

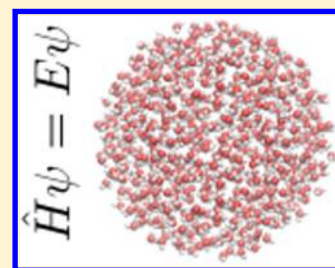
# A Variational Linear-Scaling Framework to Build Practical, Efficient Next-Generation Orbital-Based Quantum Force Fields

Timothy J. Giese, Haoyuan Chen, Thakshila Dissanayake, George M. Giambaşu, Hugh Heldenbrand, Ming Huang, Erich R. Kuechler, Tai-Sung Lee, Maria T. Panteva, Brian K. Radak, and Darrin M. York\*

BioMaPS Institute and Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, New Jersey 08854-8087, United States

## S Supporting Information

**ABSTRACT:** We introduce a new hybrid molecular orbital/density-functional modified divide-and-conquer (mDC) approach that allows the linear-scaling calculation of very large quantum systems. The method provides a powerful framework from which linear-scaling force fields for molecular simulations can be developed. The method is variational in the energy and has simple, analytic gradients and essentially no break-even point with respect to the corresponding full electronic structure calculation. Furthermore, the new approach allows intermolecular forces to be properly balanced such that nonbonded interactions can be treated, in some cases, to much higher accuracy than the full calculation. The approach is illustrated using the second-order self-consistent charge density-functional tight-binding model (DFTB2). Using this model as a base Hamiltonian, the new mDC approach is applied to a series of water systems, where results show that geometries and interaction energies between water molecules are greatly improved relative to full DFTB2. In order to achieve substantial improvement in the accuracy of intermolecular binding energies and hydrogen bonded cluster geometries, it was necessary to extend the DFTB2 model to higher-order atom-centered multipoles for the second-order self-consistent intermolecular electrostatic term. Using generalized, linear-scaling electrostatic methods, timings demonstrate that the method is able to calculate a water system of 3000 atoms in less than half of a second, and systems of up to 1 million atoms in only a few minutes using a conventional desktop workstation.



## 1. INTRODUCTION

There is great interest in the scientific community in developing molecular simulation force fields that are able to explicitly treat electronic degrees of freedom in order to model chemical and biochemical phenomena that are inherently related to electronic structure. Example applications include simulations of reactive chemical events such as catalysis,<sup>1</sup> photochemistry, and electron transfer processes,<sup>2–4</sup> as well as the calculation of spectroscopic observables to aid in the interpretation of experiments.<sup>5,6</sup> Toward that end, quantum chemical methods are required that are fast and accurate and can scale to the large system sizes and long time scales required by the applications. The scaling requirement has been a long-standing arena of active research effort in the development of so-called “linear-scaling”— $O(N)$  or  $O(N \log N)$ —electronic structure methods, for which much success has been achieved.<sup>7–13</sup>

Nonetheless, there remain challenges that prevent these established linear-scaling electronic structure methods from being used as practical tools, and having broad impact, in truly large-scale molecular dynamics simulations. First, despite having favorable scaling properties, the computational cost of conventional linear-scaling methods is typically still prohibitive for long-time simulations of large systems, although some applications are beginning to emerge.<sup>14–21</sup> This is related to the fact that these methods typically have break-even points of a few hundred atoms, only after which do the methods have any computational advantage relative to their full nonlinear-scaling counterparts.<sup>22</sup>

Second, for condensed phase simulations, methods need to accurately treat both strong intramolecular (i.e., bonding) interactions and subtle, typically weak intermolecular (i.e., non-bonding) interactions. The latter needs to be delicately balanced such that bulk properties, binding affinities, and solvation free energies are treated accurately and consistently. This balance remains a challenge for modern density-functional methods, even without consideration of extension to linear-scaling.<sup>23</sup> Third, to make these methods practical for molecular dynamics simulations, rigorous analytic gradients are needed that require only relatively minor overhead relative to the calculation of the energy itself.<sup>21,24</sup>

In this work, we introduce a new modified divide-and-conquer (mDC) approach that allows the linear-scaling calculation of very large quantum systems, including analytic gradients, with essentially no break-even point. The extended atomic orbital (AO) buffer space used in traditional divide-and-conquer (DC) is replaced by a density-overlap intermolecular interaction model<sup>25</sup> that is treated self-consistently and can be tuned for high accuracy. The approach is illustrated using the second-order self-consistent field density-functional tight-binding (SCC-DFTB or DFTB2) model.<sup>26</sup> Using DFTB2 as a base Hamiltonian, the new mDC approach is applied to a series of water systems, where results show that geometries and interaction energies between

Received: November 17, 2012

Published: January 9, 2013

water molecules are greatly improved relative to full DFTB2, whereas timings demonstrate that the method may be practical for large-scale molecular dynamics (MD) simulations. The method is demonstrated to be highly scalable and able to calculate systems of up to 1 million atoms in only a few minutes using a conventional desktop workstation.

## 2. METHODS

**2.1. Essential Background.** Next generation MO-based force fields are founded upon linear-scaling electronic structure theory to reduce the scaling bottlenecks associated with the underlying *ab initio* model; however, unlike traditional linear-scaling electronic structure methods, the Hamiltonian form is modified with empirical potentials and parameters to increase performance and make the resulting method more accurate. The concept of using electronic structure theory explicitly as a “next generation quantum force field” was introduced by Gao,<sup>27</sup> who demonstrated its accuracy for liquid simulations<sup>28</sup> and explored its feasibility for use in treating biological molecules.<sup>29</sup> There are other approaches, such as SIBFA<sup>30,31</sup> and GEM,<sup>30,32–34</sup> that may also be considered quantum force fields but do not explicitly use MOs. These density-based models approximate the exchange-repulsion between fragments using a density-overlap model.<sup>35</sup> The present work makes explicit use of MOs, like those models described by Gao, and treats the inter-region coupling with a density-overlap model that is influenced by the SIBFA and GEM methods.

Consider the case of a quantum mechanical method based on a single determinant wave function that is variationally optimized to minimize the total electronic energy (e.g., Hartree–Fock and most modern density-functional or semiempirical quantum methods fit this category). The two most expensive mathematical operations performed during the self-consistent field (SCF) procedure on large systems are the construction of the Fock matrix elements and its diagonalization. Several methods have been introduced for computing the Fock matrix with  $O(N \log N)$  effort.<sup>36–39</sup> The dominant part of this effort is spent performing the electrostatic interactions<sup>39–46</sup> because of their significance even between well-separated atoms. Most methods for overcoming the  $O(N^3)$ -scaling of the Fock matrix diagonalization are exploitations of the sparse matrices that arise due to the negligible overlap between well-separated atoms. As a simple illustration of how this property reduces the scaling of the diagonalization, consider a system composed of two regions *A* and *B*. When the two regions are sufficiently separated, their inter-region overlap is negligibly small, and the inter-region Fock matrix elements necessarily vanish. In this case, the wave function of the system can then be accurately approximated by a Hartree product of antisymmetrized determinants of molecular fragments.<sup>27</sup> The eigenvalues **E** and eigenvectors **C** of the system then satisfy

$$\begin{pmatrix} \mathbf{C}^A & 0 \\ 0 & \mathbf{C}^B \end{pmatrix}^T \begin{pmatrix} \mathbf{F}^A & 0 \\ 0 & \mathbf{F}^B \end{pmatrix} \begin{pmatrix} \mathbf{C}^A & 0 \\ 0 & \mathbf{C}^B \end{pmatrix} = \begin{pmatrix} \mathbf{E}^A & 0 \\ 0 & \mathbf{E}^B \end{pmatrix} \quad (1)$$

where  $\mathbf{F}^A \cdot \mathbf{C}^A = \mathbf{S}^A \cdot \mathbf{C}^A \cdot \mathbf{E}^A$ ,  $\mathbf{S}^A$  is the intraregion overlap, and region *B* diagonalizes analogously. In other words, diagonalization of the full system is equivalent to the diagonalization of the smaller systems. The scaling thus reduces to  $O(NK^3)$ , where *K* is the size of a single region. This does not imply that the systems are isolated, however.  $\mathbf{F}^A$  and  $\mathbf{F}^B$  are the nonzero diagonal blocks of the Fock matrix, and these include the potentials arising from

both regions, which is longer-ranged than the inter-region overlap.

When the two regions above are not well-separated, they remain coupled by inter-region overlap, and one must concoct a scheme to account for it. For example, the method used in the X-Pol model is to replace the explicit coupling with empirical Lennard-Jones or Buckingham potentials<sup>29,47–53</sup> or through perturbative corrections.<sup>54,55</sup> The Fragment Molecular Orbital (FMO) method<sup>56–62</sup> computes the variational energy of decoupled regions and then reintroduces the inter-region coupling via many-body expansion corrections solved in the presence of the fixed electrostatic potential of the other regions as determined from the decoupled calculation. This makes the FMO total energy *nonvariational* and thus complicates the implementation of its analytic gradients.<sup>57</sup> Other fragment-based methods exist that treat the coupling with many-body perturbative expansions.<sup>63–65</sup> The divide-and-conquer (DC) scheme<sup>16,17,19,22,24,66–82</sup> accounts for the coupling by explicitly extending the AO space of the region to include its near-neighbor regions, called its “buffer.” The resulting molecular orbitals (MOs) extend beyond the region and into its buffer. Special care must then be taken to partition the resulting electron density to avoid its double counting when the region is treated as its neighbor’s buffer. The fragment density functional method is similar to DC: each region is surrounded by a buffer, and the system outside of the buffer is treated by empirical point charges.<sup>51,83</sup>

Rigorous analytic gradients of buffered DC require the solution of the coupled perturbed Hartree–Fock/Kohn–Sham equations of the entire system.<sup>24</sup> The solution of these equations is difficult to achieve in a linear-scaling fashion. Fortunately, the approximate analytic formulas for the buffered DC method are sufficient so as long as the buffer is large ( $\approx 4$  to  $5$  Å) but produce unacceptably large errors when the size of the buffer is small. The mDC method described in the present work circumvents the need to solve the coupled-perturbed equations by replacing the explicit buffer space with a density-overlap intermolecular interaction model. Without an explicit buffer, the gradient expressions for each region are no more difficult than those encountered in traditional single-determinant methods.

**2.2. The mDC Energy.** In the context of mDC, we consider a “region” to be a fragment of the system whose AOs are chosen to be completely localized on that fragment. In this manner, the Fock matrix for a fragment can be diagonalized to produce its own set of local MOs. The mDC total energy is then described as the sum of the energy of each region  $E^A$  and the inter-region interactions.

$$E = \sum_A E^A(\mathbf{P}^{\alpha,A}, \mathbf{P}^{\beta,A}, \mathbf{q}^A; \mathbf{R}^A) + J(\mathbf{q}; \mathbf{R}) - \sum_A J(\mathbf{q}^A; \mathbf{R}^A) + \sum_{B \neq A} \sum_{a \in A} \sum_{b \in B} E_{\text{inter},ab}(\mathbf{q}_a, \mathbf{q}_b; R_{ab}) \quad (2)$$

where *A* and *B* index regions, *a* and *b* index atoms,  $\alpha$  and  $\beta$  denote electron spin, **R** is the full set of atomic coordinates, **q** is the full set atomic multipole moments, and  $\mathbf{q}_a$  and  $\mathbf{q}^A$  are the multipole moments of atom *a* and the collection of atomic multipole moments in region *A*, respectively.

$$\mathbf{P}_{ij}^{\sigma,A} = \sum_{k \in A} n_k^{\sigma,A} C_{ik}^{\sigma,A} C_{jk}^{\sigma,A} \quad (3)$$

is the single-particle spin-resolved density matrix of region  $A$ ;  $n_k^{\sigma,A}$  is the spin-resolved occupation number of orbital  $k$  in region  $A$ ;  $C_{ik}^{\sigma,A}$  is the  $i$ th MO coefficient of orbital  $k$  in region  $A$ , which satisfies

$$\mathbf{F}^{\sigma,A} \cdot \mathbf{C}^{\sigma,A} = \mathbf{S}^A \cdot \mathbf{C}^{\sigma,A} \cdot \mathbf{E}^{\sigma,A} \quad (4)$$

The multipole moments of atom  $a$  are

$$q_{l_a m_a} = Z_a \delta_{l_a,0} \delta_{m_a,0} - \int \rho_a(\mathbf{r}) C_{l_a m_a}(\mathbf{r} - \mathbf{R}_a) d^3 r \quad (5)$$

where  $Z_a$  is the core nuclear charge of atom  $a$ ,  $C_{lm}(\mathbf{r})$  is a real regular solid harmonic, and  $\rho_a(\mathbf{r})$  is the partitioned atomic density of atom  $a$ , which shall be discussed in more detail later.

$$J(\mathbf{q}; \mathbf{R}) = \frac{1}{2} \mathbf{q}^T \cdot \mathbf{J} \cdot \mathbf{q} \quad (6)$$

is the Coulomb electrostatic energy between point-multipoles, i.e.,

$$J_{(l_a m_a), (l_b m_b)} = \int \int \frac{\delta_{l_a m_a}(\mathbf{r} - \mathbf{R}_a) \delta_{l_b m_b}(\mathbf{r}' - \mathbf{R}_b)}{|\mathbf{r} - \mathbf{r}'|} d^3 r d^3 r' \quad (7)$$

where

$$\delta_{lm}(\mathbf{r} - \mathbf{R}_a) = \frac{C_{lm}(\nabla_a)}{(2l-1)!!} \delta(\mathbf{r} - \mathbf{R}_a) \quad (8)$$

is a point-multipole function,  $C_{lm}(\nabla_a)$  is a spherical tensor gradient operator<sup>84</sup> acting on the coordinates of atom  $a$ , and  $\delta(\mathbf{r})$  is a Dirac-delta function.

$E^A(\mathbf{P}^{\alpha,A}, \mathbf{P}^{\beta,A}, \mathbf{q}^A; \mathbf{R}^A)$  is the quantum energy of region  $A$ , which for this work is the DFTB2 model described in numerous articles.<sup>26,85–88</sup> In brief,

$$\begin{aligned} E_{\text{DFTB2}}(\mathbf{P}^A, \mathbf{q}^A; \mathbf{R}^A) \\ = \sum_{b>a} E_{\text{rep}}(R_{ab}) + \sum_{ij} P_{ij}^A H_{ij} + \frac{1}{2} \sum_{ab} q_{0,0,a} q_{0,0,b} \gamma_{ab}(R_{ab}) \end{aligned} \quad (9)$$

where  $\mathbf{P}^A = \mathbf{P}^{\alpha,A} + \mathbf{P}^{\beta,A}$  is the total density matrix,  $\mathbf{H}$  is the tight-binding matrix constructed from two-center splines and one-center parameters,  $E_{\text{rep}}(R_{ab})$  is a pairwise repulsive potential stored on splines, and

$$\gamma_{ab}(R_{ab}) = \int \int \frac{\varphi_a^s(\mathbf{r} - \mathbf{R}_a; \xi_a) \varphi_b^s(\mathbf{r} - \mathbf{R}_b; \xi_b)}{|\mathbf{r} - \mathbf{r}'|} d^3 r d^3 r' \quad (10)$$

is the Coulomb interaction between the auxiliary Slater monopole representation of the response density, i.e.,

$$\varphi_a^s(\mathbf{r}; \xi_a) = \frac{\xi_a^3}{8\pi} e^{-\xi_a r} \quad (11)$$

In this work,  $E_{\text{inter},ab}(\mathbf{q}_a, \mathbf{q}_b; R_{ab})$  is chosen to be the OPNQ van der Waals (vdW) model described in ref 25, where

$$\begin{aligned} E_{\text{vdW}}(\mathbf{q}_a, \mathbf{q}_b; R_{ab}) &= E_{\text{exch}}(q_{0,0,a}, q_{0,0,b}; R_{ab}) \\ &+ E_{\text{disp}}(q_{0,0,a}, q_{0,0,b}; R_{ab}) \end{aligned} \quad (12)$$

The exchange energy

$$\begin{aligned} E_{\text{exch}}(q_a, q_b, R_{ab}) &= s_a s_b \int \varphi_a^s(\mathbf{r} - \mathbf{R}_a; \varsigma_a(q_a)) \\ &\times \varphi_b^s(\mathbf{r} - \mathbf{R}_b; \varsigma_b(q_b)) d^3 r \end{aligned} \quad (13)$$

is the overlap of Slater monopoles with a charge-dependent exponent

$$\varsigma_a(q_a) = \varsigma_a(0) e^{-h_a q_a} \quad (14)$$

where  $s_a$ ,  $\varsigma_a(0)$ , and  $h_a$  are parameters. The dispersion energy is

$$E_{\text{disp}}(q_a, q_b, R_{ab}) = -S_6(b_{ab}, R_{ab}) \frac{C_6(q_a, q_b)}{R_{ab}^6} \quad (15)$$

where  $b_{ab} \equiv b_{ab}(q_a, q_b, R_{ab})$

$$b_{ab}(q_a, q_b, R_{ab}) = -\frac{d}{dR_{ab}} \ln E_{\text{exch}}(q_a, q_b, R_{ab}) \quad (16)$$

is the Tang and Toennies (TT) damping exponent,

$$C_6(q_a, q_b) = \frac{3}{2} \frac{\eta_a(q_a) \eta_b(q_b)}{\eta_a(q_a) + \eta_b(q_b)} \alpha(q_a) \alpha(q_b) \quad (17)$$

is a charge-dependent dispersion coefficient,

$$\eta_a(q_a) = \sqrt{\frac{N_{\text{eff},a}(q_a)}{\alpha_a(q_a)}} \quad (18)$$

$$N_{\text{eff},a}(q_a) = \begin{cases} N_{\text{val},a}(q_a) N_{\text{eff},a}(0) / N_{\text{val},a}(0), & \text{if } N_{\text{val},a}(q_a) > 0 \\ 0, & \text{otherwise} \end{cases} \quad (19)$$

is the effective number of valence electrons,

$$N_{\text{val},a}(q_a) = N_{\text{val},a}(0) - q_a \quad (20)$$

is the number of valence electrons and  $N_{\text{val},a}(0)$  is the number of valence electrons in the neutral atom,

$$\alpha_a(q_a) = \alpha_a(0) e^{-f_a q_a} \quad (21)$$

is the charge-dependent dipole polarizability, where  $\alpha_a(0)$  and  $f_a$  are treated as parameters and the  $N_{\text{eff},a}(0)$  parameters are taken from ref 89. The TT damping function is

$$S_6(q_a, q_b, R_{ab}) = 1 - e^{-b_{ab} R_{ab}} \sum_{k=0}^6 \frac{(b_{ab} R_{ab})^k}{k!} \quad (22)$$

**2.3. The mDC Fock Matrix and SCF Procedure.** The spin-resolved Fock matrix of region  $A$  is

$$F_{ij}^{\sigma,A} = \left. \frac{\partial E^A}{\partial P_{ij}^{\sigma,A}} \right|_{\mathbf{q}, \mathbf{R}} + \sum_{a \in A} \sum_{lm \in a} p_{lm_a} \frac{dq_{lm_a}}{dP_{ij}^{\sigma,A}} \quad (23)$$

where  $\mathbf{P}$  is a vector of “multipolar potentials”, i.e.,

$$p_{lm_a} = \left. \frac{\partial E}{\partial q_{lm_a}} \right|_{\mathbf{P}, \mathbf{R}} \quad (24)$$

The dependence of the inter-region interactions on the density matrices occurs entirely through the atomic multipole moments, and therefore the Fock matrix need only be corrected by reverse-mapping the multipolar potentials. The mapping of the density

1. Construct initial guess orbitals.
2. Collect eigenvalues from regions, populate, and send occupation numbers.
3. Compute the density matrix of each region.
4. Collect multipole moments of each atom from each region.
5. Transform multipole moments to multipolar potentials from global electrostatics and short-range empirical inter-region interactor.
6. Compute the Fock matrix for each region and return the intra-region energy.
7. Compute the error matrix of each region. If the error is small, then exit.
8. Compute new orbitals and go to (2).

Figure 1. The mDC SCF procedure.

matrices to the multipole moments is a *choice*. In this work, we suppose

$$\rho_a(\mathbf{r} - \mathbf{R}_a) = \sum_{b \neq a} \sum_{i \in a} \frac{1}{2} P_{ij}^A \chi_i(\mathbf{r} - \mathbf{R}_a) \chi_j(\mathbf{r} - \mathbf{R}_b) + \sum_{ij \in a} P_{ij}^A \chi_i(\mathbf{r} - \mathbf{R}_a) \chi_j(\mathbf{r} - \mathbf{R}_a) \quad (25)$$

where  $\chi_i(\mathbf{r}) = \chi_i(r) Y_{lm_i}(\Omega)$  is an atomic orbital basis function and  $Y_{lm}(\Omega)$  is a real spherical harmonic. Furthermore, we limit the multipole expansion of the two-center densities to monopole so that the resulting multipole expansions consist of Mulliken charges and the high-order moments of the one-center densities. Inserting eq 25 into eq 5 with this constraint produces

$$q_{0,0_a} = Z_a - \sum_{ij \in a} P_{ij}^A S_{ij}^A - \sum_{b \neq a} \sum_{i \in a} \sum_{j \in b} P_{ij}^A S_{ij}^A / 2 \quad (26)$$

and

$$q_{l_a m_a} = - \sum_{ij \in a} P_{ij}^A W_{(l_a m_a), (l_j m_j)}^{(l_a)} M_{ij}^{(l_a)} \quad (27)$$

for  $l_a > 0$ , where

$$W_{(l_a m_a), (l_j m_j)}^{(l_a)} = \sqrt{\frac{4\pi}{2l_a + 1}} \int Y_{l_a m_a}(\Omega) Y_{l_j m_j}(\Omega) Y_{l_a m_a}(\Omega) d\Omega \quad (28)$$

are tabulated constants and

$$M_{ij}^{(l)} = \int_0^\infty \chi_i(r) \chi_j(r) r^{l+2} dr \quad (29)$$

is dependent on the radial behavior of the AOs. In the present work, there are only two cases to consider:

$$M_{O_{2s} O_{2p}}^{(1)} = \int_0^\infty O_{2s}(r) O_{2p}(r) r^3 dr \quad (30)$$

and

$$M_{O_{2p} O_{2p}}^{(2)} = \int_0^\infty O_{2p}(r) O_{2p}(r) r^4 dr \quad (31)$$

where  $O_{2s}(r)$  and  $O_{2p}(r)$  are the radial-components of the 2s and 2p AOs of oxygen. These integrals are readily computed from quadrature using the AOs resulting from atomic calculations; however, we treat the value of these integrals as parameters to tune the accuracy of the intermolecular interactions, which leads to significant improvement.

The derivatives required for mapping the multipolar potentials into the Fock matrix are then

$$\frac{dq_{0,0_a}}{dP_{ij}^{\sigma,A}} = \begin{cases} -S_{ij}^A & \text{if } ij \in a \\ -S_{ij}^A/2 & \text{if } i \in a \text{ and } j \notin a \text{ or vice versa} \end{cases} \quad (32)$$

and

$$\frac{dq_{l_a m_a}}{dP_{ij}^{\sigma,A}} = -W_{(l_a m_a), (l_j m_j)}^{(l_a)} M_{ij}^{(l_a)} \quad (33)$$

for  $l_a > 0$  and  $ij \in a$ .

The Fock matrix is used within the SCF procedure to generate new guess orbitals as described in Figure 1.

The above mapping of the density matrix to the auxiliary basis representation is simple in the present demonstration. Nonetheless, more sophisticated and rigorous procedures have been described in our previous works,<sup>90,91</sup> which also mapped the two-center density components to high-order Gaussian multipoles.<sup>92</sup> These methods may provide tools that extend the current work and enable the development of other linear-scaling quantum methods to be developed using higher-level quantum base models with less parameters.

**2.4. The mDC Gradients.** As described above, the total mDC energy is variationally minimized, and thus the analytic gradient formulas are no more complicated than those encountered in traditional single-determinant methods. Suppose atom  $c$  is in region  $A$ ; then its gradient in the  $X$  direction is

$$\begin{aligned} \frac{dE}{dX_c} &= \left. \frac{\partial E}{\partial X_c} \right|_{\mathbf{p}, \mathbf{q}} + \left( \sum_{\sigma \in (\alpha, \beta)} \sum_{ij \in A} \frac{\partial E^A}{\partial P_{ij}^{\sigma,A}} \right)_{\mathbf{q}, \mathbf{R}} \\ &+ \sum_{lm \in c} \left. \frac{\partial E}{\partial q_{lm}} \right|_{\mathbf{p}, \mathbf{R}} \left( \frac{dq_{lm}}{dP_{ij}^{\sigma,A}} \right) \frac{dP_{ij}^{\sigma,A}}{dX_c} \\ &= \left. \frac{\partial E}{\partial X_c} \right|_{\mathbf{p}, \mathbf{q}} + \sum_{\sigma \in (\alpha, \beta)} \sum_{ij \in A} F_{ij}^{\sigma,A} \frac{dP_{ij}^{\sigma,A}}{dX_c} \\ &= \left. \frac{\partial E}{\partial X_c} \right|_{\mathbf{p}, \mathbf{q}} - \sum_{ij \in A} Q_{ij}^A \frac{dS_{ij}^A}{dX_c} \end{aligned} \quad (34)$$

where

$$Q_{ij}^A = \sum_{\sigma \in (\alpha, \beta)} \sum_k n_k^{\sigma,A} E_{kk}^{\sigma,A} C_{ik}^{\sigma,A} C_{jk}^{\sigma,A} \quad (35)$$



**2.5. Computational Details.** We examine the binding energies and geometries of small water clusters and measure wall-clock timings of large “water spheres,” like those used in stochastic boundary simulations. The spheres were constructed from an equilibrated 216 TIP3P water box (18.86 Å cube) which was replicated periodically and cut to the desired size. The water cluster reference binding energies result from unconstrained geometry optimization with MP2(FULL)/6-311++G\*\* (referred to as “MP2” henceforth) using the Gaussian 09 (revision A.02) software package.<sup>93</sup> These energies are adiabatic MP2 electronic energies; i.e., vibrational corrections are not applied. The MP2 water dimer interaction energy is  $-6.13$  kcal/mol, which is in reasonable agreement with the experimental estimate<sup>94</sup> of  $-5.4 \pm 0.7$  kcal/mol; however, counterpoise-corrected CCSD(T) in the complete basis set limit predicts<sup>95</sup> an interaction energy of  $-5.0$  kcal/mol. One might therefore consider the MP2 value to be too attractive. We’ve chosen to parametrize the model against the MP2 data to compensate for DFTB2’s under-prediction of the water polarizability. For comparison, the TIP3P water model produces a dimer interaction energy of  $-6.59$  kcal/mol while still being considered useful for condensed phase simulations. The tables and figures of this manuscript will compare results to MP2, MP2+CP, and  $\langle\text{MP2}\rangle$ , where MP2+CP is the counterpoise corrected MP2(FULL)/6-311++G\*\* energies and  $\langle\text{MP2}\rangle$  is the average between the MP2 and MP2+CP energies. MP2+CP and  $\langle\text{MP2}\rangle$  produce dimer interaction energies of  $-4.47$  and  $-5.30$  kcal/mol, respectively. On the basis of the water dimer interaction energy of  $-5.0$  kcal/mol, the MP2 results are too attractive, and the MP2+CP results are not attractive enough. We demonstrate that the parametrized model

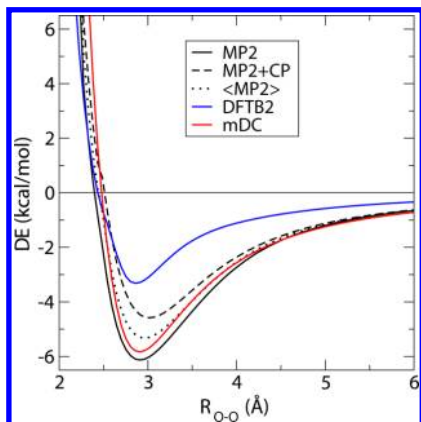


Figure 2. Water dimer PESs.

agrees more closely with the reference results than does standard DFTB2 regardless of which of these *ab initio* results is used as a reference.

To determine the parameters in the mDC model, we constructed a water dimer MP2 binding energy curve as a function of oxygen separation (see Figure 2) and adjusted the  $M_{\text{O}_2\text{O}_{2p}}^{(1)}$  parameter so that the long-range tail was reproduced. The  $M_{\text{O}_{2p}\text{O}_{2p}}^{(2)}$  parameter was adjusted so that the “wag angle” (angle  $b$  in Figure 3 and Table 1) of the water dimer was reproduced. The

Table 1. Water Dimer Geometry and Interaction Energies

method	$\Delta E$ , kcal/mol	$R_{\text{O}-\text{O}}$ , Å	$a$ , deg	$b$ , deg
MP2	-6.13	2.91	2.1	44.1
MP2+CP	-4.47			
$\langle\text{MP2}\rangle$	-5.30			
DFTB2	-3.32	2.86	3.6	65.5
mDC	-5.83	2.91	0.7	43.0
mDC(q)	-2.62	2.91	7.1	14.3
TIP3P	-6.59	2.77	4.3	21.0

remaining vdW parameters were adjusted to reproduce the water cluster relative energies (Tables 3 and 4) and water dimer binding energy curve. For comparison purposes, we also introduce a model called mDC(q), which is constructed in the same way as mDC; however, the  $M_{\text{O}_2\text{O}_{2p}}^{(1)}$  and  $M_{\text{O}_{2p}\text{O}_{2p}}^{(2)}$  parameters are zero so that the inter-region interactions involve Mulliken charges only.

Table 2 demonstrates the correctness of the gradient expression provided in section 2.4 by comparing the analytic gradients to those computed from finite differentiation of the mDC energy. The displacement value was  $3 \times 10^{-5}$  Bohr.

Table 3 shows the MP2 binding energy ( $\Delta E = E((\text{H}_2\text{O})_n) - nE(\text{H}_2\text{O})$ ) for various cluster sizes. The MP2+CP binding energies are  $\Delta E = E((\text{H}_2\text{O})_n) - \sum_{i=1}^n E_i(\text{H}_2\text{O}; (\text{H}_2\text{O})_n)$ , where  $E_i(\text{H}_2\text{O}; (\text{H}_2\text{O})_n)$  is the energy of the  $i$ th water in the basis of the full cluster. The  $\langle\text{MP2}\rangle$   $\Delta E$ 's are the average between MP2 and MP2+CP. The “rms” columns are the root-mean-square of the geometry optimized coordinates relative to the starting coordinates, where the starting coordinates are the DFTB2 isolated waters superimposed onto the optimized MP2 geometry. In this manner, the rms values shown are not biased due to the inherent difference in DFTB2 and MP2 isolated water geometries. Table 4 summarizes the  $\Delta E$  and  $\Delta\Delta E$  errors relative to MP2, MP2+CP, and  $\langle\text{MP2}\rangle$  for each cluster size. The  $\Delta\Delta E$ 's are computed as the difference in cluster binding energy relative to the minimum energy cluster configuration. “mse,” “mue,” and

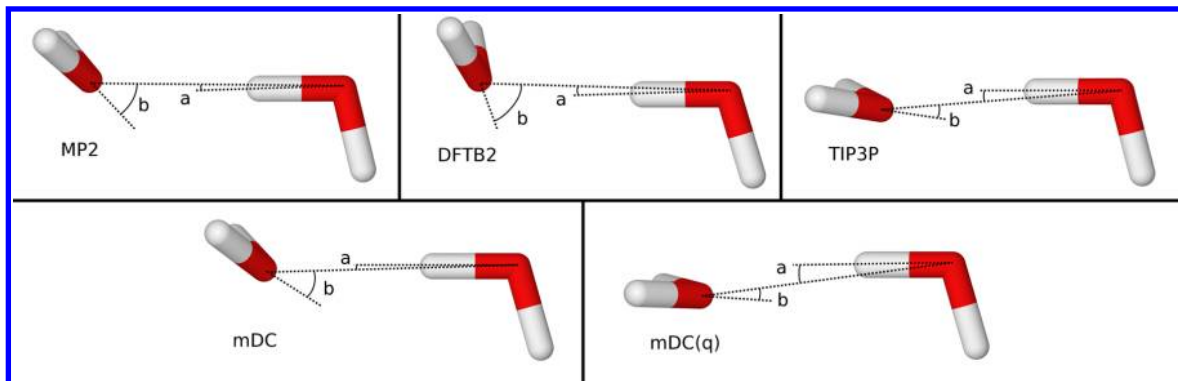


Figure 3. Comparison of water dimer geometries.

**Table 2. Comparison between Analytic Gradients and Finite Difference-Computed Gradients (kcal mol<sup>-1</sup> Å<sup>-1</sup>) for a Water Dimer with Coordinates (Å): O<sub>1</sub> (−6.30251, 2.12723, 0.82374); H<sub>1a</sub> (−6.89182, 2.55639, 1.44404); H<sub>1b</sub> (−5.73654, 1.58707, 1.375233); O<sub>2</sub> (−5.61042, 3.12260, −1.66172); H<sub>2a</sub> (−6.13634, 2.53942, −2.20902); and H<sub>2b</sub> (−5.80420, 2.83043, −0.77104)**

	analytic			analytic-numerical		
	<i>x</i>	<i>y</i>	<i>z</i>	<i>x</i>	<i>y</i>	<i>z</i>
O <sub>1</sub>	2.864 162	3.921 650	−7.135 612	0.000 000 03	0.000 000 10	0.000 000 00
H <sub>1a</sub>	0.105 627	−1.017 738	0.375 008	−0.000 000 12	−0.000 000 03	0.000 000 09
H <sub>1b</sub>	−0.983 719	−0.083 198	0.422 835	0.000 000 03	−0.000 000 03	0.000 000 02
O <sub>2</sub>	−4.968 891	−6.631 666	10.682 508	−0.000 000 06	0.000 000 01	−0.000 000 24
H <sub>2a</sub>	1.082 556	1.367 119	−1.174 105	−0.000 000 05	−0.000 000 15	−0.000 000 01
H <sub>2b</sub>	1.900 265	2.443 833	−3.170 632	−0.000 000 08	0.000 000 03	0.000 000 28

**Table 3. Cluster Binding Energies for Each Method and mDC and DFTB2 Root Mean Square Values<sup>a</sup>**

cluster	ΔE					rms	
	MP2	MP2+CP	⟨MP2⟩	mDC	DFTB2	mDC	DFTB2
(H <sub>2</sub> O) <sub>2</sub>	−6.13	−4.47	−5.30	−5.83	−3.32	0.03	0.19
(H <sub>2</sub> O) <sub>3</sub> uud	−17.92	−13.88	−15.90	−15.11	−9.62	0.20	0.31
(H <sub>2</sub> O) <sub>3</sub> uuu	−17.11	−13.28	−15.19	−13.93	−8.79	0.26	0.31
(H <sub>2</sub> O) <sub>4</sub> 1_udud	−31.57	−24.35	−27.96	−26.87	−17.69	0.17	0.16
(H <sub>2</sub> O) <sub>4</sub> 2_uudd	−30.53	−23.40	−26.97	−25.75	−16.94	0.18	0.22
(H <sub>2</sub> O) <sub>4</sub> 6	−24.15	−18.34	−21.25	−20.81	−13.07	0.40	0.75
(H <sub>2</sub> O) <sub>4</sub> 7	−24.07	−18.27	−21.17	−20.60	−13.53	0.48	2.03
(H <sub>2</sub> O) <sub>4</sub> 8	−18.99	−13.50	−16.25	−18.37	−10.25	0.17	0.11
(H <sub>2</sub> O) <sub>5</sub> Puckered_ring	−41.93	−32.10	−37.01	−35.53	−23.03	0.20	0.39
(H <sub>2</sub> O) <sub>5</sub> Envelope	−40.35	−30.78	−35.57	−34.10	−23.03	1.08	1.23
(H <sub>2</sub> O) <sub>5</sub> Tetramer_trimer	−40.15	−30.87	−35.51	−34.16	−22.37	0.23	0.32
(H <sub>2</sub> O) <sub>5</sub> Tricycle	−39.08	−29.94	−34.51	−32.61	−22.47	0.34	0.96
(H <sub>2</sub> O) <sub>5</sub> Tetramer_1	−37.55	−28.58	−33.06	−32.03	−20.75	0.20	0.53
(H <sub>2</sub> O) <sub>5</sub> Double_trimer	−35.66	−27.38	−31.52	−29.91	−19.21	0.37	0.66
(H <sub>2</sub> O) <sub>5</sub> Cage	−33.05	−24.63	−28.84	−29.34	−18.87	0.18	0.21
(H <sub>2</sub> O) <sub>6</sub> Prism	−52.18	−39.92	−46.05	−44.54	−29.60	0.20	0.21
(H <sub>2</sub> O) <sub>6</sub> Chair	−52.02	−40.01	−46.01	−44.38	−28.23	0.26	0.21
(H <sub>2</sub> O) <sub>6</sub> Book	−51.75	−39.64	−45.69	−43.79	−28.88	0.26	0.31
(H <sub>2</sub> O) <sub>6</sub> Open_bag	−51.47	−39.22	−45.34	−43.31	−28.67	0.28	0.37
(H <sub>2</sub> O) <sub>6</sub> Prism_book	−51.09	−38.66	−44.87	−42.96	−28.53	0.42	0.54
(H <sub>2</sub> O) <sub>6</sub> Boat	−50.83	−38.86	−44.85	−43.22	−27.59	0.42	0.23
(H <sub>2</sub> O) <sub>6</sub> Twisted_boat	−50.80	−38.91	−44.85	−43.14	−27.73	0.41	1.80
(H <sub>2</sub> O) <sub>6</sub> Intermediate_bag	−50.73	−38.71	−44.72	−42.93	−27.88	1.35	1.33
(H <sub>2</sub> O) <sub>6</sub> Double_tetramer	−50.70	−39.26	−44.98	−42.45	−28.10	0.47	0.42
(H <sub>2</sub> O) <sub>6</sub> Pentamer_down_1	−50.68	−38.85	−44.77	−43.07	−27.87	0.32	0.54
(H <sub>2</sub> O) <sub>6</sub> Pentamer_planar_1	−50.47	−38.66	−44.57	−43.24	−27.80	0.50	0.39
(H <sub>2</sub> O) <sub>6</sub> Cage	−50.31	−38.61	−44.46	−42.60	−28.52	0.28	0.45
(H <sub>2</sub> O) <sub>6</sub> Double_envelope	−48.83	−37.59	−43.21	−41.84	−27.35	0.37	0.41
(H <sub>2</sub> O) <sub>6</sub> Tetramer_trimer	−48.80	−37.53	−43.16	−40.92	−26.74	0.34	0.65
(H <sub>2</sub> O) <sub>6</sub> Tricycle_1a	−47.65	−36.35	−42.00	−40.40	−27.54	0.35	1.18
(H <sub>2</sub> O) <sub>6</sub> Closed_bag	−47.08	−35.93	−41.50	−40.10	−26.32	0.37	0.51
(H <sub>2</sub> O) <sub>6</sub> Tricycle_1b	−46.79	−35.45	−41.12	−39.96	−26.33	0.22	0.69
(H <sub>2</sub> O) <sub>6</sub> Triple_trimer	−43.01	−33.13	−38.07	−36.10	−27.67	0.46	2.88

<sup>a</sup>All energies are kcal mol<sup>-1</sup>. Coordinate rms values (au) are relative to the structure before geometry optimization, which are the DFTB2 isolated waters superimposed onto the optimized MP2 cluster geometry.

“max” are the mean signed, mean unsigned, and maximum error, respectively. The “rms” columns display the mean and maximum rms values relative to the optimization starting coordinates.

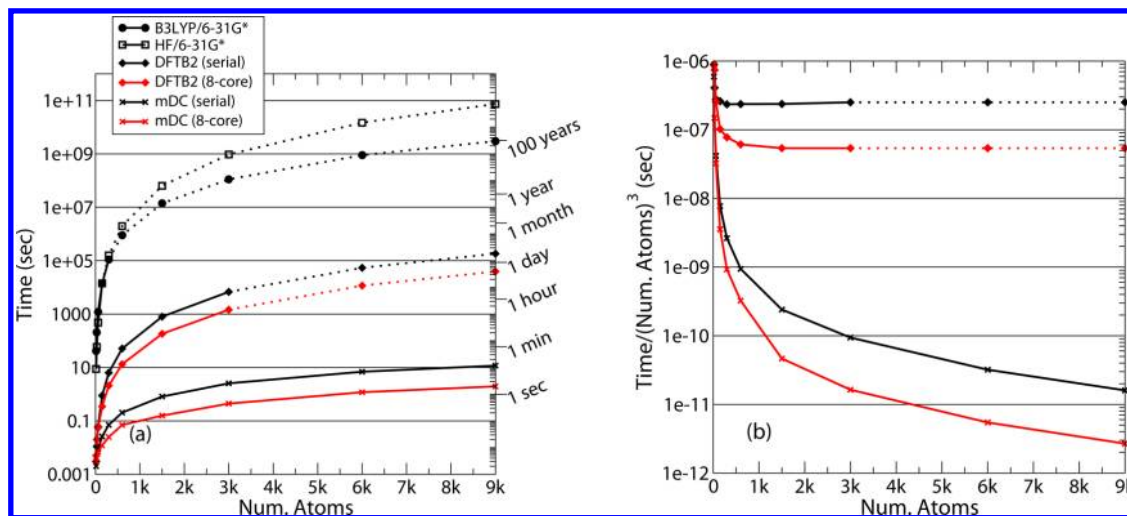
Figures 4 and 5 display the amount of wall-clock time required to perform a SCF calculation for various models. All of these timings (except for the B3LYP/6-31G\* timings) were performed on a Dell Precision T5500 desktop computer equipped with 12 GB of memory and dual quad-core Intel Xeon E5520 processors clocked at 2.27 GHz. Some of the calculations in these figures

were performed on a single core while others used all eight cores via OpenMP parallelization; see the figure captions for details. The B3LYP/6-31G\* timings shown in Figure 4a were performed on a single AMD Opteron (Model 8356) core using the GAMESS (ref 96) software package. The HF/6-31G\* timings were performed using the PSI (ref 97) program on a single Xeon E5520 core. The GAMESS and PSI calculations perform “direct SCF,” i.e., integrals are recomputed to avoid writing large integral files to disk. All other program options were left to default values.

Table 4. Summary of mDC and DFTB2 Cluster Binding Energy ( $\Delta E$ ) and Cluster Relative Energies ( $\Delta\Delta E$ ) Errors (kcal mol<sup>-1</sup>) Relative to the MP2, MP2+CP, and  $\langle$ MP2 $\rangle$  Reference Values<sup>a</sup>

ref.	cluster		$\Delta E$		$\Delta\Delta E$		rms	
			mDC	DFTB2	mDC	DFTB2	mDC	DFTB2
MP2	(H <sub>2</sub> O) <sub>4</sub>	mse	3.38	11.57	-1.65	-2.90		
		mue	3.38	11.57	1.69	2.90	0.28	0.66
		max	4.78	13.89	-4.08	-5.15	0.48	2.03
	(H <sub>2</sub> O) <sub>5</sub>	mse	5.73	16.86	-0.78	-2.38		
		mue	5.73	16.86	0.80	2.38	0.37	0.61
		max	6.47	18.90	-2.69	-4.73	1.08	1.23
	(H <sub>2</sub> O) <sub>6</sub>	mse	7.57	21.88	-0.08	-0.74		
		mue	7.57	21.88	0.36	1.15	0.40	0.73
		max	8.26	23.78	-0.81	-7.24	1.35	2.88
MP2+CP	(H <sub>2</sub> O) <sub>4</sub>	mse	-2.91	5.28	-0.49	-1.74		
		mue	2.91	5.28	0.69	1.74		
		max	-4.87	6.67	-2.35	-3.42		
	(H <sub>2</sub> O) <sub>5</sub>	mse	-3.34	7.79	0.11	-1.49		
		mue	3.34	7.79	0.54	1.49		
		max	-4.72	9.07	-1.28	-3.32		
	(H <sub>2</sub> O) <sub>6</sub>	mse	-4.09	10.22	0.55	-0.11		
		mue	4.09	10.22	0.55	0.92		
		max	-4.61	11.78	1.65	-4.86		
$\langle$ MP2 $\rangle$	(H <sub>2</sub> O) <sub>4</sub>	mse	0.23	8.42	-1.07	-2.32		
		mue	1.09	8.42	1.13	2.32		
		max	-2.13	10.28	-3.22	-4.28		
	(H <sub>2</sub> O) <sub>5</sub>	mse	1.19	12.33	-0.33	-1.93		
		mue	1.34	12.33	0.52	1.93		
		max	1.90	13.99	-1.98	-4.02		
	(H <sub>2</sub> O) <sub>6</sub>	mse	1.74	16.05	0.24	-0.43		
		mue	1.74	16.05	0.33	1.01		
		max	2.54	17.78	1.02	-6.05		

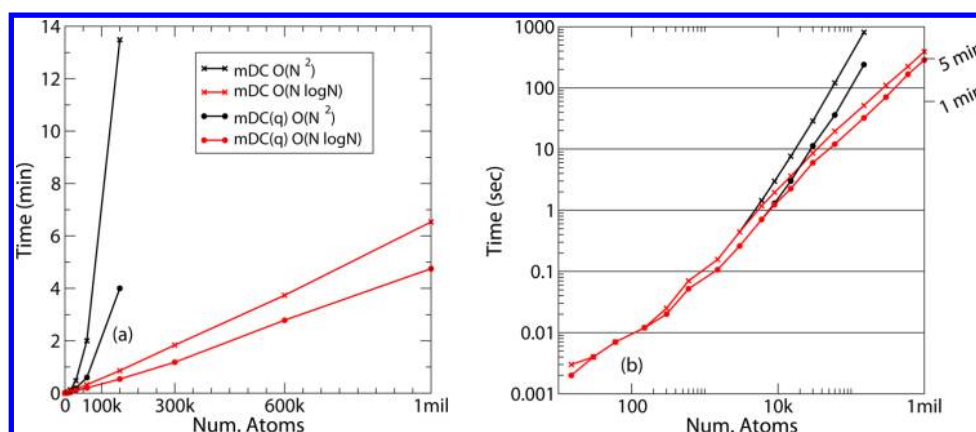
<sup>a</sup>The coordinate root mean square statistics (au) are relative to the structure before geometry optimization, which are the DFTB2 isolated waters superimposed onto the optimized MP2 cluster geometry.



**Figure 4.** SCF evaluation timings for spheres of water of various sizes. (a) Timings shown on a log scale. (b) Timings divided by the cube of the number of atoms. Part b is meant to illustrate the size at which the matrix diagonalization begins to adversely dominate the performance of the SCF procedure. Solid lines, observed timings; dotted lines, extrapolated estimates. Black lines, serial evaluation; red lines, OpenMP parallelized with eight threads on an eight-processing-core workstation. Squares, HF/6-31G\* evaluated with PSI (ref 97); circles, B3LYP/6-31G\* evaluated with GAMESS (ref 96); diamonds, standard DFTB2; X's, mDC. The B3LYP calculations were performed on an AMD Opteron (Model 8356). All other calculations were performed on an Intel Xeon E5520 workstation.

The solid lines are measured timings, and the dotted lines are extrapolations. For example, our implementation of the standard DFTB2 method requires too much memory beyond 3000 atoms

to be run on our desktop, and so we rely on extrapolation for an estimate. There are very few B3LYP and HF data points available due to the enormous cost of these methods; therefore, the



**Figure 5.** SCF evaluation timings for spheres of water of various sizes. (a) Timings shown on a linear scale. (b) Timings shown on a double log scale. Black lines,  $O(N^2)$  brute-force electrostatics; red lines,  $O(N \log N)$  adaptive FMM electrostatics. X's, mDC with point multipoles; circles, mDC with point charges. All calculations were performed on an Intel Xeon E5520 workstation and performed using eight cores.

extrapolated timings of these models cannot be regarded as being quantitatively reliable, but they are instead intended to offer a qualitative guess at the order of magnitude that one would expect. Figure 5 compares the use of direct evaluation of the electrostatics to the adaptive fast multipole method (FMM) described in ref 98.

### 3. RESULTS AND DISCUSSION

**3.1. The Effect of Intermolecular Tight-Binding Energy on the Water Dimer Geometry.** Standard DFTB2 does a good job at reproducing the geometries and relative binding energies of water clusters. It is particularly impressive that DFTB2 predicts the correct geometry of the water dimer fairly accurately even though the explicit electrostatic interactions occur through monopolar charges only. When we remove the interwater overlap and associated Fock matrix elements [mDC(q)], however, the hydrogen bond angle “flattens” in a manner similar to what one observes with the TIP3P water model (see Figure 3 and Table 1). From this we infer that DFTB2 retains the proper dimer wag angle, not through a second order electrostatic interaction but through the multipolar character of the two-center AO basis products and the resulting effect on the inter-region coupling through the tight-binding energy. Note, that despite getting a fairly accurate water dimer binding geometry, the water dimer adiabatic binding energy with DFTB2 is considerably under-bound by 2.8 kcal/mol relative to MP2 or 1.2 kcal/mol relative to MP2+CP.

We were interested in exploring the possibility of creating a mDC method that improves the accuracy of the intermolecular interaction energies relative to DFTB2 while retaining good geometries. Initial tests using models based on monopolar charge representations for the second-order term, while able in some cases to obtain improved energies, were not successful in reliably reproducing the dimer (and cluster) geometries. The solution to the geometry problem came upon considering models that included higher-order multipole electrostatic interactions. Our approach was thus to expand the density matrix to higher-order atomic multipoles and use those expansions only for the inter-region interaction. Since these higher-order multipoles do not directly alter the intraregion energy, this strategy allows us to avoid having to reparametrize the DFTB2 model for intramolecular interactions, for which the model has proved to be highly successful, and instead concentrate on intermolecular interactions. The multipole expansions do indirectly affect the

energy of a water through the variational procedure that allows coupling of electrostatic interactions (through the multipole expansions) with other water molecules. The variational inclusion of the multipole moment expansions allows us to compute the mDC analytic gradients accurately with simple formulas (see Table 2).

It was found that expanding the oxygen density up to quadrupole terms was necessary to obtain improved water dimer angles. Adjusting the  $M_{O_{2p}O_{2p}}^{(2)}$  parameter reduces the “b” angle error (Table 1) from  $29.8^\circ$  [mDC(q)] to  $1.1^\circ$  (mDC). We choose not to reduce this error to zero so that we can better reproduce the binding energies and geometries of larger clusters (section 3.2). The  $M_{O_{2p}O_{2p}}^{(1)}$  parameter was then adjusted to reproduce the long-range tail ( $R_{oo} > 5 \text{ \AA}$ ) of the water dimer (Figure 2). The tail of the water dimer is not exactly reproduced because we found that the relative energies of larger water clusters were benefited by slightly overpolarizing the mDC water. From this simple procedure, the dipole moments of the isolated waters are (D) 2.19 (MP2), 1.63 (DFTB2), and 2.34 (mDC), and the quadrupole moments are (au) 2.14 (MP2), 1.11 (DFTB2), and 2.14 (mDC).

**3.2. Reproduction of Water Cluster Relative Energies and Geometries.** Table 3 compares DFTB2 and mDC water cluster geometries and relative energies to MP2, MP2+CP, and (MP2). In general, DFTB2 does an excellent job of retaining the proper water cluster minima. Only a few DFTB2 clusters degenerate into lower energy clusters (e.g., SHOH\_Envelope) or other hydrogen bonding arrangements (e.g., 4HOH\_7, 6HOH\_Twisted\_Boat, 6HOH\_Intermediate\_bag, 6HOH\_Tri-cycle\_1a, and 6HOH\_Triple\_trimer.) mDC does not rearrange the hydrogen bonds in any case where DFTB2 does not and retains the MP2 hydrogen bonding network in most cases where DFTB2 reorganizes. The exceptions to this are the SHOH\_Envelope and 6HOH\_Intermediate\_bag structures, which both DFTB2 and mDC undergo a reorganization.

Table 4 summarizes the  $\Delta E$  and  $\Delta\Delta E$  differences between the models. The mDC model reduces the  $\Delta\Delta E$  unsigned error by factors of 1.7, 3.0, and 3.2 for tetramer, pentamer, and hexamer clusters, respectively, when compared to MP2 and by factors of 2.5, 2.8, and 1.7 when compared to MP2+CP. The rms errors are also reduced by nearly a factor of 2.

**3.3. Computational Cost.** Figure 4a displays timing results for *ab initio*, DFTB2, and mDC. For a 9000 atom system, an *ab initio* SCF procedure would take 100 years to complete, assuming



a cubic extrapolation of observed timings. Standard DFTB2 would take a day and mDC takes 2–10 s. Figure 4b redisplay the DFTB2 and mDC timings divided by the cube of the number of atoms. Therefore, the horizontal lines in this figure represent perfect  $O(N^3)$  scaling. The DFTB2 timings become dominated by the  $O(N^3)$  diagonalizations at around 600 atoms.

Overcoming the diagonalization bottleneck is only one obstacle toward achieving linear scaling. Electrostatic interactions scale  $O(N^2)$ , but methods for its  $O(N \log N)$  evaluation exist, e.g., various linear-scaling Ewald methods for periodic systems<sup>99–102</sup> and the FMM for nonperiodic systems.<sup>98</sup> Figure 5a compares the SCF timings between brute-force electrostatic with an adaptive FMM.<sup>98</sup> The mDC timings with the FMM show linearity to 1 million atoms. We also show the timings for the mDC(q) model to measure the cost of including higher-order multipoles in the method. For large systems, the use of multipoles increases the cost by less than a factor of 2. When comparing the timings between mDC and mDC(q), it is important to note that they differ only in their treatment of the electrostatic interactions, and the differences are not due to our modification of DFTB2 to expand the density to higher-order atomic multipoles. The atoms in mDC(q) continue to have point quadrupole moments, and those expansions are still mapped from the density matrix and the potentials reverse-mapped into the Fock matrix. It only so happens that the  $M_{O_{2s}O_{2p}}^{(1)}$  and  $M_{O_{2s}O_{2p}}^{(2)}$  parameters have been set to zero and the electrostatic interactions have been optimized for point-charges.

Figure 5b shows Figure 5a on a double-log scale to make the smaller systems more visible and show the points at which the FMM becomes faster than brute-force electrostatic evaluation. The “break-even” points, relative to full evaluation of electrostatic interactions, occur at 4000 atoms (point multipoles) and 8000 atoms (point charges). Prior to these system sizes, full electrostatic evaluation is quite efficient. To put the scale of Figure 5 into perspective, a desktop computer computed a 1 million atom system with mDC in 6.5 min. Assuming a cubic scaling, standard DFTB2 and B3LYP/6-31G\* would take 1700 years and 131 million years, respectively.

#### 4. CONCLUSION

We have described a variational modified divide-and-conquer algorithm with charge-dependent inter-region interactions that eliminates the need for a buffer space, is extremely efficient and accurate, and has simple gradient formulas. We have applied this method with the DFTB2 Hamiltonian and demonstrated that intermolecular interactions in the mDC method require higher-order multipole expansions to reproduce the correct geometry and binding for a broad range of water clusters. The resulting mDC model well-reproduces the reference MP2 quadrupole moment of water and greatly improves the binding energies and geometries of water clusters relative to the full DFTB2 method. This strategy allows intermolecular interactions to be tuned to obtain even higher accuracy than the DFTB2 method without sacrificing the quality of the intramolecular geometries and energies. This balance of intra- and intermolecular interactions may be important for the development of linear-scaling quantum mechanical force fields for molecular simulations. DFTB2 and mDC program execution timings were compared with *ab initio* programs using modestly simple Hamiltonians, from which the performance of mDC was shown to be superior. It was shown that a complete SCF calculation of a 3000 atom system, using

quadrupole expansions and the FMM, takes 0.44 s, whereas a 1 million atom system was possible to calculate in 6.5 min.

#### ■ ASSOCIATED CONTENT

##### Supporting Information

Parameters for the mDC model are provided. This material is available free of charge via the Internet at <http://pubs.acs.org/>.

#### ■ AUTHOR INFORMATION

##### Corresponding Author

\*E-mail: [york@biomaps.rutgers.edu](mailto:york@biomaps.rutgers.edu).

##### Notes

The authors declare no competing financial interest.

#### ■ 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) Lassila, J. K.; Zalatan, J. G.; Herschlag, D. *Annu. Rev. Biochem.* **2011**, *80*, 669–702.
- (2) Steinbrecher, T.; Koslowski, T.; Case, D. A. *J. Phys. Chem. B* **2008**, *112*, 16935–16944.
- (3) Bressler, C.; Chergui, M. *Annu. Rev. Phys. Chem.* **2010**, *61*, 263–282.
- (4) Castner, E. W., Jr.; Margulis, C. J.; Maroncelli, M.; Wishart, J. F. *Annu. Rev. Phys. Chem.* **2011**, *62*, 85–105.
- (5) Barone, V.; Improta, R.; Rega, N. *Acc. Chem. Res.* **2008**, *41*, 605–616.
- (6) Bühl, M.; van Mourik, T. *WIREs Comput. Mol. Sci.* **2011**, *1*, 634–647.
- (7) Yang, W. *Phys. Rev. A* **1991**, *44*, 7823–7826.
- (8) Yang, W.; Lee, T.-S. *J. Chem. Phys.* **1995**, *103*, 5674–5678.
- (9) Scuseria, G. E.; Ayala, P. Y. *J. Chem. Phys.* **1999**, *111*, 8330–8343.
- (10) Shao, Y.; White, C. A.; Head-Gordon, M. *J. Chem. Phys.* **2001**, *114*, 6572–6577.
- (11) Morokuma, K. *Philos. Trans. R. Soc. London, Ser. A* **2002**, *360*, 1149–1164.
- (12) Goedecker, S.; Scuseria, G. E. *IEEE Comput. Sci. Eng.* **2003**, *5*, 14–21.
- (13) He, X.; Zhang, J. Z. H. *J. Chem. Phys.* **2006**, *124*, 184703.
- (14) Gogonea, V.; Suárez, D.; van der Vaart, A.; Merz, K. M., Jr. *Curr. Opin. Struct. Biol.* **2001**, *11*, 217–223.
- (15) Lewis, J. P.; Liu, S.; Lee, T.-S.; Yang, W. *J. Comput. Phys.* **1999**, *151*, 242–263.
- (16) Khandogin, J.; York, D. M. *J. Phys. Chem. B* **2002**, *106*, 7693–7703.
- (17) Monard, G.; Bernal-Uruchurtu, M. I.; van der Vaart, A.; Merz, K. M., Jr.; Ruiz-López, M. F. *J. Phys. Chem. A* **2005**, *109*, 3425–3432.
- (18) Shao, Y.; et al. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172–3191.
- (19) Hu, H.; Lu, Z.; Elstner, M.; Hermans, J.; Yang, W. *J. Phys. Chem. A* **2007**, *111*, 5685–5691.
- (20) Zhou, T.; Huang, D.; Caflisch, A. *Curr. Top. Med. Chem.* **2010**, *10*, 33–45.
- (21) Hu, X.; Jin, Y.; Zeng, X.; Hu, H.; Yang, W. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7700–7709.
- (22) He, X.; Merz, K. M., Jr. *J. Chem. Theory Comput.* **2010**, *6*, 405–411.
- (23) Vydrov, O. A.; Van Voorhis, T. *J. Chem. Theory Comput.* **2012**, *8*, 1929–1934.

- (24) Kobayashi, M.; Kunisada, T.; Akama, T.; Sakura, D.; Nakai, H. *J. Chem. Phys.* **2011**, *134*, 034105.
- (25) Giese, T. J.; York, D. M. *J. Chem. Phys.* **2007**, *127*, 194101.
- (26) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260–7268.
- (27) Gao, J. J. *Phys. Chem.* **1997**, *101*, 657–663.
- (28) Gao, J. J. *Chem. Phys.* **1998**, *109*, 2346–2354.
- (29) Xie, W.; Gao, J. *J. Chem. Theory Comput.* **2007**, *3*, 1890–1900.
- (30) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P. *J. Chem. Theory Comput.* **2007**, *3*, 1960–1986.
- (31) Gresh, N.; Claverie, P.; Pullman, A. *Theor. Chem. Acc.* **1984**, *66*, 1–20.
- (32) Piquemal, J.; Cisneros, G.; Reinhardt, P.; Gresh, N.; Darden, T. A. *J. Chem. Phys.* **2006**, *124*, 104101.
- (33) Cisneros, G. A.; Piquemal, J.; Darden, T. A. *J. Chem. Phys.* **2006**, *125*, 184101.
- (34) Cisneros, G. A.; Elking, D.; Piquemal, J.-P.; Darden, T. A. *J. Phys. Chem. A* **2007**, *111*, 12049–12056.
- (35) Cisneros, G. A.; Darden, T. A.; Gresh, N.; Reinhardt, P.; Parisel, O.; Pilmé, J.; Piquemal, J.-P. *Multi-Scale Quantum Models for Biocatalysis*; Springer Verlag: New York, 2009; Chapter Design of next generation force fields from ab initio computations: Beyond point charges electrostatics, pp 137–172.
- (36) Schwegler, E.; Challacombe, M. *J. Chem. Phys.* **1999**, *111*, 6223–6229.
- (37) Challacombe, M. *J. Chem. Phys.* **2000**, *113*, 10037–10043.
- (38) Beck, T. L. *Rev. Mod. Phys.* **2000**, *72*, 1041–1080.
- (39) Watson, M. A.; Sałek, P.; Macak, P.; Helgaker, T. *J. Chem. Phys.* **2004**, *121*, 2915–2931.
- (40) White, C. A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M. *Chem. Phys. Lett.* **1994**, *230*, 8–16.
- (41) Kutte, R.; Aprá, E.; Nichols, J. *Chem. Phys. Lett.* **1995**, *238*, 173–179.
- (42) Burant, J. C.; Strain, M. C.; Scuseria, G. E.; Frisch, M. J. *Chem. Phys. Lett.* **1996**, *248*, 43–49.
- (43) Strain, M. C.; Scuseria, G. E.; Frisch, M. J. *Science* **1996**, *271*, 51–53.
- (44) Schwegler, E.; Challacombe, M.; Head-Gordon, M. *J. Chem. Phys.* **1997**, *106*, 9708–9717.
- (45) Challacombe, M.; Schwegler, E. *J. Chem. Phys.* **1997**, *106*, 5526–5536.
- (46) Jung, Y.; Sodt, A.; Gill, P. W.; Head-Gordon, M. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6692–6697.
- (47) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 1342–1348.
- (48) Xie, W.; Song, L.; Truhlar, D. G.; Gao, J. *J. Chem. Phys.* **2008**, *128*, 234108.
- (49) Xie, W.; Orozco, M.; Truhlar, D. G.; Gao, J. *J. Chem. Theory Comput.* **2009**, *5*, 459–467.
- (50) Isegawa, M.; Gao, J.; Truhlar, D. G. *J. Chem. Phys.* **2011**, *135*, 084107.
- (51) Zhang, P.; Truhlar, D. G.; Gao, J. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7821–7829.
- (52) Wang, Y.; Sosa, C. P.; Cembran, A.; Truhlar, D. G.; Gao, J. *J. Phys. Chem. B* **2012**, *116*, 6781–6788.
- (53) Gao, J.; Wang, Y. *J. Chem. Phys.* **2012**, *136*, 071101.
- (54) Cembran, A.; Bao, P.; Wang, Y.; Song, L.; Truhlar, D. G.; Gao, J. *J. Chem. Theory Comput.* **2010**, *6*, 2469–2476.
- (55) Jacobson, L. D.; Herbert, J. M. *J. Chem. Phys.* **2011**, *134*, 094118.
- (56) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999**, *313*, 701–706.
- (57) Nagata, T.; Brorsen, K.; Fedorov, D. G.; Kitaura, K.; Gordon, M. S. *J. Chem. Phys.* **2011**, *134*, 124115.
- (58) Pruitt, S. R.; Fedorov, D. G.; Kitaura, K.; Gordon, M. S. *J. Chem. Theory Comput.* **2010**, *6*, 1–5.
- (59) Nagata, T.; Fedorov, D. G.; Kitaura, K.; Gordon, M. S. *J. Chem. Phys.* **2009**, *131*, 024101–024112.
- (60) Fedorov, D. G.; Kitaura, K. *J. Phys. Chem. A* **2007**, *111*, 6904–6914.
- (61) Fedorov, D. G.; Olson, R. M.; Kitaura, K.; Gordon, M. S.; Koseki, S. *J. Comput. Chem.* **2004**, *25*, 872–880.
- (62) Nakano, T.; Kaminuma, T.; Sato, T.; Fukuzawa, K.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. *Chem. Phys. Lett.* **2002**, *351*, 475–480.
- (63) Le, H.-A.; Tan, H.-J.; Ouyang, J. F.; Bettens, R. P. A. *J. Chem. Theory Comput.* **2012**, *8*, 469–478.
- (64) Elliott, P.; Cohen, M. H.; Wasserman, A.; Burke, K. *J. Chem. Theory Comput.* **2009**, *5*, 827–833.
- (65) Mayhall, N. J.; Raghavachari, K. *J. Chem. Theory Comput.* **2012**, *8*, 2669–2675.
- (66) Yang, W. *J. Mol. Struct.* **1992**, *255*, 461–479.
- (67) Ammar, G. S.; Reichel, L.; Sorensen, D. C. *ACM Trans. Math Soft.* **1992**, *18*, 292–307.
- (68) Lee, C.; Yang, W. *J. Chem. Phys.* **1992**, *96*, 2408–2411.
- (69) Zhao, Q.; Yang, W. *J. Chem. Phys.* **1995**, *102*, 9598–9603.
- (70) Dixon, S. L.; Merz, K. M., Jr. *J. Chem. Phys.* **1996**, *104*, 6643–6649.
- (71) Lee, T.-S.; York, D. M.; Yang, W. *J. Chem. Phys.* **1996**, *105*, 2744–2750.
- (72) York, D. M.; Lee, T.-S.; Yang, W. *J. Am. Chem. Soc.* **1996**, *118*, 10940–10941.
- (73) Dixon, S. L.; Merz, K. M., Jr. *J. Chem. Phys.* **1997**, *107*, 879–893.
- (74) Pan, W.; Lee, T.-S.; Yang, W. *J. Comput. Chem.* **1998**, *19*, 1101–1109.
- (75) Lee, T.-S.; Yang, W. *Int. J. Quantum Chem.* **1998**, *69*, 397–404.
- (76) van der Vaart, A.; Suárez, D.; Merz, K. M., Jr. *J. Chem. Phys.* **2000**, *113*, 10512–10523.
- (77) van der Vaart, A.; Valentin, G.; Dixon, S. L.; Merz, K. M., Jr. *J. Comput. Chem.* **2000**, *21*, 1494–1504.
- (78) Frauenheim, T.; Seifert, G.; Elstner, M.; Niehaus, T.; Köhler, C.; Amkreutz, M.; Sternberg, M.; Hajnal, Z.; Di Carlo, A.; Suhai, S. *J. Phys.: Condens. Matter* **2002**, *14*, 3015–3047.
- (79) Wollacott, A. M.; Merz, K. M., Jr. *J. Chem. Theory Comput.* **2007**, *3*, 1609–1619.
- (80) Song, G.-L.; Li, Z. H.; Liu, Z.-P.; Cao, X.-M.; Wang, W.; Fan, K.-N.; Xie, Y.; Schaefer, H. F., III. *J. Chem. Theory Comput.* **2008**, *4*, 2049–2056.
- (81) Kobayashi, M.; Nakai, H. *J. Chem. Phys.* **2009**, *131*, 114108.
- (82) Alizadegan, R.; Hsia, K. J.; Martinez, T. J. *J. Chem. Phys.* **2010**, *132*, 034101.
- (83) Zhu, T.; He, X.; Zhang, J. Z. H. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7837–7845.
- (84) Giese, T. J.; York, D. M. *J. Chem. Phys.* **2008**, *129*, 016102.
- (85) Elstner, M.; Frauenheim, T.; Kaxiras, E.; Seifert, G.; Suhai, S. *Phys. Status Solidi B* **2000**, *217*, 357–376.
- (86) Yang, Y.; Yu, H.; York, D.; Elstner, M.; Qiang, C. J. *J. Chem. Theory Comput.* **2008**, *4*, 2067–2084.
- (87) Kaminski, S.; Gaus, M.; Phatak, P.; von Stetten, D.; Elstner, M.; Mroginiski, M. A. *J. Chem. Theory Comput.* **2010**, *6*, 1240–1255.
- (88) Gaus, M.; Cui, Q.; Elstner, M. *J. Chem. Theory Comput.* **2011**, *7*, 931–948.
- (89) Pellenq, R.; Nicholson, D. *Mol. Phys.* **1998**, *95*, 549–570.
- (90) Giese, T. J.; York, D. M. *J. Chem. Phys.* **2011**, *134*, 194103.
- (91) Giese, T. J.; York, D. M. *Theor. Chem. Acc.* **2012**, *131*, 1145.
- (92) Giese, T. J.; York, D. M. *J. Chem. Phys.* **2008**, *128*, 064104.
- (93) Frisch, M. J. et al. *Gaussian 09*, revision A.02; Gaussian, Inc.: Wallingford, CT, 2009.
- (94) Dyke, T. R.; Mack, K. M.; Muentner, J. S. *J. Chem. Phys.* **1977**, *66*, 498–510.
- (95) Kloppe, W.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Duijneveldt, F. B. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2227–2234.
- (96) Schmidt, M. W.; Baldrige, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A., Jr. *J. Comput. Chem.* **1993**, *14*, 1347–1363.
- (97) Crawford, T. D.; Sherrill, C. D.; Valeev, E. F.; Fermann, J. T.; King, R. A.; Leininger, M. L.; Brown, S. T.; Janssen, C. L.; Seidl, E. T.; Kenny, J. P.; Allen, W. D. *J. Comput. Chem.* **2007**, *28*, 1610–1616.
- (98) Giese, T. J.; York, D. M. *J. Comput. Chem.* **2008**, *29*, 1895–1904.

- (99) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (100) York, D. M.; Yang, W. *J. Chem. Phys.* **1994**, *101*, 3298–3300.
- (101) Sagui, C.; Darden, T. A. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 155–179.
- (102) Nam, K.; Gao, J.; York, D. M. *J. Chem. Theory Comput.* **2005**, *1*, 2–13.