gaussians, being defined by partial derivatives of a spherical Gaussian either with respect to local coordinates defined by a reference (local) site frame[8,13] or with respect to global coordinates, provide a straightforward solution to the transformation of the fitting coefficients by using the chain rule. Furthermore, we show that the use of Hermite Gaussians leads to a natural definition of distributed point multipoles centered on the fitting sites.

Additionally, we present three reciprocal space methods that are related to the Ewald compensation methods used in calculations for periodic systems.[35] Of these, the two fast Fourier transform (FFT) based methods enable efficient linear scaling calculation of intermolecular interactions using Hermite Gaussian functions.

In all cases the ABSs used in the calculations are composed of uncontracted Gaussians arranged in s and spd shells only as in Ref. [18]. Shells of s functions correspond to Gaussians of angular momentum zero only. In the case of the spd shells, each Gaussian exponent is used to form $1s$, $3p$, and 5 (independent) d functions resulting in contributions only up to the quadrupole level for both the Hermite and multipole calculations. The calculations, however, generalize to arbitrary levels of angular momentum in a manner similar to Ref. [36].

We present intermolecular Coulomb interactions for the ten water dimers,[27,28] as well as one-dimensional (1D) electrostatic surface scans for a water dimer, a formamide dimer, a stacked benzene dimer, and a water-benzene dimer. Electronic densities at the B3LYP/6-31G* and B3LYP/aug-cc-pVTZ levels are fitted using four different ABSs. For the one-dimensional surface scans only densities at the B3LYP/6-31G* fitted with three ABSs are considered. Results for exchange repulsion at the same levels are also presented. All results are compared to reference values calculated with CSOV.[15,23] In addition, the Coulomb interaction energy for a set of water boxes of increasing size is calculated to predetermined accuracy using reciprocal space methods and timing results are discussed.

II. METHODS

In this section we present the theory and computational details employed in the present study. In Sec. II A we provide a brief explanation of the density fitting method and its implementation including the methods employed to control numerical instabilities. In Sec. II B we describe the transformation of the Cartesian Gaussian expansion coefficients to local Hermite coefficients, followed by the methodology to transform Hermite coefficients between coordinate frames in Sec. II C. Sec. II D presents the procedure to calculate site multipoles from local Hermite coefficients. Subsequently in Sec. II E the methods used to calculate the Coulomb and exchange-repulsion terms in the DF formalism are explained. The description of reciprocal space methods for computational speedup using Hermite Gaussians is presented in Sec. II F. Finally, in Sec. II G we describe the particular details of the calculations.

A. Density fitting

The use of ABSs for density fitting is a field of intense study. This method relies on the use of auxiliary basis functions to expand the molecular electron density. Briefly, the electron density may be fitted by minimizing the self-energy of the error in the density according to some metric $\hat{O}$:[20,21,37,38]

$$E_{\text{self}} = \langle \rho(r) - \tilde{\rho}(r) | \hat{O} | \rho(r) - \tilde{\rho}(r) \rangle, \tag{1}$$

where the approximate density $\tilde{\rho}(r)$ is expanded in an ABS based on Gaussian functions

$$\tilde{\rho}(r) = \sum_k x_k k(r). \tag{2}$$

The operator $\hat{O}$ can take several forms, typically the Coulomb operator $1/r$ is employed.[30] The minimization of Eq. (1) (self-interaction error) with respect to the expansion coefficients x_l leads to a linear system of equations:

$$\frac{\partial E_{\text{self}}}{\partial x_l} = -2 \sum_{\mu,\nu} P_{\mu\nu} \langle \mu\nu | \hat{O} | l \rangle + 2 \sum_k x_k \langle k | \hat{O} | l \rangle, \tag{3}$$

where $P_{\mu\nu}$ is the density matrix. This equation may be used for the determination of the coefficients

$$\mathbf{x} = \mathbf{G}^{-1} \mathbf{j}, \tag{4}$$

where $\mathbf{j} = P_{\mu\nu} \langle \mu\nu | \hat{O} | l \rangle$ and $\mathbf{G} = \langle k | \hat{O} | l \rangle$.

Since $\mathbf{G}$ is symmetric and positive definite it may be diagonalized to obtain its inverse. In our previous studies we had employed a modification of the matrix inverse similar to a singular value decomposition (SVD) procedure[29] by setting the inverse of the eigenvalue to zero if it is below a certain cutoff. However, this method produces undesirable numerical instabilities (noise) when the number of basis functions starts to grow when using Gaussian functions as we and others have discussed previously.[18,39] In addition, we have observed that these instabilities are also present when using only s-type (spherical) functions[22] albeit to a lower extent.

In the present implementation we have opted to use the Tikhonov regularization formalism,[29] similar to the constrained density fitting algorithm recently proposed by Misquitta and Stone.[40] In this method redundant basis set contributions are penalized by instead minimizing the following equation:

$$E_{\text{self}} + \lambda \sum_k x_k^2. \tag{5}$$

Additionally, Jung *et al.* have recently showed that the use of a damped Coulomb operator $\hat{O} = \text{erfc}(\beta r)/r$ can be used for the fitting procedure.[30] These authors have employed this kernel to localize the integrals in order to increase the calculation speed of three center Coulomb integrals in a quantum mechanical program. For our purposes, the damped Coulomb operator may also be employed to attenuate the near singular behavior due to long range interactions present in $\mathbf{G}$.

These expressions have been implemented by the authors in a FORTRAN90 code developed for the present study. All the required (real space) integrals are calculated using the McMurchie-Davidson recursions.[33] Besides implementing the noise control features explained above, this code generates the fitted coefficients in local frames for a Hermite basis and the corresponding site multipoles by means of the methods explained in the subsections below.

B. Hermite coefficients

One advantage of using fitted densities expressed in a linear combination of Gaussian functions is that the choice of Gaussian functions for the ABS need not be restricted to Cartesian Gaussians.[32] In the present work we have chosen to use (normalized) Hermite Gaussian functions for the calculation of the intermolecular interactions:

$$\Lambda_{tuv}(r, \alpha, \mathbf{S}) = \left(\frac{\alpha}{\pi} \right)^{3/2} \left(\frac{\partial}{\partial S_x} \right)^t \left(\frac{\partial}{\partial S_y} \right)^u \left(\frac{\partial}{\partial S_z} \right)^v \exp(-\alpha r_S^2), \tag{6}$$

where $\mathbf{S}$ denotes the expansion points.

The use of Hermite Gaussians in the calculation of the intermolecular interactions results in improved efficiency by the use of the MD recursions,[33] since the expensive Cartesian-Hermite transformation is avoided. Obtaining the Hermite expansion coefficients from the fitted Cartesian coefficients is straightforward since Hermite polynomials form a basis for the linear space of polynomials.[41]

Let hc_{tuv} and cc_{tuv} be the Hermite and Cartesian coefficients of order tuv, respectively, for an *spd* function with exponent α from the fitting basis. The zeroth order coefficient is

$$hc_{000} = cc_{000} + \frac{(cc_{200} + cc_{020} + cc_{002})}{2\alpha}, \tag{7}$$

the first order coefficients ($t + u + v = 1$) are

$$hc_{tuv} = -\frac{cc_{tuv}}{2\alpha}, \tag{8}$$

and the second order coefficients ($t + u + v = 2$) correspond to

$$hc_{tuv} = \frac{cc_{tuv}}{4\alpha^2}. \tag{9}$$

C. Coefficient transformations

As explained in Sec. I, in order to be able to use one set of fitting coefficients for the calculation of intermolecular interaction energies the coefficients must be transformed to each molecular fragment orientation. Additionally, the generalized Particle mesh Ewald (PME) approach outlined below requires that the coefficients be transformed to scaled fractional coordinates. The former transformation can be achieved by defining a *global* orthogonal coordinate system (global frame) and *local* orthogonal coordinate systems (local frame) for each fitting site in the fragment. Hermite Gaussians in the two coordinate systems, Λ_{tuv}^g and Λ_{tuv}^l, can be simply related using the chain rule. The same approach
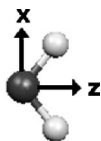
FIG. 1. Local frame definition for the oxygen atom in water.

works for the transformation to scaled fractional coordinates. This method has been previously developed for point dipoles and generalized to higher order multipoles.[36,42]

For the local frames two reference axes are selected for each site, based on the geometry of the molecule. For instance, in Fig. 1, for the local frame on the oxygen atom the first axis ($\mathbf{p}_1 = \mathbf{z}$) is along the bisector of the two hydrogen atoms, while the second axis ($\mathbf{p}_2 = \mathbf{x}$) points to one of the hydrogen atoms. Once chosen, an orthogonal local frame represented by the orthogonal unit vectors $\hat{x}$, $\hat{y}$, and $\hat{z}$ can be defined as follows: $\hat{z} = \mathbf{p}_1/|\mathbf{p}_1|$; $\mathbf{u} = \mathbf{p}_2 - (\mathbf{p}_2 \cdot \hat{z})$ and $\hat{x} = \mathbf{u}/|\mathbf{u}|$; $\hat{y} = \hat{z} \times \hat{x}$. This transformation has been explained in detail (as are the additional force terms on frame defining sites due to infinitesimal local frame rotations) in Refs. 36 and 42. These frame definitions are similar to the definitions used in Ref. 7 and in the OPEP code.[43] Note that in the case of Hermite functions, the use of the chain rule to calculate the coefficient transformation matrix is straightforward since the 3×3 coordinate frame transformation matrices are constants.

D. Site multipoles

In this subsection we present the methodology to obtain Cartesian point multipoles from the Hermite coefficients obtained in the fitting procedure. Note that since in the present work the highest order used for a Gaussian Hermite is 2 we will concentrate on multipoles expanded only up to quadrupole order. However, higher order multipoles can be obtained if an ABS with higher angular momentum is used, as explained below.

Challacombe *et al.* have shown that Hermite Gaussians have a simple relation to elements of the Cartesian multipole tensor.[44] Expanding on that work, once the Hermite coefficients have been determined, they may be employed to calculate point multipoles centered at the expansion sites. Thus, if hc_{tuv} represents the coefficient of a Hermite Gaussian of order Λ_{tuv}, then if this Hermite is normalized we have

$$hc_{000} \int \Lambda_0 dr = hc_{000}. \tag{10}$$

This guarantees that higher order multipole integrals will integrate to integer numbers, for example, for the dipole integral in the x direction, $\mathbf{d}_x$:

$$hc_{100} \int \int \int x\Lambda_{100} dx dy dz = hc_{100} \int x\frac{\partial}{\partial S_x}\Lambda_0 dx$$
$$= -hc_{100} \int x\frac{\partial}{lx}\Lambda_0 dx. \tag{11}$$

Integrating by parts this latter integral is evaluated as

$$-hc_{100} \int x\Lambda_0 dx\Big|_{-\infty}^{\infty} + hc_{100} \int \Lambda_0 dx = hc_{100}. \tag{12}$$

In the case of quadrupole integrals, $\mathbf{Q}_{x_a x_b}$, two cases need to be considered. When $a=b$, for example, for the x^2 quadrupole term, $\mathbf{Q}_{x^2}$,

$$hc_{200} \int \int \int x^2\Lambda_{200} dx dy dz = hc_{200}x^2\frac{\partial}{\partial x}\Lambda_0\Big|_{-\infty}^{\infty}$$
$$- 2hc_{200} \int x\frac{\partial}{\partial x}\Lambda_0 dx$$
$$= -2hc_{200}x\,\Lambda_0\Big|_{-\infty}^{\infty}$$
$$+ 2hc_{200} \int \Lambda_0 dx = 2hc_{200}. \tag{13}$$

When $a \neq b$, for example, for the xy quadrupole term, $\mathbf{Q}_{xy}$,

$$hc_{110} \int \int \int xy\Lambda_{110} dx dy dz = hc_{110}x\,\frac{\partial}{\partial x}\Lambda_0\Big|_{-\infty}^{\infty}$$
$$+ hc_{110}y\,\frac{\partial}{\partial y}\Lambda_0\Big|_{-\infty}^{y}$$
$$+ hc_{110} \int \int \Lambda_0 dx dy$$
$$= hc_{110}. \tag{14}$$

Note that in the case of the quadrupolar term when $a=b$ there will be a contribution from the s-type Hermites, since this involves an even function; for example, for the x^2 quadrupole term

$$hc_{000} \int \int \int x^2\Lambda_0 dx dy dz = hc_{000} \int x^2 e^{-\alpha x^2} dx$$
$$= \frac{1}{2\alpha}hc_{000}. \tag{15}$$

However, this latter contribution is not present in the traceless point quadrupole expansion, nor does it contribute to the multipolar electrostatic interaction. The off-diagonal quadrupolar terms have no contribution from the zeroth order Hermites (nor do the dipolar terms) because the integral of an odd function is zero. In practice, following Stone's definition,[7] we have used traceless quadrupoles.

Using the above distributed multipole definitions it is easy to calculate the penetration error in the site-site Coulomb interaction energy and to show that this error vanishes rapidly as the intersite distance increases. Thus density fitting, using a Hermite Gaussian ABS, leads to distributed multipoles that connect naturally with an accurate evaluation of the exact Coulomb interaction energy. This connection will be useful in the generation of damping functions[15] to account for the penetration error when using these multipoles (work in progress). Recently we have demonstrated the superiority of fitted density compared to point charges or distributed multipoles in quantum mechanical/molecular mechanical (QM/MM) implementations.[45] Multipoles defined

as above fit naturally into a multiscale treatment of the QM/MM problem, where the MM region nearest the active QM region is treated by density fitting using Hermite Gaussians, and more distant MM sites are treated by the corresponding multipoles together with their damping functions.

A further advantage of this approach to distributed multipoles is that, unlike some conventional multipole expansions,[46,47] the (spherical) multipole expansion obtained from Hermite Gaussians in this way is intrinsically finite of order $t+u+v$ (i.e., the highest angular momentum in the ABS), similar to the multipoles obtained by Volkov and Coppens.[48] This is easily seen since Cartesian polynomials of order $t+u+v$ can be expanded in a solid spherical harmonic expansion of the same and lower orders,[7] leading to a similar expansion for the Hermite Gaussians:

$$\Lambda_{tuv} = \left(\sum_{l=0}^{t+u+v} \sum_{m=-l}^{l} D_{tuv}^{lm} r^l C_{lm}(\theta, \varphi) \right) \exp(-\alpha r_S^2). \qquad (16)$$

Now, evaluating the spherical multipole contribution of Λ_{tuv} using Eq. (16) we obtain

$$
\begin{aligned}
Q_{l_2 m_2} &= \int r^{l_2+2} dr \int \Lambda_{tuv} C_{l_2 m_2}(\theta, \phi) \sin\theta d\theta d\varphi \\
&= \sum_{l_1=0}^{t+u+v} \sum_{m_1=-l_1}^{l_1} \int D_{tuv}^{l_1 m_1} r^{l_1} r^{l_2+2} \exp(-\alpha r^2) dr \\
&\quad \times \int C_{l_1 m_1}(\theta, \varphi) C_{l_2 m_2}(\theta, \varphi) \sin\theta d\theta d\varphi \\
&= \sum_{l_1=0}^{t+u+v} \sum_{m_1=-l_1}^{l_1} \int D_{tuv}^{l_1 m_1} r^{l_1} r^{l_2+2} \exp(-\alpha r^2) dr \, \delta_{l_1 l_2} \delta_{m_1 m_2}
\end{aligned}
$$

$$(17)$$

and thus for $l_2 > t+u+v$, $Q_{l_2 m_2} = 0$.

E. Frozen-core intermolecular interactions

Following the GEM-0 approach, the total intermolecular interaction energy can be computed as the sum of four intermolecular contributions:[24] Coulomb, exchange repulsion, polarization, and charge transfer. The first two are referred to as the frozen-core contribution while the latter are termed the induction term. In the present paper we will concentrate on the frozen core part exclusively.

In order to calculate the intermolecular Coulomb interaction between two molecules it is necessary to determine the nuclear-nuclear (N-N), nuclear-electron (N-e), and electron-electron (e-e) contributions to the Coulomb energy[49,50]

$$E_{\text{Coulomb}} = E_{N\text{-}N} + E_{N\text{-}e} + E_{e\text{-}e}. \qquad (18)$$

By making use of the approximate molecular electron density [Eq. (2)] we have previously shown[18] that the intermolecular Coulomb energy can be calculated by

$$
\begin{aligned}
E_{\text{Coulomb}} &= \sum_i \sum_j \frac{Z_{A_i} Z_{B_j}}{r_{A_i B_j}} + \sum_i \int \frac{Z_{A_i} \tilde{\rho}^B(r_2)}{r_{A_i 2}} dr_2 \\
&\quad + \sum_j \int \frac{Z_{B_j} \tilde{\rho}^A(r_1)}{r_{1 B_j}} dr_1 \\
&\quad + \int \int \frac{\tilde{\rho}^A(r_1) \tilde{\rho}^B(r_2)}{r_{12}} dr_1 dr_2,
\end{aligned}
$$

$$(19)$$

where Z_{A_i} represents the nuclei on molecule A, $\tilde{\rho}^A$ represents the approximate density of molecule A, Z_{B_j} represents the nuclei on molecule B, and $\tilde{\rho}^B$ represents the approximate density of molecule B. Note that in our force field formulation we use only the fitted density for the calculations, which differs from other DF methods used in DFT procedures where the fitted density is interacted with the density matrix. The use of only the fitted density (i.e., no density matrix) results in a less accurate representation.[39] However, this can be overcome by using extra points to place more ABSs such as at bond midpoints and/or lone pairs.[18,22]

For the calculation of the exchange-repulsion contribution, we have used the overlap model initially proposed by Wheatley and Price,[25] which we have shown to be applicable using the DF formalism.[22] This model relies on the observed proportionality between the overlap of the charge density and the exchange-repulsion energy

$$E_{\text{exchange-repulsion}} = K\Omega_\rho, \qquad (20)$$

where Ω_ρ is the overlap of charge density and K is a fitting parameter obtained as the slope of a linear regression of the overlap of charge density versus the corresponding *ab initio* exchange-repulsion energy values. By inserting the fitted densities of the interacting monomers into Ω_ρ we obtain

$$\Omega_\rho = \int \rho^A(r_A) \rho^B(r_B) dr \approx \int \tilde{\rho}^A(r_A) \tilde{\rho}^B(r_B) dr. \qquad (21)$$

F. Increasing computational efficiency using reciprocal space methods for Hermite Gaussians

In this section we describe the acceleration of the Coulomb and exchange-repulsion interaction calculations by using reciprocal space methods. For simplicity we first consider the overlap of charge density in an isolated cluster, describing how we split this into direct and reciprocal space contributions and then how each is calculated. Following this we discuss Coulomb interactions in an isolated cluster, using Ewald-type counterion and coion densities. Finally we discuss the extension of these results to a system under periodic boundary conditions (PBC).

To begin, consider a cluster of N identical molecules, e.g., water, each with an identical local frame density expansion over a set of expansion sites **S**. For this system the molecular density in the local frame can be expressed by

$$\tilde{\rho}_{mol}^l(\mathbf{r}) = \sum_{s \in \mathbf{S}} \tilde{\rho}_s^l(r) = \sum_{s \in \mathbf{S}} \sum_{\alpha \in A_s} \tilde{\rho}_{s,\alpha}^l(r)$$

$$= \sum_{s \in \mathbf{S}} \sum_{\alpha \in A_s} \sum_{t,u,v} hc_{tuv}^l(\alpha,s) \Lambda_{tuv}^l(r,\alpha,s), \qquad (22)$$

where $s \in \mathbf{S}$ are the expansion points, A_s are the set of primitive Gaussian exponents on each expansion site, and hc_{tuv}^l are the expansion coefficients in the local frame. Thus, the total density of the system in the global frame is

$$\tilde{\rho}_{tot}(r) = \sum_{n=1}^N \sum_{s \in \mathbf{S}_n} \sum_{\alpha \in A_{\mathbf{S}}} \sum_{t,u,v} hc_{tuv}^g(\alpha,s) \Lambda_{tuv}^g(r,\alpha,s). \qquad (23)$$

The overlap calculation, $\Omega_{tot} = \int \tilde{\rho}_{tot}(r) \tilde{\rho}_{tot}(r) d^3r$, can be accelerated by performing part of it in reciprocal space, using a method inspired by Fusti-Molnar and Pulay[51] and Parrinello and co-workers.[52,53] For a given α_0 (called the splitting exponent) we can split $\tilde{\rho}_{tot}$ into $\tilde{\rho}_{tot,c}$ and $\tilde{\rho}_{tot,d}$, where c and d denote compact and diffuse Hermite Gaussian functions, respectively.

$$\tilde{\rho}_{tot,c}(r) = \sum_{n=1}^N \sum_{s \in \mathbf{S}_n} \sum_{\alpha \in A_{\mathbf{S}}; \alpha \geq \alpha_0} \sum_{t,u,v} hc_{tuv}^g(\alpha,s) \Lambda_{tuv}^g(r,\alpha,s), \qquad (24)$$

$$\tilde{\rho}_{tot,d}(r) = \sum_{n=1}^N \sum_{s \in \mathbf{S}_n} \sum_{\alpha \in A_{\mathbf{S}}; \alpha < \alpha_0} \sum_{t,u,v} hc_{tuv}^g(\alpha,s) \Lambda_{tuv}^g(r,\alpha,s). \qquad (25)$$

Using this definition the total overlap becomes

$$\Omega_{tot} = \int_{R^3} \tilde{\rho}_{tot,c}(r) \tilde{\rho}_{tot,c}(r) dr + 2 \int_{R^3} \tilde{\rho}_{tot,c}(r) \tilde{\rho}_{tot,d}(r) dr$$

$$+ \int_{R^3} \tilde{\rho}_{tot,d}(r) \tilde{\rho}_{tot,d}(r) dr, \qquad (26)$$

where the first integral is calculated in direct space and the latter two will be reexpressed in order to calculate them in reciprocal space. Note that the reciprocal contribution to the intramolecular (self) overlap needs to be calculated in direct space and subtracted when using this method. To employ reciprocal space methods the density of the cluster needs to be extended periodically. To this end, for any tolerance δ we can define an effective radius R_δ of the cluster with

$$\int_{r > R_\delta} \tilde{\rho}_{tot}(r) dr < \delta. \qquad (27)$$

Let $T > 2R_\delta$, and for any site $s \in \mathbf{S}_n$ and any α, t, u, and v, define the periodic extension of Λ_{tuv}^g by

$$\Lambda_{tuv}^{g,P}(r,\alpha,s) = \sum_{l_1} \sum_{l_2} \sum_{l_3} \Lambda_{tuv}^g(r + T(l_1,l_2,l_3), \alpha, s). \qquad (28)$$

Note that $\Lambda_{tuv}^{g,P}$ is periodic with period T. Define the periodic extension of $\tilde{\rho}_{tot,c}$ and $\tilde{\rho}_{tot,d}$ using $\Lambda_{tuv}^{g,P}$ and expressions similar to Eqs. (24) and (25). Then using these periodic extensions and Parseval's relations[54] we can approximate Ω_{tot} as

$$\Omega_{tot} \approx \int_{R^3} \tilde{\rho}_{tot,c}(r) \tilde{\rho}_{tot,c}(r) dr + 2 \int_{\mathcal{T}^3} \tilde{\rho}_{tot,c}^P(r) \tilde{\rho}_{tot,d}^P(r) dr$$

$$+ \int_{\mathcal{T}^3} \tilde{\rho}_{tot,d}^P(r) \rho(r) \tilde{\rho}_{tot,d}^P(r) dr$$

$$= \int_{R^3} \tilde{\rho}_{tot,c}(r) \tilde{\rho}_{tot,c}(r) dr + T^3 \sum_{\mathbf{m}} \hat{\tilde{\rho}}_{tot,c}^P(\mathbf{m}) \hat{\tilde{\rho}}_{tot,d}^P(-\mathbf{m})$$

$$+ T^3 \sum_{\mathbf{m}} \hat{\tilde{\rho}}_{tot,c}^P(-\mathbf{m}) \hat{\tilde{\rho}}_{tot,d}^P(\mathbf{m}) + T^3 \sum_{\mathbf{m}} \hat{\tilde{\rho}}_{tot,d}^P(\mathbf{m}) \hat{\tilde{\rho}}_{tot,d}^P(-\mathbf{m}), \qquad (29)$$

where $\mathcal{T}^3$ denotes the cube of side T centered at the origin, $\mathbf{m} = (m_1, m_2, m_3)/T$, and $\hat{\tilde{\rho}}_{tot,c}^P(\mathbf{m})$ is given by

$$\hat{\tilde{\rho}}_{tot,c}^P(\mathbf{m}) = \frac{1}{T^3} \int_{\mathcal{T}^3} \tilde{\rho}_{tot,c}^P(r) \exp(-2\pi i \mathbf{m} \cdot r) d^3r$$

$$= \frac{1}{T^3} \int_{R^3} \tilde{\rho}_{tot,c}(r) \exp(-2\pi i \mathbf{m} \cdot r) d^3r, \qquad (30)$$

and $\hat{\tilde{\rho}}_{tot,d}^P(\mathbf{m})$ is defined similarly.

Thus, recalling Eq. (23), Ω reduces to the efficient real space evaluation of the overlap of compact Hermite Gaussian functions and the efficient approximation of the Fourier coefficients of compact and diffuse Hermite Gaussian functions.

1. Direct space integrals

The first integral on the right hand side (RHS) of Eq. (29) is calculated in direct space. For the direct space evaluation we use the MD recursion[33] as explained above. The MD recursion was developed to evaluate Coulombic integrals and later extended to treat overlap.[34] To keep the paper self-contained we prove a more general version of this recursion, applicable to Cartesian derivatives of any smooth function of r.

For any function $f(r)$, $r = (x^2 + y^2 + z^2)^{1/2}$ we want an efficient recursion for the derivatives $(\partial/\partial x)^t (\partial/\partial y)^u (\partial/\partial z)^v f(r)$. Following the notation in the original MD paper, let $R(0,0,0,0) = f(r)$, $R(0,0,0,n+1) = 1/r(\partial/\partial r)R(0,0,0,n)$, and $R(t,u,v,n) = (\partial/\partial x)^t (\partial/\partial y)^u (\partial/\partial z)^v R(0,0,0,n)$. In particular, $(\partial/\partial x)^t (\partial/\partial y)^u (\partial/\partial z)^v f(r) = R(t,u,v,0)$. Then

$$R(t+1,u,v,n) = xR(t,u,v,n+1) + tR(t-1,u,v,n+1),$$

$$R(t,u+1,v,n) = yR(t,u,v,n+1) + uR(t,u-1,v,n+1),$$

$$R(t,u,v+1,n) = zR(t,u,v,n+1) + vR(t,u,v-1,n+1). \qquad (31)$$

This can be proven as follows:

$$R(t+1,u,v,n) = \left(\frac{\partial}{\partial x}\right)^{t+1}\left(\frac{\partial}{\partial y}\right)^{u}\left(\frac{\partial}{\partial z}\right)^{v}R(0,0,0,n)$$

$$= \left(\frac{\partial}{\partial y}\right)^{u}\left(\frac{\partial}{\partial z}\right)^{v}\left(\frac{\partial}{\partial x}\right)^{t}\left(\frac{\partial}{\partial x}\right)R(0,0,0,n)$$

$$= \left(\frac{\partial}{\partial y}\right)^{u}\left(\frac{\partial}{\partial z}\right)^{v}\left(\frac{\partial}{\partial x}\right)^{t}\left(\frac{\partial}{\partial r}\right)R(0,0,0,n)\frac{\partial r}{\partial x}$$

$$= \left(\frac{\partial}{\partial y}\right)^{u}\left(\frac{\partial}{\partial z}\right)^{v}\left(\frac{\partial}{\partial x}\right)^{t}\left(\frac{\partial}{\partial r}\right)\frac{x}{r}\left(\frac{\partial}{\partial r}\right)R(0,0,0,n)$$

$$= \left(\frac{\partial}{\partial y}\right)^{u}\left(\frac{\partial}{\partial z}\right)^{v}\left(\frac{\partial}{\partial x}\right)^{t}xR(0,0,0,n+1). \quad (32)$$

Using Leibniz' rule on the last line of Eq. (32) we obtain Eq. (31). The recursions for $R(t,u+1,v,n)$ and $R(t,u,v+1,n)$ are obtained similarly. The use of this method allows the determination of all the direct space integrals used in this study, i.e., overlap, Coulomb, and damped Coulomb, for each of which the required starting function $f(\mathbf{r})$ is known. Ahlrichs has recently shown that a similar generalization can be obtained for the Obara-Saika recursion.[55]

2. Reciprocal space integrals

In order to evaluate the two sums on the RHS of Eq. (29) we need to obtain the Fourier coefficients of $\tilde{\rho}_{\text{tot},c}^{P}(\mathbf{m})$ and $\tilde{\rho}_{\text{tot},d}^{P}(\mathbf{m})$. Note that these need to be calculated separately to avoid redundant calculation of the compact-compact interactions, which in any case do not converge for reasonable grid sizes.

The first procedure to obtain these coefficients is the explicit evaluation of the Fourier transform:

$$\hat{\Lambda}_{tuv}^{g,P}(\mathbf{m},\alpha,s) = \frac{1}{T^3}\left(\frac{\alpha}{\pi}\right)^{3/2}\left(\frac{\partial}{\partial s_x}\right)^{t}\left(\frac{\partial}{\partial s_y}\right)^{u}\left(\frac{\partial}{\partial s_z}\right)^{v}$$

$$\times \exp\left(-\frac{\pi^2\mathbf{m}^2}{\alpha}\right)\exp(-2\pi i\mathbf{m}\cdot s). \quad (33)$$

Then $\hat{\tilde{\rho}}_{\text{tot}}^{P}(\mathbf{m})$ can be obtained by linear combination of $\hat{\Lambda}_{tuv}^{g,P}(\mathbf{m},\alpha,s)$. Note that in general the number of reciprocal vectors $\mathbf{m}$ needed in Eq. (29) to converge the sums to a given tolerance scales linearly with the number of molecules N.[73] Since the number of basis elements $\Lambda_{tuv}^{g,P}$ also scales linearly with N, this method scales as order N^2 as in regular Ewald summation. One way to avoid the $O(N^2)$ scaling is to employ FFTs, which reduce the scaling to $N\log N$. Two methods that employ FFT's have been implemented as explained below.

The second method we have implemented for the determination of the Fourier coefficients is the PME method. The use of PME is accomplished by rewriting the complex exponential as

$$\exp(-2\pi i\mathbf{m}\cdot s) = \exp\left(-2\pi i\frac{m_1}{M}u_1\right)\exp\left(-2\pi i\frac{m_2}{M}u_2\right)$$

$$\times \exp\left(-2\pi i\frac{m_3}{M}u_3\right), \quad (34)$$

where, e.g., $u_1 = M(s_x/T)$ and T/M is the grid step size (in a. u.) of the PME grid. Next we approximate $\exp(-2\pi i(m_1/M)u_1)$ by the Euler spline $g_q(m_1/M,u_1)$ as defined by Essmann et al.[56] and optimized by Toukmaji et al.[42] Using this definition for the complex exponential we obtain

$$\hat{\Lambda}_{tuv}^{g,P}(\mathbf{m},\alpha,s)_{\text{PME}} = \frac{1}{T^3}\left(\frac{\alpha}{\pi}\right)^{3/2}\exp\left(-\frac{\pi^2\mathbf{m}^2}{\alpha}\right)$$

$$\times \left(\frac{\partial}{\partial s_x}\right)^{t}\left(\frac{\partial}{\partial s_y}\right)^{u}\left(\frac{\partial}{\partial s_z}\right)^{v}g_q(u_1)g_q(u_2)g_q(u_3). \quad (35)$$

The partial derivatives with respect to Cartesian coordinates can be reexpressed in terms of $\partial/\partial u_1, \partial/\partial u_2, \partial/\partial u_3$ and the coefficients of the Hermites transformed to the (u_1,u_2,u_3) coordinate system. Methods for the efficient evaluation of these approximations have been explained in detail by Sagui et al.[36] Note that separate grids are needed for each exponent on the ABS, except when it only contains s-type functions with constant coefficients.

The third method is to directly sample the density on the grid as in the fast Fourier-Poisson (FFP) method of York and Yang[57] or the very similar Fourier transform Coulomb (FTC) method by Fusti-Molnar and Pulay.[51] In this case $\tilde{\rho}_{\text{tot},c}$ cannot be directly sampled, as discussed by Fusti-Molnar and Pulay. These authors solved this problem by using a modified low-pass filter operation on the compact basis elements. We have instead chosen to use a Gaussian splitting method inspired by Shan et al.[58] to render the core Hermite Gaussians more diffuse and the diffuse more compact.

Specifically, consider an individual term $\Lambda_{tuv}^{g,P}(\mathbf{m},\alpha_c,s_1)\Lambda_{t'u'v'}^{g,P}(\mathbf{m},\alpha_d,s_2)$ from the second sum of Eq. (29), where $\alpha_c \geq \alpha_0$ and $\alpha_d < \alpha_0$.

For any $\bar{\alpha} > \alpha_0$ replace α_c by $\alpha_c' = \alpha_c\bar{\alpha}/(\bar{\alpha}+\alpha_c)$ and α_d by $\alpha_d'' = \alpha_d\bar{\alpha}/(\bar{\alpha}-\alpha_d)$. In practice we have found that a good choice is $\bar{\alpha} = 2\alpha_0$ because in this case α_c' and α_d'' all tend to cluster near $\bar{\alpha}$ so that a uniform grid for sampling can be chosen. Then we note that

$$\hat{\Lambda}_{tuv}^{g,P}(\mathbf{m},\alpha_c,s_1)\hat{\Lambda}_{t'u'v'}^{g,P}(-\mathbf{m},\alpha_d,s_2)$$

$$= \hat{\Lambda}_{tuv}^{g,P}(\mathbf{m},\alpha_c',s_1)\hat{\Lambda}_{t'u'v'}^{g,P}(-\mathbf{m},\alpha_d'',s_2), \quad (36)$$

where $\Lambda_{tuv}^{g,P}(r,\alpha_c',s_1)$ and $\Lambda_{t'u'v'}^{g,P}(r,\alpha_d'',s_2)$ can be directly sampled.

In the case of diffuse-diffuse interactions

$$\hat{\Lambda}_{tuv}^{g,P}(\mathbf{m}, \alpha_{d_1}, s_1)\hat{\Lambda}_{t'u'v'}^{g,P}(-\mathbf{m}, \alpha_{d_2}, s_2)$$

$$= \hat{\Lambda}_{tuv}^{g,P}(\mathbf{m}, \alpha''_{d_1}, s_1)\hat{\Lambda}_{t'u'v'}^{g,P}(-\mathbf{m}, \alpha''_{d_2}, s_2)\exp\left(\frac{-2\pi^2\mathbf{m}^2}{\bar{\alpha}}\right),$$

(37)

thus, the diffuse-diffuse interactions are corrected in reciprocal space by the factor $\exp(-2\pi m^2/\bar{\alpha})$.

Note that in general the calculation of the exchange interaction requires different proportionality constants K for each molecular pair, thus different grids for different molecules. Nevertheless, these reciprocal space methods can be useful if most of the molecules are of one or two types such as a solute in a solvent bath. In the systems we have tested so far we have seen only modest reduction in computing time (see below). However, for systems with diffuse functions (thus, many direct space interactions), we expect greater savings by the use of these reciprocal space methods.

3. Coulomb interaction with Ewald splitting

To extend the above reciprocal space methods to Coulomb interactions in an isolated cluster we use the truncated Coulomb kernel $g_\delta(r)$ defined by $g_\delta(r) = 1/r, r \leq 2R_\delta$ and $g_\delta(r) = 0, r > 2R_\delta$ following Fusti-Molnar and Pulay.[51] In addition we now choose T twice as large as in the overlap case, i.e., satisfying $T > 4R_\delta$ so that periodic images of the cluster do not interact via $g_\delta(r)$. Finally we introduce counter- and co-ion densities[35] (Ewald splitting method) to allow efficient cutoffs in the direct space compact-compact interactions. Here we discuss the e-e Coulombic integrals (N-e and N-N terms are handled similarly). Recalling that compact and diffuse Hermites are defined in terms of the exponent α_0, the site density is reexpressed as

$$\tilde{\rho}_s^g(r) = \tilde{\rho}_{s,c}^g(r) + \tilde{\rho}_{s,d}^g(r) = (\tilde{\rho}_{s,c}^g(r) - \tilde{\rho}_{s,\alpha_0}^g(r))$$
$$+ (\tilde{\rho}_{s,\alpha_0}^g(r) + \tilde{\rho}_{s,d}^g(r)),$$

(38)

where

$$\tilde{\rho}_{s,\alpha_0}^g(r) = \sum_{tuv}\left(\sum_{\alpha \in A_{\mathbf{S}}; \alpha \geq \alpha_0} hc_{tuv}^g(\alpha, s)\right)\Lambda_{tuv}^g(\alpha_0, s).$$

(39)

By using this splitting scheme, $\langle\tilde{\rho}_{s_1,c}^g(r) - \tilde{\rho}_{s_1,\alpha_0}^g(r)\|\tilde{\rho}_{s_2,c}^g(r) - \tilde{\rho}_{s_2,\alpha_0}^g(r)\rangle$ can be neglected if $|s_1 - s_2|$ exceeds twice the Gaussian extent of $\tilde{\rho}_{s,\alpha_0}^g(r)$ since $\tilde{\rho}_{s,c}^g(r) - \tilde{\rho}_{s,\alpha_0}^g(r)$ has no nonzero multipole moments.[31]

The compact-diffuse and diffuse-diffuse integrals in reciprocal space are treated similarly to the overlap case by using Parseval's relations, the fact that the convolution of $g_\delta(r)$ with a periodic density is itself periodic, the convolution product relation in reciprocal space,[54] and the Fourier transform of $g_\delta(r)$ to obtain, e.g.,

$$\int_{R^3}\int_{R^3}\frac{\tilde{\rho}_{\text{tot},d}(r_1)\tilde{\rho}_{\text{tot},d}(r_2)}{|r_1 - r_2|}dr_1 dr_2$$

$$\approx \int_{T^3}\int_{T^3}\tilde{\rho}_{\text{tot},d}^P(r_1)g_\delta(r_2 - r_1)\tilde{\rho}_{\text{tot},d}^P(r_2)dr_1 dr_2$$

$$= T^3\sum_{\mathbf{m}}\hat{\tilde{\rho}}_{\text{tot},d}^P(\mathbf{m})\hat{\tilde{\rho}}_{\text{tot},d}^P(-\mathbf{m})\frac{1 - \cos(4\pi|\mathbf{m}|R_\delta)}{\pi\mathbf{m}^2},$$

(40)

where $|\mathbf{m}|$ denotes the norm of the vector $\mathbf{m} = (m_1, m_2, m_3)$.

The compact-compact integrals in real space can be further accelerated by

$$\langle\tilde{\rho}_{s_1,c}^g(r) - \tilde{\rho}_{s_1,\alpha_0}^g(r)\|\tilde{\rho}_{s_2,c}^g(r) - \tilde{\rho}_{s_2,\alpha_0}^g(r)\rangle$$

$$= \sum_{\alpha' \in A_{s_2}; \alpha' \geq \alpha_0}\sum_{\alpha \in A_{s_1}; \alpha \geq \alpha_0}\langle\tilde{\rho}_{s_1,\alpha}^g(r)\|\tilde{\rho}_{s_2,\alpha'}^g(r) - \delta_{s_2,\alpha'}\rangle$$

$$- \sum_{\alpha' \in A_{s_2}; \alpha' \geq \alpha_0}\langle\tilde{\rho}_{s_1,\alpha_0}^g(r)\|\tilde{\rho}_{s_2,\alpha'}^g(r) - \delta_{s_2,\alpha'}\rangle$$

$$- \sum_{\alpha \in A_{s_1}; \alpha \geq \alpha_0}\langle\tilde{\rho}_{s_1,\alpha}^g(r)\|\tilde{\rho}_{s_2,\alpha_0}^g(r) - \delta_{s_2,\alpha_0}\rangle + \langle\tilde{\rho}_{s_1,\alpha_0}^g(r)|$$

$$\times|\tilde{\rho}_{s_2,\alpha_0}^g(r) - \delta_{s_2,\alpha_0}\rangle,$$

(41)

where for any α,

$$\delta_{s_2,\alpha} = \sum_{tuv} hc_{tuv}^g(\alpha, s_2)\Lambda_{tuv}(r, \alpha_\infty, s_2),$$

(42)

and α_∞ corresponds to a very large exponent (effectively replacing the Hermite with exponent α by its multipole approximation). Thus as above, the individual terms on the right hand side of Eq. (41) can be neglected as soon as $s_1 + s_2$ is larger than the sum of the relevant Gaussian extents.[31]

4. Extension to periodic boundary conditions

In the case of a periodic system with unit cell U (triclinic in general) the fitted density is already periodic and can be split as above into compact and diffuse components. The overlap and Coulomb integrals extend over U by definition and Parseval's relations hold for the compact-diffuse and diffuse-diffuse interactions. The overlap calculation converges in direct space using cutoff methods but as with the cluster calculation is accelerated considerably by reciprocal space methods.

The Coulomb interaction converges slowly in direct space, if the unit cell is neutral, and the result is conditional on summation order. For point charges the Ewald sum can be interpreted (in the limit) as the Coulomb energy of the unit cell embedded in a large array of periodic copies of itself, which is further embedded in a conducting medium,[59] and this interpretation holds as well for the present case of nuclei plus fitted electron density. The same Ewald splitting strategy is used as described above for finite clusters with $g_\delta(r)$ replaced by $1/r$; however, in contrast to that case, under PBC the system must be neutral, due to the infinite range of the non-truncated Coulomb interaction. In practice, each of the compact and diffuse contributions is neutralized by a uniform plasma. With reciprocal space methods and overall net-

neutral systems this is achieved automatically by the neglect of the $\mathbf{m}=0$ term. If real space methods such as multigrid[60] are used instead to solve Poisson's equation, the grid must be explicitly neutralized.

G. Computational details

Relaxed one-electron molecular densities for the test cases were obtained from *ab initio* calculations performed with CADPAC.[61] For the determinations of the ten water dimers, the electronic densities were calculated at the B3LYP (Refs. [62] and [63]) levels with the 6-31G* and aug-cc-pVTZ basis sets. Electronic densities for all molecules in the 1D surface scans were obtained at the B3LYP/6-31G* level only. The geometries for the ten water dimers were based on the geometries reported by van Duijneveldt–van de Rijdt *et al.*.[28] All molecular geometries were optimized at the B3LYP/aug-cc-pVTZ level. Reference values for all systems were calculated at the corresponding levels of theory and basis set with the CSOV method,[23,24] using the same molecular geometries. For the CSOV electrostatic and exchange-repulsion analysis, all calculations were performed with a modified version of the HONDO95 package.[24,50]

The density fit was done with three different ABSs. The first two correspond to DGAUSS A1 and P1 Coulomb fitting ABS.[64,65] The final ABS, denoted g03, was obtained from the automatic fitting utility from GAUSSIAN03.[66] The g03 ABS was obtained by using the 6-311G** basis sets for the corresponding atoms.[18] Due to the extensive size of this last ABS, the f function orbitals were deleted and only the s and spd functions were retained. In the case of the P1 set, the original basis sets contain only s and d basis functions; in this study the d functions were expanded to represent spd shells.[64,65] Additional calculations with the CFIT basis set[67] for the ten water dimers were performed, using the same "transformation" as for the g03 ABS. ABSs were placed on all atomic positions, as well as bond midpoints (MPs) on bonds involving hydrogen for all fitting procedures based on our previous study for intermolecular Coulomb interactions.[18]

The calculated density matrices were used as input for the program described in Sec. II A to obtain the local Hermite coefficients for the corresponding ABSs. As explained in Sec. II A, there are two parameters that can be used to control noise in the fit: λ for the Tikhonov regularization and β for the damped Coulomb kernel (see Sec. II A). Note that for the latter, when $\beta=0$ the operator reduces to the usual Coulomb kernel. The procedure to obtain the optimal parameters for the fitting procedure was done in the following manner. The fitting coefficients and point multipoles for a fragment are obtained for a given λ and β. These coefficients and multipoles are tested by calculating the intermolecular Coulomb interaction for a single homodimer of that fragment and comparing the resulting Coulomb interactions to the corresponding CSOV value. This procedure is repeated until the optimal parameters are found.[74]

Given the fact that our method requires the rotation of the densities, it was necessary to obtain density matrices converged with very tight criteria; 10^{-12} rms change in elements in the density matrix. In order to determine if the calculation
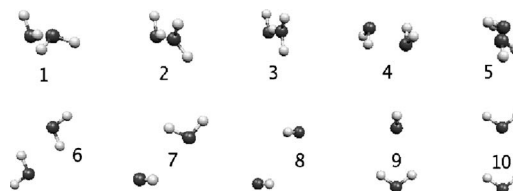

FIG. 2. Orientation for the ten water dimers (Ref. [25]).

of *ab initio* density matrices in different orientations affected the fitting coefficients, we decided to average coefficients calculated from density matrices of the same fragments in 99 random orientations. The results obtained for these average coefficients are compared to results calculated with coefficients from a single fit for the 1D surface scans for homodimers only.

Note that since all the ABSs employed for the fit in all our calculations include only up to d-type functions, the resulting multipoles extracted from the calculated Hermite coefficients only include contributions up to quadrupole moments.

For the timing results we prepared a series of TIP3P (Ref. [68]) water boxes containing 64, 128, 256, 512, and 1024 molecules and equilibrated them at the common density of 0.997 g/ml at room temperature and 1 atm using the TINKER package.[69] After this, the geometry of the optimized water molecule, as explained above, was superposed on the individual waters in the box in order to be able to employ the calculated Hermite coefficients.

III. RESULTS

In this section we present the results obtained with the DF method compared to reference *ab initio* calculations. Fitting coefficients and point multipoles were calculated for water, formamide, and benzene. These coefficients were employed for the calculation of the frozen-core contributions for the test systems which include ten water dimers (Fig. [2]), as well as 1D surface scans for the canonical water dimer, formamide dimer, stacked benzene dimer, and water-benzene dimer (Fig. [3]).

In Sec. III A the intermolecular Coulomb interactions for all test cases are discussed. Section III B contains the results for the exchange-repulsion intermolecular interactions calculated with the overlap method, with discussion and analysis relating to the optimal fitting parameters for the frozen-core contributions. Subsequently, results for some water clusters along with timing data are presented in Sec. III C, followed by concluding remarks.
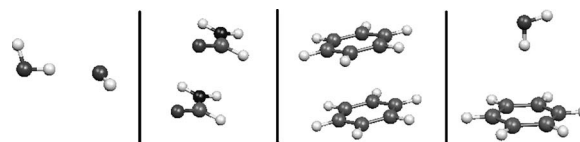

FIG. 3. Dimers used for the 1D Coulomb and exchange-repulsion intermolecular scans.

TABLE I. Average (maximum) absolute errors in Coulomb interaction with respect to CSOV for the ten water dimers (in kcal/mol).

	6-31G*				aug-cc-pVTZ			
	A1	P1	g03	CFIT	A1	P1	g03	CFIT
Hermite	0.26(0.45)	0.08(0.18)	0.02(0.09)	0.02(0.11)	0.56(1.14)	0.15(0.34)	0.12(0.28)	0.04(0.14)
Multipole	0.81(1.62)	0.79(1.73)	0.64(1.66)	1.75(3.02)	1.91(2.85)	1.65(2.29)	1.75(2.74)	1.45(2.28)

A. Intermolecular Coulomb interactions

Intermolecular Coulomb interaction results for the ten water dimers show good agreement with respect to CSOV, as can be seen in Table I (individual dimer results are reported in Supplementary Information). As can be seen, the results improve with the quality of the ABS as is to be expected. The calculated average errors for the interaction energies from Hermite coefficients are below 0.2 kcal/mol for all ABSs except for the A1 basis (see Table I), which has the smallest number of auxiliary functions per atom, in agreement with our previous results.[18] Note that in the case of the *ab initio* densities calculated with the aug-cc-PVTZ basis, the A1, P1, and g03 basis sets do not perform as well as for densities calculated with 6-31G* basis. This is to be expected

since these three ABSs are designed for double zeta basis sets. In contrast, the CFIT basis performs very well for both double zeta and triple zeta basis sets since it is designed for the latter basis sets. As can be seen from Table I the errors from the multipoles with respect to the CSOV reference values are considerably larger, around 1 kcal/mol or larger on average.

The results for the one dimensional scans for the homodimers are presented in Fig. 4–6. For the distance dependent scans (water and benzene dimers) the interactions calculated with the Hermite Gaussians show reasonable agreement with CSOV throughout the distance range. In the case of the water dimer, the short-mid- and long-range part of the curve (2.3–3.5 Å) was explored with a 0.1 Å interval;
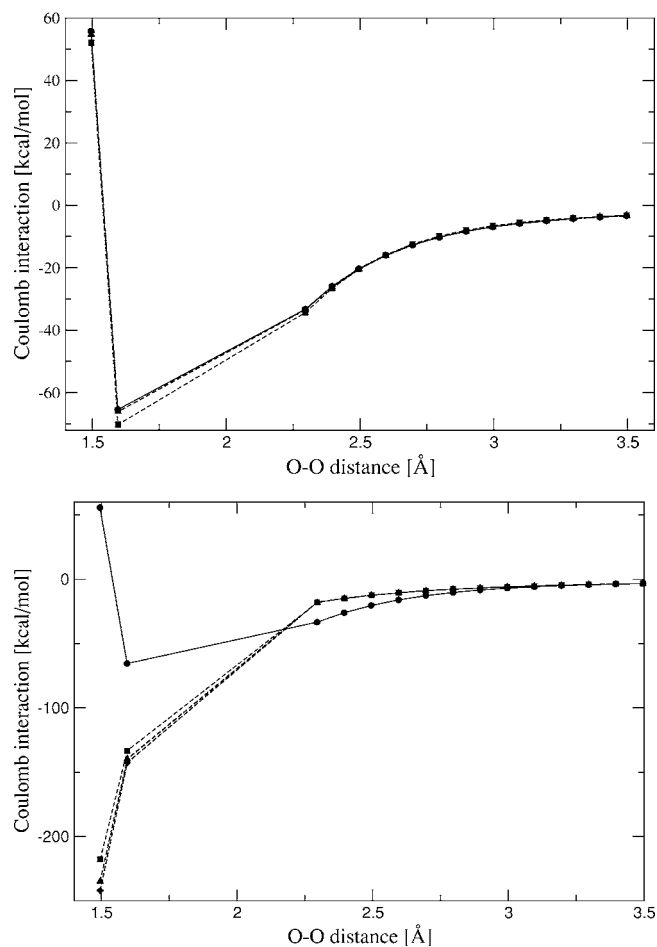

FIG. 4. Water dimer (structure 1) Coulomb interaction energies from Hermite (top) and multipoles (bottom) for a range of distances. Closed circles—CSOV; closed squares—A1 average; open squares—A1 single; closed diamonds—P1 average; open diamonds—P1 single; closed triangles—g03 average; open triangles—g03 single.
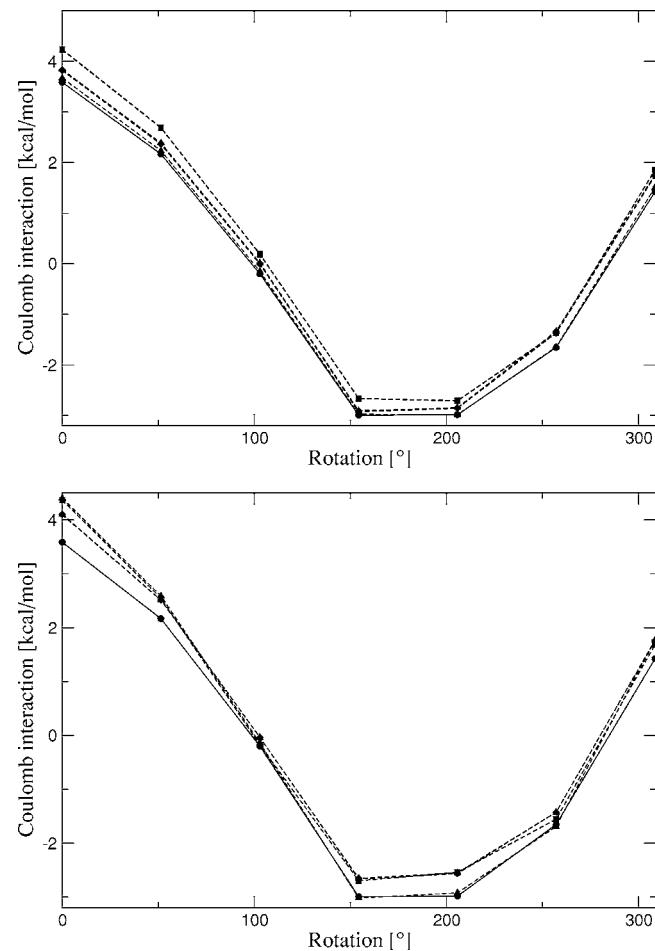

FIG. 5. Formamide dimer Coulomb interaction energies from Hermite (top) and multipoles (bottom) rotating about one of the fragments. Closed circles—CSOV; closed squares—A1 average; open squares—A1 single; closed diamonds—P1 average; open diamonds—P1 single; closed triangles—g03 average; open triangles—g03 single.
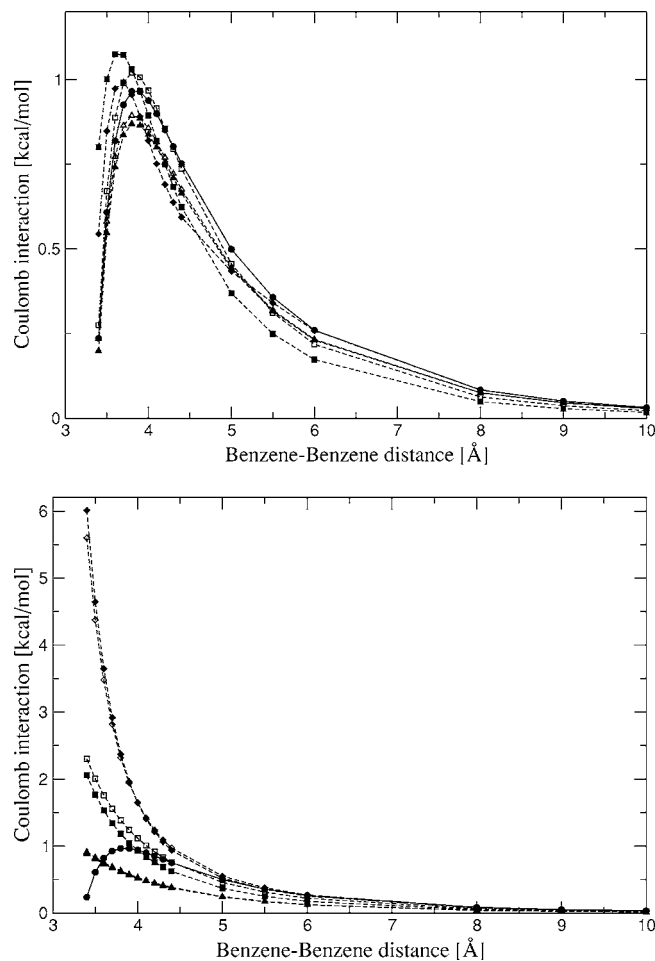
FIG. 6. Stacked benzene dimer Coulomb interaction energies from Hermite (top) and multipoles (bottom) for a range of distances. Closed circles—CSOV; closed squares—A1 average; open squares—A1 single; closed diamonds—P1 average; open diamonds—P1 single; closed triangles—g03 average; open triangles—g03 single.
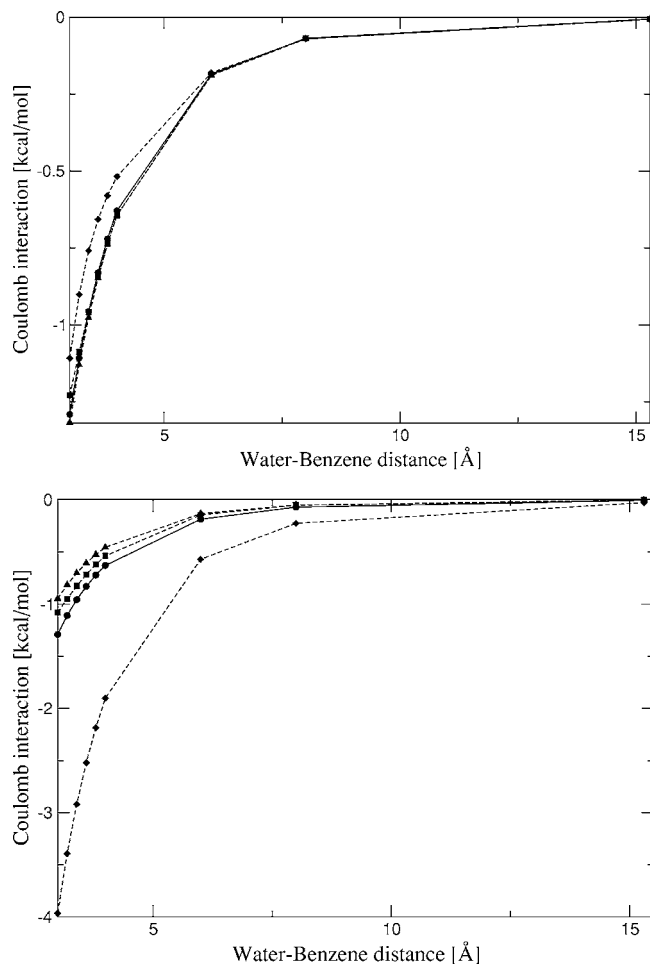

FIG. 7. Water-benzene dimer Coulomb interaction energies from Hermite (top) and multipoles (bottom) for a range of distances. Closed circles—CSOV; closed squares—A1 average; closed diamonds—P1 average; closed triangles—g03 average.

two additional points at very short range (1.5 and 1.6 Å) were also included to determine the accuracy of GEM when large penetration effects are expected. Here, as in the water dimers the accuracy at short and mid ranges goes as g03 $>$ P1 $>$ A1 for both averaged and single fit coefficients. For example, at an O-O distance of 2.3 Å the errors are $\approx$1, 0.2, and 0.1 kcal/mol out of 33 kcal/mol for A1, P1, and g03, respectively. These errors shrink rapidly as the distance increases.

The intermolecular interactions for the rotation of the formamide dimer show reasonable agreement with the reference calculations. Hermite coefficients obtained from the fit of a single fragment as well as averaged coefficients were used for these surface scans. As can be seen, in most cases there is almost no difference between the single fit and averaged coefficients. Strangely, the only noticeable difference is for the benzene dimer at close range (Fig. 6), 3.4–3.7 Å, for the average A1 coefficients which do not reproduce the curve accurately at these distances, whereas the single fit A1 coefficients do. We currently have no explanation for this difference.

The multipolar interactions for the 1D surface scans show that these converge to the correct result at long range.

In the case of the formamide dimer rotation, we obtain good results for the Hermite interactions and relatively good results for the multipolar interactions except for the P1 derived multipoles, where the distributed multipoles for benzene show a larger deviation from the CSOV results at short and medium ranges compared with the multipoles obtained with the A1 and g03 ABSs.

The 1D surface scan for a water-benzene dimer was calculated in order to test the transferability of the Hermite coefficients (see Fig. 7). All coefficients were fitted to reproduce the intermolecular Coulomb interaction of a single homodimer, as was explained in Sec. II G. Thus, testing the Hermite coefficients on a heterodimer provides a good test for their transferability. The curve shows good agreement for the A1 and g03 ABSs along the calculated distances. In the case of the P1 Hermite coefficients a deviation of around 0.2 kcal/mol was observed at close range. This is most likely due to the fact that the carbon P1 ABS is not accurate enough for our purposes. In the case of the multipoles from P1 a large deviation is observed resulting from the poor representation of the benzene multipoles. Another interesting feature of the multipole curve for the heterodimer is that the multipoles obtained from the A1 coefficients give reasonable agreement throughout the entire distance range.

TABLE II. Average (maximum) absolute errors in exchange-repulsion interaction with respect to CSOV for the ten water dimers (in kcal/mol).

6-31G*			aug-cc-PVTZ		
A1	P1	g03	A1	P1	g03
0.15(0.24)	0.18(0.55)	0.14(0.43)	0.41(0.72)	0.14(0.22)	0.21(0.79)

These results show that accurate intermolecular Coulomb energies can be obtained using the DF methodology. In general the g03 ABS shows the best agreement for all test cases. The A1 ABS also shows reasonable agreement, especially for the 1D surface scans. The P1 ABS gives very accurate results for water dimer interactions but appears to break down for intermolecular Coulomb calculations with fragments containing C atoms. This could be due to the fact that we have modified the P1 ABS such that all the d orbitals in the original ABS were replaced by spd orbitals.[18]

B. Intermolecular exchange-repulsion interactions

As shown in Table II, good agreement is observed for the exchange-repulsion interaction calculations for the ten dimers with respect to the reference CSOV values in general (see Supplementary Information for individual dimers). However, note that better results are obtained if the Hermite coefficients are fitted to reproduce the exchange-repulsion interaction for the A1 basis, which is the smallest ABS. The reason is that even though the fitted density calculated with the parameters for the Coulomb fit is good, the errors (see Supplementary Materials for color scales) extend further from the molecule than for the exchange-repulsion parameters, as can be seen in Fig. 8. This is in agreement with the fact that only the valence density is needed for the determination of the exchange repulsion using the overlap model, as explained by Wheatley and Price.[25] The exact and fitted density cubes for the comparison in Fig. 8 were obtained with density matrices from GAUSSIAN 98 at the B3LYP/6-31G* level.[70] Better results were also observed in the case of the P1 ABS with aug-cc-PVTZ by optimizing the Hermite coefficients for exchange repulsion. However, in this case the difference is much less: 0.14 kcal/mol average error with Hermite coefficients optimized for exchange repulsion (see Table II) compared with 0.19 kcal/mol average error for Hermite coefficients optimized for Coulomb. On the other

hand, for the rest of the cases, the same Hermite coefficients give the best Coulomb and exchange-repulsion interaction results.

The errors for the exchange-repulsion interactions presented in Table II are not as good as the ones obtained for GEM-0.[22] This is probably due to the fact that the training set for the linear regression in the calculation of the K parameter was significantly smaller in the present study than for the GEM-0 parametrization, only ten dimers in the present study compared to 190 random orientations. Nevertheless, note that the errors for the results obtained with the λ optimized for exchange repulsion are around 0.2–0.3 kcal/mol for the three ABSs when fitting 6-31G* density. The errors for the fit of the aug-cc-pVTZ are bigger; this is most likely due to the fact that the three ABSs used for this interaction are designed to work with double zeta basis sets as noted in Sec. III A.

The 1D homodimer surface scans are presented in Figs. 9–11. Both distance dependent scans (water and benzene dimers) show reasonable agreement with CSOV along a wide range of distances, especially for benzene. In the case of water dimer 1 (Fig. 9), the simple overlap model with fitted densities begins to break down at short to very short range. However it is important to note that there is strong dependence on the ABS quality, for example, the errors at 2.5 Å are around 30% for A1, 4% for P1, and 2% for g03. At 2.7 Å (hydrogen bonding distance) these errors are reduced to 10% for A1 and remain around 2%–3% for P1 and g03. For longer distances the errors are reduced to less than 2% (0.2 kcal/mol and below) for all auxiliary basis sets. The 1D homodimer scan for rotation (Fig. 10) shows that all three
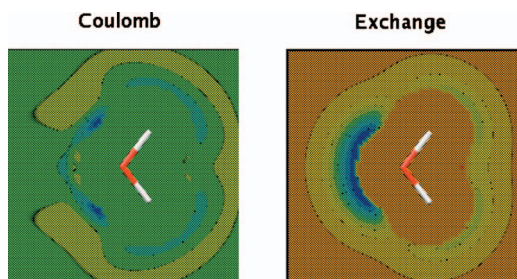

FIG. 8. (Color) Water molecule density difference maps for density fitting using A1 ABS with respect to *ab initio*. All calculations and fittings were done at the B3LYP/6-31G* level with a fine grid (120×120×120 points). All density fitting results were obtained using midpoints with O ABS.
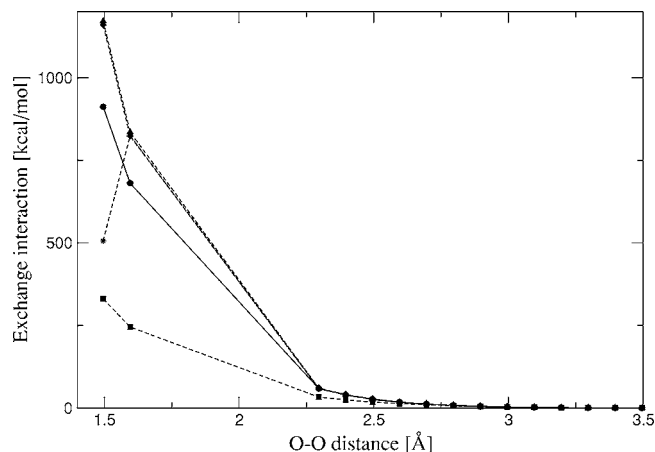

FIG. 9. Water dimer (structure 1) exchange-repulsion interaction energies for a range of distances. Closed circles—CSOV; stars—*ab initio* density; closed squares—A1 single; closed diamonds—P1 single; closed triangles—g03 single.
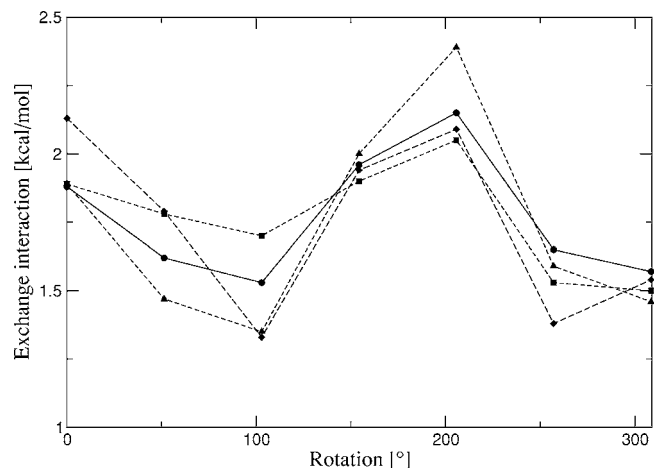
FIG. 10. Formamide exchange-repulsion interaction energies rotating about one of the fragments. Closed circles—CSOV; closed squares—A1 single; closed diamonds—P1 single; closed triangles—g03 single.

ABSs follow the shape of the reference curve. In all cases the largest errors of the intermolecular exchange interactions with respect to CSOV are around 0.2 kcal/mol.

Additionally, the intermolecular exchange repulsion for water dimer 1 has been calculated with the overlap model using the full *ab initio* density, as shown in Fig. 9. In this case the overlap model agrees very well with CSOV results in the range of 2.3–3.5 Å, with errors below 1%. At very short distances (1.6 Å and below) the errors increase substantially. Based on these results it can be seen that this simple model is very robust even at smaller intermolecular distances than those encountered in regular hydrogen bonding situations and the intermolecular exchange-repulsion interactions calculated with fitted densities will depend on the quality of the fitted densities and not on the accuracy of the overlap model.

As can be seen from Fig. 12, the exchange-repulsion interaction for the water-benzene heterodimer surface agrees with the reference curve. In all cases the average errors are below 0.15 kcal/mol. Note that in the case of the P1 ABS, very small negative values (>0.02 kcal/mol) are observed at

3.6 and 4.0 Å; this is due to the error on the benzene coefficients for this ABS. Also, in this case a small kink is observed in the CSOV curve, this is most likely due to the fact that the integration grid for the B3LYP calculation was not fine enough. This kink is also the reason why the R^2 obtained for the linear regressions to determine the K parameters are in the order of 0.95 instead of 0.99 as obtained for all the homodimer linear regressions (see Supplementary Materials). In spite of this we observe reasonable agreement with respect to the reference calculation.

C. Reciprocal space calculations

The implementation of GEM with the reciprocal space methods has been tested by calculating the intermolecular Coulomb energy and force for a series of water boxes, $(H_2O)_n$, $n=64$, 128, 256, 512, and 1024, in PBC. In practice in the molecular dynamics (md) community, the measure of the accuracy for a force field has been the force since this is the quantity that will determine the md trajectories. Additionally, there have been recent proposals for force field parametrization using force matching.[71,72]

Initially the accuracy of the forces calculated with GEM with three ABSs has been determined by comparing the calculated GEM forces with forces obtained with CSOV using the finite difference method. Table III shows the rms deviation for the ten water dimers between forces calculated with A1, P1, and g03 ABSs versus the CSOV forces from molecular density calculated with 6-31G* and aug-cc-pVTZ basis sets for both Coulomb and overlap interactions (see Supplementary Information for complete tables). In the case of the Coulomb interaction, as observed previously, the accuracy increases when better ABSs are employed for both basis sets. Furthermore, the rms between the CSOV forces obtained with the two different basis sets is 0.14, compared with the largest rms at the 6-31G* level obtained with A1 of 0.06 (see Table III). In the case of the aug-cc-pVTZ densities, A1 shows a slightly larger rms (0.15); however, as explained previously, all three ABSs are designed to work with double
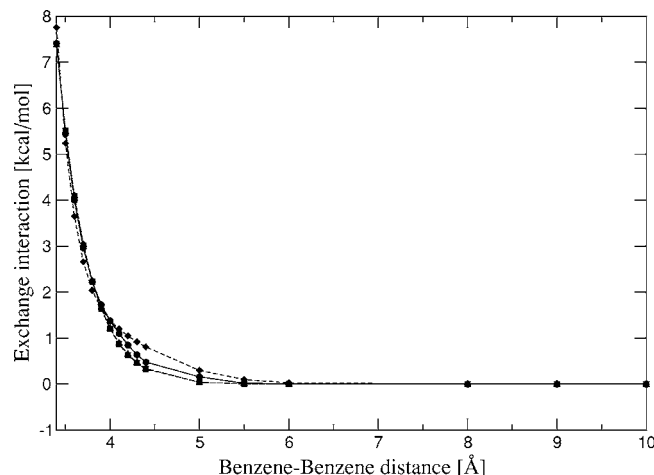

FIG. 11. Stacked benzene dimer exchange-repulsion interaction energies for a range of distances. Closed circles—CSOV; closed squares—A1 single; closed diamonds—P1 single; closed triangles—g03 single.
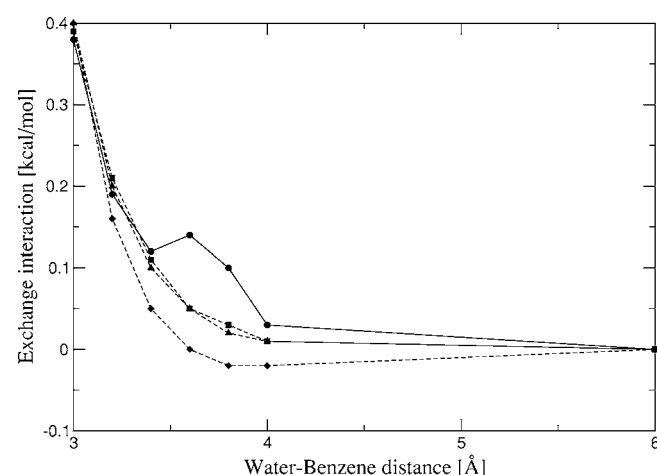

FIG. 12. Water-benzene dimer exchange-repulsion interaction energies for a range of distances. Closed circles—CSOV; closed squares—A1 average; closed diamonds—P1 average; closed triangles—g03 average.

TABLE III. Relative rms force deviation with respect to CSOV for the ten water dimers.

	6-31G*			aug-cc-pVTZ		
	A1	P1	g03	A1	P1	g03
Coulomb	0.06	0.03	0.01	0.15	0.04	0.05
Exchange	0.22	0.08	0.07	0.12	0.04	0.07

zeta basis. Both P1 and g03 show slightly larger rms with respect to CSOV with the larger basis set although it is still well below the CSOV one.

For the overlap forces the CSOV rms deviation obtained between 6-31G* and aug-cc-pVTZ is considerably larger, 0.26. In this case the rms for all ABSs using both basis sets is below the CSOV rms. Note that, as was the case above, the results for overlap are not as good as for Coulomb interactions. Also note that in this case the use of reciprocal space methods for the overlap determination results only in a modest improvement in computational efficiency. For example, the overlap energy force for the 256 water box to 10^{-4} rms force tolerance with the P1 ABS calculating all integrals in direct space takes 1.88 s; on the other hand, when using reciprocal space methods the time is reduced to 1.31 s.

Figures 13 and 14 show the CPU time for a single energy-force Coulomb calculation for the water boxes in PBC using relative rms force tolerances of 10^{-3} and 10^{-4} respectively. These tolerances were chosen to be well below the intrinsic accuracy of the ABS fits of 1%–10%. All calculations were performed on a single CPU of a dual Xeon 3.3 GHz computer with 3 Gbytes of memory. In both cases the A1 and P1 ABSs show a quasilinear behavior for both the PME and FFP methods for all boxes, while g03 gives a quasilinear behavior for FFP and linear behavior for PME up to 512 waters and quadratic afterward. Note first that unlike previous PME implementations, in the present case the reciprocal space calculations dominate the compute time. Thus the irregularities in feasible FFT grid dimensions (restricted to products of powers of 2, 3, and 5) result in nonsmooth timings as a function of system size. Furthermore, currently

the code does not allow for cutoffs bigger than 1/2 the box size which limits timing optimization in the 64 and 128 water boxes. The quadratic behavior is most probably due to an inefficient use of memory in our code due to the large Hermite-Hermite interaction list which is kept in memory, as well as the large number of FFTs required. We have decided not to report results obtained with the Ewald method since it is too slow [$O(N^2)$ scaling] compared to FFT and PME. Interestingly, from the results it is observed that the advantage of PME widens for more accurate results (10^{-4} rms) compared to FFP because in this case it is difficult to increase accuracy with the Gaussian splitting method and the increase of accuracy in FFP requires a better sampling of the tails of the Gaussians. Also, note that in the case of the Coulomb calculation, in order to achieve a relative rms of 10^{-4} the rms accuracy of the individual components (*e*-*e* and *N*-*e*) needs to be equivalent to 5×10^{-8} due to the fact that a subtraction between very large numbers has to be performed. For example, for dimer 1 the *e*-*e*, *N*-*N*, and *N*-*e* energies are approximately 18.285, 18.325, and −36.624 hartree compared with −8.32 kcal/mol for the total Coulomb interaction.

The reciprocal space methods are quite efficient. The calculation of the energy and force for the largest system tested at the highest accuracy, i.e., the 1024 water box with g03 ABS, which corresponds to 654 336 Hermite coefficients, takes only 34 s with FFP and 42 s with PME (see Fig. 14). Moreover, note that both reciprocal space methods are highly parallelizable, which would increase the computational efficiency. Additionally, we have not explored the possibility of using FFTW for PME which would allow the calculation of multiple FFTs at once, as well as a thorough
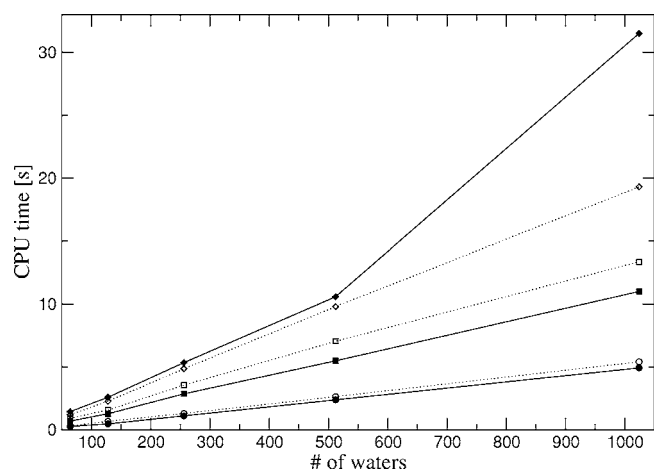

FIG. 13. Timings for water boxes with rms force tolerance of 10^{-3}. Closed circles—A1 PME; closed squares—P1 PME; closed diamonds—g03 PME; open circles—A1 FFP; open squares—P1 FFP; open diamonds—g03 FFP.
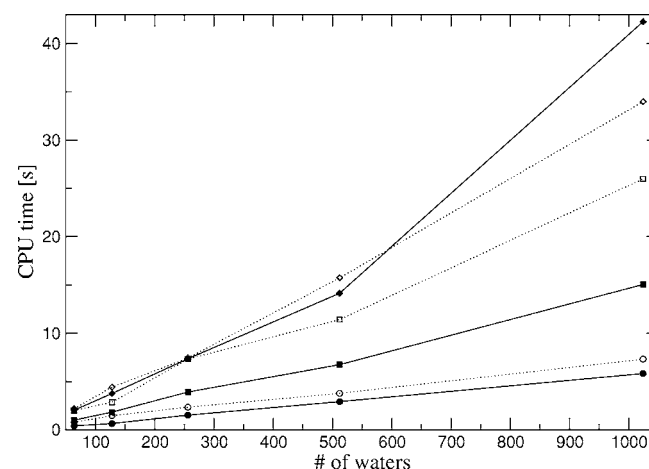

FIG. 14. Timings for water boxes with rms force tolerance of 10^{-4}. Closed circles—A1 PME; closed squares—P1 PME; closed diamonds—g03 PME; open circles—A1 FFP; open squares—P1 FFP; open diamonds—g03 FFP.

optimization of all the possible parameters involved in the FFP implementation with Gaussian splitting. Another factor that could be used to decrease the FFP computing time is the addition of a third level coarse grid when the ABSs have very diffuse auxiliary basis. Finally, it is important to mention that results obtained with both PME and FFP can be mixed (data not shown); therefore this opens up interesting possibilities such as a QM/MM implementation using GEM,[45] where the GEM section proximal to the QM can be calculated with PME or FFP, while the rest of the MM subsystem can be represented via GEM multipoles and calculated in reciprocal space with PME (work in progress).

IV. CONCLUSIONS

The Gaussian electrostatic model has been extended for use of Hermite Gaussian fitting functions of arbitrary angular momentum. In addition to increased computational speed by the use of the generalized MD recursion, the use of Hermite Gaussians also gives rise to the determination of site multipoles obtained directly from the fitting procedure. Noise problems for the fit of the expansion coefficients have been addressed by using the Tikhonov regularization method as well as by means of the damped Coulomb kernel. The use of a local reference frame for each fitting site allows the rotation of the fitting coefficients in order to perform intermolecular interaction calculations. Frozen-core intermolecular contributions (Coulomb and exchange repulsion) have been calculated for a series of systems including both homo- and heterodimers. Three different fitting basis sets were tested for accuracy. Reasonable agreement is obtained in all cases for the intermolecular interactions with errors below 0.1 kcal/mol for electrostatic and around 0.15–0.2 kcal/mol for exchange repulsion. Additionally, a significant computational speedup is achieved by using reciprocal space methods. Three such methods were implemented, FFP, PME, and regular Ewald. Results for a series of water boxes in PBC using rms accuracies of 10^{-3} and 10^{-4} have been obtained and the CPU timings have been reported. For example, a 1024 water box using the g03 ABS (the largest system and ABS studied) takes 42 s for an energy-force calculation with PME and 34 s with FFP.

[1] D. A. Case, T. E. Cheatham III, T. A. Darden, H. Gohlke, R. Luo, K. M. Merz, Jr., A. Onufirev, C. Simmerling, B. Wang, and R. J. Woods, J. Comput. Chem. **26**, 1668 (2005).
[2] W. L. Jorgensen and J. Tirado-Rives, J. Comput. Chem. **26**, 1669 (2005).
[3] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhoff, A. E. Mark, and H. J. C. Berensen, J. Comput. Chem. **26**, 1701 (2005).
[4] M. Christen, P. H. Hünenberger, D. Bakowies *et al.*, J. Comput. Chem. **26**, 1719 (2005).
[5] A. D. MacKerrell, Jr., B. Brooks, C. L. Brooks III, N. B. Roux, Y. Won, and M. Karplus, *Encyclopedia of Computational Chemistry* (Wiley, New York, NY, 1998).
[6] A. R. Leach, *Molecular Modelling: Principles and Applications*, 2nd ed. (Prentice-Hall, Harlow, UK, 2001).
[7] A. J. Stone, *The Theory of Intermolecular Forces* (Oxford University Press, Oxford, UK, 2000).
[8] A. J. Stone, Chem. Phys. Lett. **83**, 233 (1981).
[9] F. Vigné-Maeder and P. Claverie, J. Chem. Phys. **88**, 4934 (1988).
[10] G. Náray-Szabó and G. G. Ferenczy, Chem. Rev. (Washington, D.C.) **95**, 829 (1995).
[11] N. Gresh, P. Claverie, and A. Pullman, Int. J. Quantum Chem. **22**, 253 (1979).
[12] N. Gresh, J. Comput. Chem. **16**, 856 (1995).
[13] P. Ren and J. W. Ponder, J. Phys. Chem. B **107**, 5933 (2003).
[14] M. A. Freitag, M. S. Gordon, J. H. Jensen, and W. J. Stevens, J. Chem. Phys. **112**, 7300 (2000).
[15] J.-P. Piquemal, N. Gresh, and C. Giessner-Prettre, J. Phys. Chem. A **107**, 10353 (2003).
[16] N. Gresh, J.-P. Piquemal, and M. Krauss, J. Comput. Chem. **26**, 1113 (2005).
[17] A. Gavezzotti, J. Phys. Chem. B **106**, 4145 (2002).
[18] G. A. Cisneros, J.-P. Piquemal, and T. A. Darden, J. Chem. Phys. **123**, 044109 (2005).
[19] S. F. Boys and I. Shavit, *A Fundamental Calculation of the Energy Surface for the System of Three Hydrogens Atoms* (NTIS, Springfield, VA, 1959), AD212985.
[20] B. I. Dunlap, J. W. D. Connolly, and J. R. Sabin, J. Chem. Phys. **71**, 4993 (1979).
[21] A. M. Köster, P. Calaminici, Z. Gómez, and U. Reveles, *Reviews of Modern Quantum Chemistry: A Celebration of the Contribution of Robert G. Parr* (World Scientific, Singapore, 2002).
[22] J.-P. Piquemal, G. A. Cisneros, P. Reinhardt, N. Gresh, and T. A. Darden, J. Chem. Phys. **124**, 104101 (2006).
[23] P. S. Bagus, K. Hermann, and C. W. Bauschlicher, Jr., J. Chem. Phys. **80**, 4378 (1984).
[24] J.-P. Piquemal, A. Marquez, O. Parisel, and C. Giessner-Prettre, J. Comput. Chem. **26**, 1052 (2005).
[25] R. J. Wheatley and S. L. Price, Mol. Phys. **69**, 507 (1990).
[26] C. Domene, P. W. Fowler, M. Wilson, P. Madden, and R. J. Wheatley, Chem. Phys. Lett. **333**, 403 (2001).
[27] G. S. Tschumper, M. L. Leininger, B. C. Hoffman, E. F. Valeev, H. F. Schaffer III, and M. Quack, J. Chem. Phys. **116**, 690 (2002).
[28] J. G. C. M. van Duijneveldt–van de Rijdt, W. T. M. Mooij, and F. B. Duijneveldt, Phys. Chem. Chem. Phys. **5**, 1169 (2003).
[29] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran77: The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, New York, NY, 1992).
[30] Y. Jung, A. Sodt, P. M. W. Gill, and M. Head-Gordon, Proc. Natl. Acad. Sci. U.S.A. **102**, 6692 (2005).
[31] T. Helgaker, P. J. Jørgensen, and J. Olsen, *Molecular Electronic-Structure Theory* (Wiley, Chichester, England, 2000).
[32] A. M. Köster, J. Chem. Phys. **118**, 9943 (2003).
[33] L. E. McMurchie and E. R. Davidson, J. Comput. Phys. **26**, 218 (1978).
[34] M. Tachikawa and M. Shiga, Phys. Rev. E **64**, 056707 (2001).
[35] S. B. Trickey, J. A. Alfort, and J. C. Boettger, *Computational Materials Science* (Elsevier B. V., Amsterdam, 2004).
[36] C. Sagui, L. G. Pedersen, and T. A. Darden, J. Chem. Phys. **120**, 73 (2004).
[37] K. Eichkorn, O. Treutler, H. Öhm, M. Häser, and R. Ahlrichs, Chem. Phys. Lett. **240**, 283 (1995).
[38] A. M. Köster, J. Chem. Phys. **104**, 4114 (1996).
[39] R. Podeszwa, R. Bukowski, and K. Szalewicz, J. Chem. Theory Comput. **2**, 400 (2006).
[40] A. J. Misquitta and A. J. Stone, J. Chem. Phys. **124**, 024111 (2006).
[41] G. Arfken and H. Weber, *Mathematical Methods for Physicists* (Academic, San Diego, CA, 1996).
[42] A. Toukmaji, C. Sagui, J. Board, and T. A. Darden, J. Chem. Phys. **113**, 10913 (2000).
[43] J. G. Ángyán, C. Chipot, F. Dehez, C. Hätig, G. Jansen, and C. Millot, OPEP, a tool for the optimal partitioning of electric properties, Version 1.0-β, Equipe de Chimie and Biochimie Théoretiques, Unité Mixte de Recherche CNRS/UHP No. 7565, Université Henri Poincaré, France, 2005.
[44] M. Challacombe, E. Schwgler, and J. Almlöf, *Computational Chemistry: Review of Current Trends* (World Scientific, Singapore, 1996).
[45] G. A. Cisneros, J.-P. Piquemal, and T. A. Darden, J. Phys. Chem. B **110**,

13682 (2006).
[46] A. J. Stone, J. Chem. Theory Comput. **1**, 1128 (2005).
[47] P. L. A. Popelier, L. Joubert, and D. S. Kosov, J. Phys. Chem. A **105**, 8524 (2001).
[48] A. Volkov and P. Coppens, J. Comput. Chem. **25**, 921 (2004).
[49] E. Clementi, *Computational Aspects of Large Chemical Systems* (Springer, New York, NY, 1980).
[50] M. Dupuis, A. Marquez, and E. R. Davidson, HONDO95.3, quantum chemistry program exchange (QCPE), Indiana University, Bloomington, IN, 1995.
[51] L. Fusti-Molnar and P. Pulay, J. Chem. Phys. **117**, 7827 (2002).
[52] D. Lippert, J. Hutter, and M. Parrinello, Theor. Chem. Acc. **103**, 124 (1999).
[53] M. Krack and M. Parrinello, Phys. Chem. Chem. Phys. **2**, 2105 (2000).
[54] C. Cohen-Tannoudji, B. Diu, and F. Laloe, *Quantum Mechanics* (Wiley Interscience, New York, NY, 1977).
[55] R. Ahlrichs, Phys. Chem. Chem. Phys. **8**, 3072 (2006).
[56] U. Essmann, L. Perera, M. Berkowitz, T. A. Darden, H. Lee, and L. G. Pedersen, J. Chem. Phys. **103**, 8577 (1995).
[57] D. York and W. Yang, J. Chem. Phys. **101**, 3298 (1994).
[58] Y. B. Shan, J. L. Klepis, M. P. Eastwood, R. O. Dror, and D. E. Shaw, J. Chem. Phys. **122**, 054101 (2005).
[59] E. R. Smith, J. Stat. Phys. **77**, 449 (1994).
[60] C. Sagui and T. A. Darden, J. Chem. Phys. **114**, 6578 (2001).
[61] R. D. Amos, I. L. Alberts, J. S. Andrews, *et al.*, CADPAC, the Cambridge analytic derivatives package, Issue 6, Cambridge University, Cambridge, UK, 1995.
[62] A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).
[63] C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
[64] J. Andzelm and E. Wimmer, J. Chem. Phys. **96**, 1280 (1992).

[65] N. Godbout and J. Andzelm, DGAUSS, Versions 2.0, 2.1, 2.3, and 4.0, the file that contains the A1 and P1 auxiliary basis sets can be obtained from the CCL WWW site at http://www.ccl.net/cca/data/basis-sets/DGauss/index.shtml, Computational Chemistry List, Ltd., Columbus, OH, 1998.
[66] M. J. Frisch, G. W. Trucks, H. B. Schlegel *et al.*, GAUSSIAN 03, Gaussian, Inc., Pittsburgh, PA, 2003.
[67] D. E. Berholdt and R. J. Harrison, J. Chem. Phys. **109**, 1593 (1998).
[68] W. Jorgensen, J. Chandrasekhar, J. Madura, R. Impey, and M. Klein, J. Chem. Phys. **79**, 926 (1983).
[69] J. Ponder, TINKER, software tools for molecular design, Version 3.6, the most updated version for the TINKER program can be obtained from J. Ponder's WWW site at http://dasher.wustl.edu/tinker, Washington University, St. Louis, MO, 1998.
[70] M. J. Frisch, G. W. Trucks, H. B. Schlegel *et al.*, GAUSSIAN 98, Revision A.8, Gaussian, Inc., Pittsburgh, PA, 1998.
[71] R. Ercolessi and J. B. Adams, Europhys. Lett. **26**, 583 (1994).
[72] A. Laio, S. Bernard, C. L. Chiarotti, S. Scandolo, and E. Tosatti, Science **287**, 1027 (2000).
[73] Also, it is important to note that we rely on the exponents of the diffuse functions to be smaller than α_0 to ensure uniform convergence in the sums of Eq. (29). That is, if $1/\alpha_d > 1/\alpha_0$ these sums converge uniformly.
[74] See EPAPS Document No. E-JCPSA6-125-309641 for the optimal fitting parameters for all molecules tested in this study; intermolecular Coulomb and exchange-repulsion results for individual water dimers; K parameters for exchange repulsion for all fragments; scales for density comparison cubes; Coulomb and overlap forces for CSOV and the three ABSs. This document can be reached via a direct link in the online article's HTML reference section or via the EPAPS homepage (http://www.aip.org/pubservs/epaps.html).