

A dual length scale method for plane-wave-based, simulation studies of chemical systems modeled using mixed *ab initio*/empirical force field descriptions

Dawn A. Yarne

Department of Chemistry, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6323

Mark E. Tuckerman

Department of Chemistry and The Courant Institute of Mathematical Sciences, New York University, New York, New York 10003

Glenn J. Martyna

Department of Chemistry, Indiana University, Bloomington, Indiana 47405-4001

(Received 12 March 2001; accepted 14 May 2001)

Mixed *ab initio*/empirical force-field simulation studies, calculations in which one part of the system is treated using a fully *ab initio* description and another part is treated using an empirical description, are becoming increasingly popular. Here, the ability of the commonly used, plane wave-based generalized gradient approximation to density functional theory is extended to model systems in which the electrons are assumed to be localized in a single small region of space, that is, itself, embedded within a large chemically inert bath. This is accomplished by introducing two length scales, so that the rapidly varying, short range, electron–electron and electron–atom interactions, arising from the region where the electrons are localized, can be treated using an appropriately large plane wave basis, while the corresponding, slowly varying, long range interactions of the electrons with the full system or bath, can be treated using a small basis. Briefly, a novel Cardinal B-spline based formalism is employed to derive a smooth, differentiable, and rapidly convergent (with respect to the small basis) expression for the total electronic energy, which explicitly contains the two length scales. The method allows reciprocal space based techniques designed to treat clusters, wires, surfaces and solids/liquids (open, and 1-D and 2-D periodic boundary conditions, respectively) to be utilized. Other plane wave-based “mixed” methods are restricted to clusters. The new methodology, which scales as $N \log N$ at fixed size of the chemically active region, has been implemented for parallel computing platforms and tested through applications to both model and realistic problems including an enzyme, human carbonic anhydrase II solvated in an explicit bath of water molecules. © 2001 American Institute of Physics.

[DOI: 10.1063/1.1383795]

I. INTRODUCTION

Chemical systems are often characterized by a set of electronically active constituents localized in a small region of space, surrounded by and interacting with, a large bath of electronically inert components. The division of a large system into chemically interesting and chemically uninteresting regions is both an intuitively appealing and a practically useful description that can be fruitfully applied to many different problems of chemical and biological interest. For example, in the study of solution phase chemical reactions, it is advantageous to consider, explicitly, the electronic degrees of freedom of the reactants/products and perhaps, a first solvation shell, while the remainder of the solvent is modeled more approximately.¹ Similarly, in studies of enzyme catalysis, the valence electrons of the amino acids and the water molecules near the active site, as well as those of the substrate, must be modeled using a high level of theory, while the remainder of these large and complex systems can be modeled more approximately.^{1–6} Thus, simulation studies based on hybrid model descriptions promise to yield chemical insight into significant problems for low computational cost. It is, therefore, important to develop both the models

and methods required to treat mixed *ab initio*/empirical force field descriptions of chemical and biological systems accurately and efficiently.

It has been demonstrated that a wide variety of complex chemical systems can be treated effectively using an *ab initio* methodology that employs a plane-wave basis set in conjunction with the generalized gradient approximation to density functional theory (GGA–DFT).^{7–10} Of course, in realistic calculations, many basis functions or equivalently, a large plane wave energy cutoff ($\hbar^2|\mathbf{g}|^2/2m_e \leq E_{\text{cut}} \approx 70$ Ry) must be employed to ensure accuracy. The large basis set size, coupled with the fact that plane waves, naturally, admit only a single length scale, has made it difficult to employ plane-wave-based GGA–DFT to study hybrid model systems.

Consider, for example, a small simulation cell containing the electron density embedded within a large simulation cell containing the rest of the system (i.e., the bath). In order to determine the long range interaction of the electron density, with the atoms outside the small simulation cell within the plane wave formalism, the electron density in the large simulation cell must be expanded in a plane-wave basis truncated using the **same** cutoff required to describe the rapidly vary-

ing electron density in the small cell (e.g. $E_{\text{cut}} \approx 70$ Ry). Thus, the memory requirements are prohibitively large and the calculations scale poorly with the size of the large cell (at fixed small cell size). However, such a scheme does allow systems treated using 3-D periodic boundary conditions (liquids and solids) to be accurately studied. It also permits novel reciprocal space-based techniques that treat clusters, wires and surfaces properly^{11,12} (open, and 1-D and 2-D periodic boundary conditions, respectively) to be applied to “mixed” or “hybrid” model calculations. In contrast, previous plane-wave-based “hybrid” formulations are restricted to clusters.¹³

In this paper, a new method designed to treat large systems that can be decomposed into electronically active and electronically inert portions with high efficiency is presented. Two length scales are explicitly introduced into the plane wave-based GGA–DFT electronic structure formalism so that the small length scale, electronically active region can be treated differently than the long length scale, electronically inert region without loss of generality and with large gains in efficiency. This is accomplished by employing a Cardinal B-spline based formalism to derive a novel expression for the electronic energy that explicitly contains both the long and short length scales. The new expression can be evaluated efficiently using two independent plane wave energy cutoffs and is smooth, differentiable, and rapidly convergent with respect to the plane-wave cutoff associated with the long length scale, even when the plane-wave cutoff associated with the short length scale is quite large. Thus, the method scales as $N \log N$, where N is the number of atoms in the full system (at fixed size of the chemically active region), provided particle mesh Ewald techniques¹⁴ are employed to evaluate the atomic charge density in the large cell. In addition, the new methodology does not require an *ad hoc* electrostatic potential fitting scheme based on point charges derived from a particular choice of population analysis and can be utilized to treat clusters, wires, surfaces and solids/liquids without loss of generality. The validity of the new technique is demonstrated on both an analytically solvable problem, a Gaussian charge density interacting with a point charge, and realistic systems including an enzyme, human carbonic anhydrase II (HCA II) solvated by an explicit bath of water molecules. The high parallel efficiency of the method is also demonstrated on the large, complex, HCA II/water system.

II. METHODS

In the Kohn–Sham formulation^{15,16} of density functional theory, the electron density is expanded in a set of orbitals $\{\psi_i(\mathbf{r})\}$

$$n(\mathbf{r}) = \sum_{i=1}^n o_i |\psi_i(\mathbf{r})|^2 \quad (1)$$

and the energy functional is given by

$$E[n] = T_s[\{\psi_i\}] + E_H[n] + E_{xc}[n] + E_{\text{ext}}[n] \quad (2)$$

where T_s is the kinetic energy of a system of noninteracting electrons, E_H is the Hartree energy, E_{xc} is the exchange and correlation energy and the o_i are the occupation numbers of

the orbitals.¹⁷ Here, a generalized gradient approximation is employed and Eq. (2) is, hence, referred to as a GGA-density functional.

In this work, the GGA-density functional, Eq. (2), is minimized by expanding the orbitals in a finite plane-wave basis set and varying the expansion coefficients subject to the orthogonality constraints ($\langle \psi_j | \psi_i \rangle = \delta_{ij}$). The plane-wave basis set is truncated by including all plane waves with kinetic energy less than or equal to a cutoff energy, $\hbar^2 |\mathbf{g}|^2 / 2m_e \leq E_{\text{cut}}$. Finally, core electrons, which are difficult to treat in a plane-wave basis set, are replaced by atomic pseudopotentials. Typical pseudopotentials contain a long range local contribution, E_{loc} and a short range angular momentum dependent nonlocal contribution, E_{nonloc} that serves to replace the core (i.e., $E_{\text{ext}} = E_{\text{loc}} + E_{\text{nonloc}}$).²⁰

The GGA-density functional, thus, contains, only, two terms that act at long range, the Hartree, $E_H[n]$, and local pseudopotential energies,

$$E_H[n] = \frac{e^2}{2} \sum_{\vec{S}} \int_{D(\vec{h})} d\mathbf{r} \int_{D(\vec{h})} d\mathbf{r}' \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}' + \vec{h}\vec{S}|}, \quad (3)$$

$$E_{\text{loc}}[n] = \sum_{\vec{S}} \sum_{I=1}^N \int_{D(\vec{h})} d\mathbf{r} \phi_{\text{loc}I}(\mathbf{r} - \mathbf{R}_I + \vec{h}\vec{S}) n(\mathbf{r}), \quad (4)$$

where \mathbf{R}_I is the Cartesian position of the I th ion, \vec{h} is the cell matrix whose columns contain the d cell vectors, $\det \vec{h} = V$ is the volume, and $\vec{S} = \{\hat{s}_a, \hat{s}_b, \hat{s}_c\}$ is a vector of integers indexing the periodic replicas (in clusters, only $\vec{S} = \{0,0,0\}$ is allowed, while in systems periodically replicated in three spatial dimensions, the three integers span the full range).

In plane-wave-based calculations, the orbitals and, hence, the density are expanded as follows,^{8,21}

$$\psi_j(\mathbf{r}) = \frac{1}{\sqrt{V}} \sum_{\mathbf{g}} \bar{\psi}_j(\mathbf{g}) \exp(i\mathbf{g} \cdot \mathbf{r}), \quad (5)$$

$$n(\mathbf{r}) = \frac{1}{V} \sum_{\mathbf{g}} \bar{n}(\mathbf{g}) \exp(i\mathbf{g} \cdot \mathbf{r}),$$

where $\mathbf{g} = \vec{h}^{-1} \hat{\mathbf{g}}$ and the vector of integers, $\hat{\mathbf{g}} = \{g_a, g_b, g_c\}$, indexes reciprocal space. Typically, a cutoff is introduced on the sums describing the orbitals such that $\hbar^2 g^2 / 2m_e \leq E_{\text{cut}}$. (Note, the reciprocal space summation for the density is over a reciprocal space defined by the appropriately larger cutoff, $E_{\text{cut}}^{(\text{density})} = 4E_{\text{cut}}$.) It is convenient to express the Hartree and local external energies in reciprocal space,

$$E_H = \frac{e^2}{2V} \sum_{\mathbf{g}} ' |\bar{n}(\mathbf{g})|^2 \left[\frac{4\pi}{g^2} + \hat{\phi}^{(\text{screen,Coul})}(\mathbf{g}) \right] + \left(\frac{e^2}{2V} \right) \hat{\phi}^{(\text{screen,Coul})}(0) |\bar{n}(0)|^2, \quad (6)$$

$$\begin{aligned}
 E_{\text{loc}} = & \frac{1}{V} \sum_{\mathbf{g}}' \sum_{I=1}^N \bar{n}^*(\mathbf{g}) \exp(-i\mathbf{g} \cdot \mathbf{R}_I) \\
 & \times [\tilde{\phi}_{\text{loc},I}(\mathbf{g}) - eq_I \hat{\phi}^{(\text{screen,Coul})}(\mathbf{g})] + \frac{1}{V} \sum_{I=1}^N \bar{n}(0) \\
 & \times [\tilde{\phi}_{\text{loc},I}^{(0)} - eq_I \hat{\phi}^{(\text{screen,Coul})}(0)]. \quad (7)
 \end{aligned}$$

Here, $\tilde{\phi}_{\text{loc},I}$ denotes the Fourier transform of the local pseudopotential, the prime indicates that the $\mathbf{g}=0$ term is eliminated and the function, $\tilde{\phi}_{\text{loc},I}^{(0)}$ is the nonsingular part of the local pseudopotential at $\mathbf{g}=0$. The screening function, $\hat{\phi}^{(\text{screen,Coul})}(\mathbf{g})$, depends on the periodicity of the system. For a system periodically replicated in three spatial dimensions, it is identically zero. Other cases, clusters, wires and surfaces, are discussed in detail elsewhere.^{11,12}

It is clear that the standard expressions for the Hartree and local external energies given in Eqs. (6) and (7), respectively, only possess a single length scale. A second length scale can be introduced by first rewriting the real space expressions for these two energies using the identity $\text{erf}(ar) + \text{erfc}(ar) = 1$,

$$\begin{aligned}
 E_{\text{H}}[n] = & \left\{ \frac{e^2}{2} \sum_{\hat{\mathbf{s}}} \int_{D(\hat{\mathbf{h}})} d\mathbf{r} \int_{D(\hat{\mathbf{h}})} d\mathbf{r}' \right. \\
 & \times \left. \frac{n(\mathbf{r})n(\mathbf{r}') \text{erfc}(\alpha|\mathbf{r}-\mathbf{r}'+\vec{\mathbf{h}}\hat{\mathbf{S}}|)}{|\mathbf{r}-\mathbf{r}'+\vec{\mathbf{h}}\hat{\mathbf{S}}|} \right\} \\
 & + \left\{ \frac{e^2}{2} \sum_{\hat{\mathbf{s}}} \int_{D(\hat{\mathbf{h}})} d\mathbf{r} \int_{D(\hat{\mathbf{h}})} d\mathbf{r}' \right. \\
 & \times \left. \frac{n(\mathbf{r})n(\mathbf{r}') \text{erf}(\alpha|\mathbf{r}-\mathbf{r}'+\vec{\mathbf{h}}\hat{\mathbf{S}}|)}{|\mathbf{r}-\mathbf{r}'+\vec{\mathbf{h}}\hat{\mathbf{S}}|} \right\} \\
 = & E_{\text{H}}^{\text{short}}[n] + E_{\text{H}}^{\text{long}}[n], \quad (8)
 \end{aligned}$$

$$\begin{aligned}
 E_{\text{loc}}[n] = & \left\{ \sum_{\hat{\mathbf{s}}} \sum_{I=1}^N \int_{D(\hat{\mathbf{h}})} d\mathbf{r} n(\mathbf{r}) \left[\phi_{\text{loc},I}(\mathbf{r}-\mathbf{R}_I + \vec{\mathbf{h}}\hat{\mathbf{S}}) \right. \right. \\
 & \left. \left. + \frac{eq_I \text{erf}(\alpha|\mathbf{r}-\mathbf{R}_I + \vec{\mathbf{h}}\hat{\mathbf{S}}|)}{|\mathbf{r}-\mathbf{R}_I + \vec{\mathbf{h}}\hat{\mathbf{S}}|} \right] \right\} \\
 & - \left\{ \sum_{\hat{\mathbf{s}}} \sum_{I=1}^N \int_{D(\hat{\mathbf{h}})} d\mathbf{r} n(\mathbf{r}) \right. \\
 & \left. \times \left[\frac{eq_I \text{erf}(\alpha|\mathbf{r}-\mathbf{R}_I + \vec{\mathbf{h}}\hat{\mathbf{S}}|)}{|\mathbf{r}-\mathbf{R}_I + \vec{\mathbf{h}}\hat{\mathbf{S}}|} \right] \right\} \\
 = & E_{\text{loc}}^{\text{short}}[n] + E_{\text{loc}}^{(\text{long})}[n]. \quad (9)
 \end{aligned}$$

Here, the first term in the curly brackets in each equation is short range while the second term is long range. Note, both $\phi_{\text{loc},I}(\mathbf{r})$ and $-eq_I \text{erf}(ar)/r$ approach $-eq_I/r$, asymptotically where q_I is the charge on I th ion core. In the limit

$\alpha V^{1/3} \gg 1$, the sum over images in the first term of each expression (i.e., the short range parts) can be truncated at the first or nearest image with exponentially small error.

In order to proceed, it will be assumed that the electrons are localized in a particular region of the large cell described by $\vec{\mathbf{h}}$, which can be enclosed in a small cell, described by $\vec{\mathbf{h}}_s$, centered at the point, \mathbf{R}_c . That is, the orbitals and, hence, electron density are taken to vanish on the surface of $\vec{\mathbf{h}}_s$. Furthermore, it is assumed, for simplicity, that the a_s, b_s and c_s axes of $\vec{\mathbf{h}}_s$ are parallel to the a, b and c axes of $\vec{\mathbf{h}}$, such that $\vec{\mathbf{h}}^{-1} \vec{\mathbf{h}}_s = \vec{\mathbf{D}}$, a diagonal matrix. Thus, we can define

$$\begin{aligned}
 \psi_j(\mathbf{r}_s + \mathbf{R}_c) = & \psi_{j,s}(\mathbf{r}_s) \\
 n(\mathbf{r}_s + \mathbf{R}_c) = & n_s(\mathbf{r}_s), \quad (10)
 \end{aligned}$$

where the \mathbf{r}_s spans the small cell and can be expressed as $\mathbf{r}_s = \vec{\mathbf{h}}_s \mathbf{s}$ with $0 \leq s_\alpha \leq 1$ and, both, $\psi_j(\mathbf{r}) \equiv 0$ and $n(\mathbf{r}) \equiv 0$ for $\mathbf{r}_s = \mathbf{r} - \mathbf{R}_c$ outside the domain of $\vec{\mathbf{h}}_s$. The orbitals and the electron density can be expanded in a plane-wave-basis set that spans the small cell, only,

$$\begin{aligned}
 \psi_{j,s}(\mathbf{r}_s) = & \frac{1}{\sqrt{V_s}} \sum_{\hat{\mathbf{g}}_s} \tilde{\psi}_{j,s}(\hat{\mathbf{g}}_s) \exp(i\hat{\mathbf{g}}_s \cdot \mathbf{r}_s) \\
 n_s(\mathbf{r}_s) = & \frac{1}{V_s} \sum_{\hat{\mathbf{g}}_s} \bar{n}_s(\hat{\mathbf{g}}_s) \exp(i\hat{\mathbf{g}}_s \cdot \mathbf{r}_s), \quad (11)
 \end{aligned}$$

where $\hat{\mathbf{g}}_s = \vec{\mathbf{h}}_s^{-1} \hat{\mathbf{g}}$, the vector of integers, $\hat{\mathbf{g}}_s = \{g_{a,s}, g_{b,s}, g_{c,s}\}$, indexes the small reciprocal space and $V_s = \det \vec{\mathbf{h}}_s$ is the volume of the small cell. The plane-wave-energy cutoff is taken to be $E_{\text{cut}}^{(\text{short})}$ [with the cutoff on the density $E_{\text{cut}}^{(\text{short,density})} = 4E_{\text{cut}}^{(\text{short})}$].

Given that the electron density is localized in the small cell, the short range components of the Hartree and local pseudopotential energies can be evaluated straightforwardly,

$$\begin{aligned}
 E_{\text{H}}^{(\text{short})}[n] = & \frac{e^2}{2} \int_{D(\vec{\mathbf{h}}_s)} d\mathbf{r} \int_{D(\vec{\mathbf{h}}_s)} d\mathbf{r}' \frac{n_s(\mathbf{r})n_s(\mathbf{r}') \text{erfc}(\alpha|\mathbf{r}-\mathbf{r}'|)}{|\mathbf{r}-\mathbf{r}'|} \\
 = & \frac{e^2}{2V_s} \sum_{\hat{\mathbf{g}}_s}' \bar{n}_s(-\hat{\mathbf{g}}_s) \bar{n}_s(\hat{\mathbf{g}}_s) \left[\frac{4\pi}{g_s^2} \right] \\
 & \times \left[1 - \exp\left(-\frac{g_s^2}{4\alpha^2}\right) \right] + \frac{e^2\pi}{2V_s\alpha^2} |n_s(0)|^2 \quad (12)
 \end{aligned}$$

$$\begin{aligned}
 E_{\text{loc}}^{(\text{short})}[n] = & \sum_{J=1}^{N_s} \int_{D(\vec{\mathbf{h}}_s)} d\mathbf{r} n_s(\mathbf{r}) \left[\phi_{\text{loc},J}(\mathbf{r}-\mathbf{R}_J + \mathbf{R}_c) \right. \\
 & \left. + \frac{eq_J \text{erf}(\alpha|\mathbf{r}-\mathbf{R}_J + \mathbf{R}_c|)}{|\mathbf{r}-\mathbf{R}_J + \mathbf{R}_c|} \right] \\
 = & \frac{1}{V_s} \sum_{\hat{\mathbf{g}}_s}' \sum_{J=1}^{N_s} \bar{n}_s^*(\hat{\mathbf{g}}_s) \exp(i\hat{\mathbf{g}}_s \cdot [\mathbf{R}_J - \mathbf{R}_c]) \\
 & \times \left[\tilde{\phi}_{\text{loc},J}(\hat{\mathbf{g}}_s) + \frac{4\pi eq_J}{g_s^2} \exp\left(-\frac{q_s^2}{4\alpha^2}\right) \right] \\
 & + \frac{1}{V_s} \sum_{J=1}^{N_s} \bar{n}_s(0) \left[\tilde{\phi}_{\text{loc},J}^{(0)} - \frac{eq_J\pi}{\alpha^2} \right], \quad (13)
 \end{aligned}$$

where the J sum runs over the N_s ions within the small cell, the $\hat{\mathbf{g}}_s$ sum runs over the large reciprocal-space grid of the small cell, and \mathbf{R}_c is in position of the small cell inside the large. Since the full system is not periodic on $\vec{\mathbf{h}}_s$, but on $\vec{\mathbf{h}}$, Eqs. (12–13) will only yield the correct short range energy if $\alpha V_s^{1/3} \gg 1$ and $n(r_s)$ vanishes on the small cell boundary. The nonlocal pseudopotential energy is short range and is assumed to be evaluated within the small cell (only considering the N_s ions in the small cell and using the small cell reciprocal space). Similarly, the exchange correlation and the electronic kinetic energies can also be evaluated in the small cell using standard techniques.^{8,21}

Next, the expressions for the long range portions of the Hartree and local pseudopotential energies are formulated. This is accomplished by expanding the electron density localized in the small cell in terms of the plane waves of the large cell. This expansion is permitted because the electron density, localized in the small cell, obeys periodic boundary conditions in the large cell (i.e., it is zero on the surface of $\vec{\mathbf{h}}$). Thus

$$\begin{aligned} E_H^{(\text{long})}[n] &= \frac{e^2}{2} \sum_{\vec{\mathbf{S}}} \int_{D(\vec{\mathbf{h}})} d\mathbf{r} \int_{D(\vec{\mathbf{h}})} d\mathbf{r}' \\ &\quad \times \frac{n(\mathbf{r})n(\mathbf{r}') \text{erf}(\alpha|\mathbf{r}-\mathbf{r}'+\vec{\mathbf{h}}\hat{\mathbf{S}}|)}{|\mathbf{r}-\mathbf{r}'+\vec{\mathbf{h}}\hat{\mathbf{S}}|} \\ &= \frac{e^2}{2V} \sum_{\vec{\mathbf{g}}} \bar{n}(-\mathbf{g})\bar{n}(\mathbf{g}) \left[\frac{4\pi}{g^2} \exp\left(-\frac{g^2}{4\alpha^2}\right) \right. \\ &\quad \left. + \hat{\phi}^{(\text{screen,Coul})}(\mathbf{g}) \right] + \left(\frac{e^2}{2V} \right) \\ &\quad \times \left[\hat{\phi}^{(\text{screen,Coul})}(0) - \frac{\pi}{\alpha^2} \right] |n(0)|^2, \end{aligned} \quad (14)$$

$$\begin{aligned} E_{\text{loc}}^{(\text{long})}[n] &= - \sum_{\vec{\mathbf{S}}} \sum_{I=1}^N \int_{D(\vec{\mathbf{h}})} d\mathbf{r} n(\mathbf{r}) \left[\frac{e q_I \text{erf}(\alpha|\mathbf{r}-\mathbf{R}_I+\vec{\mathbf{h}}\hat{\mathbf{S}}|)}{|\mathbf{r}-\mathbf{R}_I+\vec{\mathbf{h}}\hat{\mathbf{S}}|} \right] \\ &= - \frac{e}{V} \sum_{\vec{\mathbf{g}}} \bar{n}^*(\mathbf{g}) S(\mathbf{g}) \left[\frac{4\pi}{g^2} \exp\left(-\frac{g^2}{4\alpha^2}\right) \right. \\ &\quad \left. + \hat{\phi}^{(\text{screen,Coul})}(\mathbf{g}) \right] \\ &\quad - \frac{e}{V} \bar{n}_s(0) S(0) \left[\hat{\phi}^{(\text{screen,Coul})}(0) - \frac{\pi}{\alpha^2} \right], \end{aligned} \quad (15)$$

where

$$S(\mathbf{g}) = \sum_I q_I \exp(i\mathbf{g} \cdot \mathbf{R}_I) \quad (16)$$

is the atomic charge density and

$$\begin{aligned} \bar{n}(\mathbf{g}) &= \int_{D(\vec{\mathbf{h}})} d\mathbf{r} \exp[-i\mathbf{g} \cdot \mathbf{r}] n(\mathbf{r}) \\ &= \int_{D(\vec{\mathbf{h}}_s)} d\mathbf{r}_s \exp[-i\mathbf{g} \cdot \mathbf{r}_s] n(\mathbf{r}_s + \mathbf{R}_c) \\ &= \int_{D(\vec{\mathbf{h}}_s)} d\mathbf{r}_s \exp[-i\mathbf{g} \cdot (\mathbf{r}_s - \mathbf{R}_c)] n_s(\mathbf{r}_s) \end{aligned} \quad (17)$$

are the plane wave expansion coefficients of the electron density in the reciprocal space of the large cell, $\mathbf{g} = \vec{\mathbf{h}}^{-1} \hat{\mathbf{g}}$. The integral in Eq. (17) can be extended to cover the domain described by the large cell without loss of generality because $n(\mathbf{r}_s + \mathbf{R}_c) \equiv 0$ outside of the small cell. Note, $\bar{n}(\mathbf{g}) = \bar{n}_s(\mathbf{g}_s)$, if $\vec{\mathbf{h}}_s \equiv \vec{\mathbf{h}}$ and $\mathbf{R}_c = 0$. Methods for the efficient evaluation of Eq. (17) and, hence, Eqs. (14) and (15) are developed as follows.

First, it is clear from the long range/short range decomposition of the Hartree and local pseudopotential energies that a different plane-wave cutoff can be introduced to treat each part. That is, one cutoff, $E_{\text{cut}}^{(\text{short})}$, can be used to evaluate the short range components of the energy [Eqs. (12) and (13)], and another, $E_{\text{cut}}^{(\text{long})}$ can be used to evaluate the long range components [Eqs. (14) and (15)]. While the long range/short range decomposition is general, it is expected that the short range contributions will be obtained by integration over functions that rapidly vary spatially, while the long range contributions will be obtained by integration over a slowly varying function. Therefore, the short range energy contributions must be evaluated using a large reciprocal space cutoff [i.e., the standard $E_{\text{cut}}^{(\text{density,short})} = 4E_{\text{cut}}^{(\text{short})}$] while the long range part can be evaluated using a small cutoff, $E_{\text{cut}}^{(\text{long})} \ll E_{\text{cut}}^{(\text{short})}$ [with $E_{\text{cut}}^{(\text{density,long})} = 4E_{\text{cut}}^{(\text{long})}$]. Thus, by splitting the electronic energy into two parts, large gains in efficiency are possible.

Next, consider the case that the number of particles in the small cell, N_s , and the small cell volume, V_s , are much less than their large cell counterparts ($N_s \ll N$ and $V_s \ll V$) as would be the case for a large, chemically inert bath surrounding a chemically active subsystem. The computational cost of evaluating the short range local pseudopotential and short range Hartree, exchange correlation, nonlocal pseudopotential and the electronic kinetic energy as well as the overlap matrix, $\langle \psi_{j,s} | \psi_{i,s} \rangle$, scales like $\sim N_s^3$. The computational cost of evaluating the long range part of the Hartree and local pseudopotential energies depends on the computational cost of evaluating the atomic charge density, $S(\mathbf{g})$, and the plane wave expansion of the density in the large cell [see Eq. (17)]. Since the atomic charge density can be evaluated in $N \log N$ using Particle Mesh Ewald techniques,¹⁴ if Eq. (17) could also be evaluated in $N \log N$, the computational cost of the method would then be $N \log N$ at fixed $\vec{\mathbf{h}}_s$ and N_s . (The present approach yields a linear scaling method because, at fixed particle density and plane wave cutoff, the number of plane waves in large cell basis set increases linearly with particle number, N .)

In order to achieve linear scaling, the electron density must be interpolated from the small cell where it is described

by a plane wave expansion with a large cutoff, $E_{\text{cut}}^{(\text{density,short})}$, to the large cell where it is described by a plane wave expansion with a small cutoff, $E_{\text{cut}}^{(\text{density,long})}$, efficiently. First, consider the Fourier components of the density,

$$\bar{n}(\mathbf{g}) = \int_{D(\mathbf{h})} d\mathbf{r} \exp[-i\mathbf{g}\cdot\mathbf{r}]n(\mathbf{r}). \tag{18}$$

If $n(\mathbf{r})$ can be expressed in a finite plane wave basis,

$$n(\mathbf{r}) \equiv \frac{1}{V} \sum_{\hat{g}_a=-P_a/2+1}^{P_a/2} \sum_{\hat{g}_b=-P_b/2+1}^{P_b/2} \sum_{\hat{g}_c=-P_c/2+1}^{P_c/2} \exp(i\mathbf{g}\cdot\mathbf{r})\bar{n}(\mathbf{g}), \tag{19}$$

then the Fourier coefficients can also be determined (exactly) from a discrete sum over a real space grid,

$$\begin{aligned} \bar{n}(\mathbf{g}) \equiv & \frac{V}{P_a P_b P_c} \sum_{\hat{s}_a=0}^{P_a-1} \sum_{\hat{s}_b=0}^{P_b-1} \sum_{\hat{s}_c=0}^{P_c-1} e^{-2\pi i \hat{g}_a \hat{s}_a / P_a} \\ & \times e^{-2\pi i \hat{g}_b \hat{s}_b / P_b} e^{-2\pi i \hat{g}_c \hat{s}_c / P_c} n(\vec{\mathbf{h}}_s). \end{aligned} \tag{20}$$

Here, P_a , P_b , and P_c are both the number of reciprocal lattice points along each direction and the number of points discretizing the a, b, c axes of the cell, while $s_\alpha = \hat{s}_\alpha / P_\alpha$. Importantly, Eq. (20) and its inverse, Eq. (19), can be evaluated using a 3-D Fast Fourier transform (3D-FFT) in order $N \log N$. A spherical cutoff can be introduced in reciprocal space by simply assuming that $n(\mathbf{r})$ is described by a basis in which $\bar{n}(\mathbf{g}) \equiv 0$ when $\hbar^2 |\mathbf{g}|^2 / 2m_e > E_{\text{cut}}^{(\text{density})}$.

Next, consider a function, $f(\mathbf{r})$ with plane-wave expansion coefficients,

$$\begin{aligned} \bar{f}(\mathbf{g}) &= \int_{D(\mathbf{h})} d\mathbf{r} \exp[-i\mathbf{g}\cdot\mathbf{r}]f(\mathbf{r}) \\ &= V \int_0^1 ds_a \int_0^1 ds_b \int_0^1 ds_c \\ & \times e^{-2\pi i \hat{g}_a s_a} e^{-2\pi i \hat{g}_b s_b} e^{-2\pi i \hat{g}_c s_c} f(\vec{\mathbf{h}}_s), \end{aligned} \tag{21}$$

that can be described on a finite reciprocal space [cf. Eq. (20)]. In order to express the plane wave expansion coefficients accurately, in terms of a sum over an arbitrary set of equally spaced discrete points in real space as opposed to the continuous integrals given in Eq. (21), or the discretization required by Eq. (20), it useful to introduce the Euler exponential spline,^{22,23}

$$\begin{aligned} \exp\left(\frac{2\pi i \hat{g}_\alpha u}{\tilde{P}_\alpha}\right) &= d_m(\hat{g}_\alpha, \tilde{P}_\alpha) \sum_{\hat{s}=-\infty}^{\infty} M_m(u-\hat{s}) \\ & \times \exp\left(\frac{2\pi i \hat{g}_\alpha \hat{s}}{\tilde{P}_\alpha}\right) + \mathcal{O}\left(\frac{2|\hat{g}_\alpha|}{\tilde{P}_\alpha}\right)^m, \\ d_m(\hat{g}_\alpha, \tilde{P}_\alpha) &= \frac{\exp(2\pi i(m-1)/\tilde{P}_\alpha)}{[\sum_{j=0}^{m-2} M_m(j+1)\exp(2\pi i \hat{g}_\alpha j / \tilde{P}_\alpha)]}, \end{aligned} \tag{22}$$

where \hat{s} is an integer, u is a real number, “ $m-1$ ” is the spline order assumed to be odd and the $M_m(u)$ are the Cardinal B-splines,

$$\begin{aligned} M_2(u) &= 1 - [u - 1], \\ M_m(u) &= \left[\frac{u}{m-1}\right]M_{m-1}(u) + \left[\frac{m-u}{m-1}\right]M_{m-1}(u-1), \\ M_m(u) &\neq 0, \quad 0 < u < m, \\ M_m(u) &= 0, \quad u \leq 0, \quad u \geq m, \\ \sum_{\hat{s}=-\infty}^{\infty} M_m(u-\hat{s}) &= 1, \\ \frac{dM_m(u)}{du} &= M_{m-1}(u) - M_{m-1}(u-1). \end{aligned} \tag{23}$$

Inserting the Euler exponential spline into Eq. (21) yields a well-defined approximation to $\bar{f}(\mathbf{g})$,

$$\begin{aligned} \bar{f}(\mathbf{g}) \approx & [V d_m^*(\hat{g}_a, \tilde{P}_a) d_m^*(\hat{g}_b, \tilde{P}_b) d_m^*(\hat{g}_c, \tilde{P}_c)] \\ & \times \sum_{\hat{s}_a=0}^{\hat{P}_a-1} \sum_{\hat{s}_b=0}^{\hat{P}_b-1} \sum_{\hat{s}_c=0}^{\hat{P}_c-1} e^{-2\pi i \hat{g}_\alpha \hat{s}_\alpha / \hat{P}_\alpha} \\ & \times e^{-2\pi i \hat{g}_b \hat{s}_b / \hat{P}_b} e^{-2\pi i \hat{g}_c \hat{s}_c / \hat{P}_c} f^{(\text{conv})}(\vec{\mathbf{h}}_s), \end{aligned} \tag{24}$$

where

$$\begin{aligned} f^{(\text{conv})}(\vec{\mathbf{h}}_s) &= \int_0^1 ds'_a \int_0^1 ds'_b \int_0^1 ds'_c \\ & \times \sum_{k_a=-\infty}^{\infty} \sum_{k_b=-\infty}^{\infty} \sum_{k_c=-\infty}^{\infty} f(\vec{\mathbf{h}}_s') \\ & \times M_m([s'_a - k_a]\hat{P}_a - \hat{s}_a) \\ & \times M_m([s'_b - \hat{s}_b]\hat{P}_b - \hat{s}_b) M_m([s'_c - k_c]\hat{P}_c - \hat{s}_c), \end{aligned} \tag{25}$$

is the interpolation of $f(\mathbf{r})$ onto the discrete real space grid defined by $s_\alpha = \hat{s}_\alpha / \tilde{P}_\alpha$ and $0 \leq \hat{s}_\alpha \leq \tilde{P}_\alpha - 1$.

Equation (24) can be evaluated using a 3D-FFT in order $N \log N$ provided the interpolation function, $f^{(\text{conv})}(\vec{\mathbf{h}}_s)$, defined on the discrete real space, can be constructed in a computationally efficient manner. In addition, Eq. (24) is smooth and possesses $m-2$ continuous derivatives. Note, if $\tilde{P}_\alpha > m+1$, then each point in the continuous space, $\{s'_a, s'_b, s'_c\}$, is mapped to m^3 unique points on the discrete grid indexed by $\{\hat{s}_a, \hat{s}_b, \hat{s}_c\}$ due to the finite support of the $M_m(u)$ [see Eq. (23)].

It is now a simple matter to generate a computationally efficient and well-defined approximation to the Fourier coefficients, $\bar{n}(\mathbf{g})$, of an electron density $n(\mathbf{r})$ that is assumed to be nonzero only in the small cell described by $\vec{\mathbf{h}}_s$. First, given that $\bar{n}_s(\mathbf{g}_s)$, defined in Eq. (11), exists on a finite reciprocal space, the identity given in Eq. (20) holds. Thus, the discrete form of the density can be inserted into Eq. (25) and the integrals performed using trapezoidal rule integration without loss of generality to yield the desired interpolation from the small cell to the large cell,

$$\begin{aligned}
n^{(\text{conv})}(\vec{\mathbf{h}}\mathbf{s}) &= \left[\frac{V_s}{V} \right] \left[\frac{1}{P_{a,s} P_{c,s} P_{c,s}} \right] \\
&\times \sum_{\hat{s}'_a=0}^{P_{a,s}-1} \sum_{\hat{s}'_b=0}^{P_{b,s}-1} \sum_{\hat{s}'_c=0}^{P_{c,s}-1} \sum_{k_a=-\infty}^{\infty} \sum_{k_b=-\infty}^{\infty} \sum_{k_c=-\infty}^{\infty} n_s(\vec{\mathbf{h}}_s s') \\
&\times M_m([s'_a + S_{a,s} - k_a] \tilde{P}_a - \hat{s}_a) M_m([s'_b + S_{b,s} - k_b] \\
&\times \tilde{P}_b - \hat{s}_b) M_m([s'_c + S_{c,s} - k_c] \tilde{P}_c - \hat{s}_c). \quad (26)
\end{aligned}$$

Here, $\{P_{a,s}, P_{b,s}, P_{c,s}\}$ are defined by the size of the small cell reciprocal space [through the plane-wave-energy cutoff, $E_{\text{cut}}^{(\text{short,density})} = 4E_{\text{cut}}^{(\text{short})}$], $s'_\alpha = \hat{s}'_\alpha D_{\alpha,\alpha} / P_{\alpha,s}$, $\mathbf{S}_s = \vec{\mathbf{h}}^{-1} \mathbf{R}_c$, and $V_s/V = \det \tilde{\mathbf{D}}$ while the $\{\tilde{P}_a, \tilde{P}_b, \tilde{P}_c\}$ are defined by the size of the large cell reciprocal space [through the cutoff, $E_{\text{cut}}^{(\text{long,density})} = 4E_{\text{cut}}^{(\text{long})}$].

The desired plane wave expansion of the electronic density in the large cell reciprocal space, $\bar{n}(\mathbf{g})$, is constructed by inserting $n^{(\text{conv})}(\vec{\mathbf{h}}\mathbf{s})$ into Eq. (24) and performing a 3D-FFT. Note, in the limit, $\tilde{P}_a = P_{a,s}$, $\tilde{P}_b = P_{b,s}$, $\tilde{P}_c = P_{c,s}$ or $E_{\text{cut}}^{(\text{short})} = E_{\text{cut}}^{(\text{long})}$, and $\vec{\mathbf{h}} = \vec{\mathbf{h}}_s$, then $\bar{n}_s(\mathbf{g}_s) \equiv \bar{n}(\mathbf{g})$ because Eq. (20) is exact for a finite reciprocal space and the Euler exponential splines are exact at the knots.^{22,23} Importantly, Eq. (26) can be evaluated in order $N_s m^3$ and the (dense) discrete real space grid spanning the small cell, $\vec{\mathbf{h}}_s$, and the (sparse) discrete real space grid spanning the large cell, $\vec{\mathbf{h}}$, need not be commensurate. In addition, the separable form of the $M_m(p)$, which is a consequence of the choice $\vec{\mathbf{h}}^{-1} \vec{\mathbf{h}}_s = \tilde{\mathbf{D}}$, allows the required $M_m(p)$ to be evaluated independently in order $m N_s^{1/3}$. Thus, the overall computational cost of constructing $\bar{n}(\mathbf{g})$ is $N \log N$ (dominated by the FFT). Finally, the resulting $\bar{n}(\mathbf{g})$ [i.e., obtained by inserting Eq. (26) into Eq. (24)] is continuously differentiable with respect to the expansion coefficients of the orbitals, $\tilde{\psi}_{j,s}(\mathbf{g}_s)$, defined in Eq. (11).

III. RESULTS

Three systems have been selected to test the new method. The first system, a Gaussian charge density interacting with a single point charge, can be solved analytically. It was chosen because it provides a clear and unambiguous demonstration of the accuracy of the new approach. The second system, a single “*ab initio*” water molecule embedded within a bath of water molecules described by the TIP3P force field, does not possess an analytical solution but it is a sufficiently small system that accurate results can be generated using standard methods, yet, it is sufficiently realistic to demonstrate the behavior of the new technique on a non-trivial problem. The third system, an enzyme, human carbonic anhydrase II (HCA II), solvated in liquid water, is a large complex system that demonstrates the true utility of the new method, as it cannot be approached using standard techniques. That is, in this latter system, the active site region and a small number of water molecules are described within the GGA-DFT, while the remainder of the system ($\sim 30,000$ atoms) is treated using an empirical force field.

TABLE I. The interaction of a Gaussian charge density, $\kappa = 3.779\,454 \text{ \AA}^{-1}$, with a point charge at distance, r_0 from its center is presented as a function of large cell plane-wave cutoff and B-spline interpolation order. The large cell size was fixed at $L_l = 20 \text{ \AA}$ on edge. The small cell size was fixed at $L_s = 4 \text{ \AA}$ on edge and the small cutoff was fixed at $E_{\text{cut}}^{\text{short}} = 120 \text{ Ry}$. The electrostatic division parameter was set to be $\alpha = 6/L_s$ and $\Delta E_{\text{ext}} = E_{\text{ext}} - E_{\text{ext}}^{(\text{exact})}$.

r_0 (\AA)	$E_{\text{cut}}^{(\text{long})}$ (Rydberg)	m	E_{ext} (Hartree)	ΔE_{ext} (Kelvin)
4	4	4	-0.132 296	1
		6	-0.132 297	1
		8	-0.132 297	1
		4	-0.132 293	0
		6	-0.132 293	0
		8	-0.132 293	0
6	4	4	-0.088 186	3
		6	-0.088 185	3
		8	-0.088 185	3
		4	-0.088 198	1
		6	-0.088 198	1
		8	-0.088 198	1
8	4	4	-0.066 126	7
		6	-0.066 125	7
		8	-0.066 125	7
		4	-0.066 149	1
		6	-0.066 149	1
		8	-0.066 149	1

First, in order to validate the new methodology, consider a simple Gaussian electron density,

$$n(\mathbf{r}) = \left(\frac{\kappa^2}{\pi} \right)^{3/2} \exp(-\kappa^2 r^2). \quad (27)$$

The interaction of this density with a point charge located an arbitrary distance, r_0 , away from its center can be determined, analytically, $E_{\text{ext}}^{(\text{exact})} = \text{erf}(\kappa r_0)/r_0$. In Table I the convergence of the total external energy to the analytical value is presented as a function of the large cell plane-wave cutoff and Cardinal B-spline interpolation order, for various choices of r_0 . The calculations were performed using the cluster boundary condition technique of reference¹¹ and fixed small cell plane-wave cutoff [$E_{\text{cut}}^{(\text{short})} = 120 \text{ Ry}$]. In general, it can be seen that low B-spline interpolation orders and small plane-wave cutoffs in the large cell, $E_{\text{cut}}^{(\text{long})}$, are sufficient to produce accurate results, indicating that the new method is both valid and efficient.

Next, a more realistic problem, a system of 64 water molecules in a cubic box, subject to periodic boundary conditions, is examined. Although relatively small, a 64 molecule system has been shown to be a reasonable model of liquid water. Therefore, a system of 64 empirical TIP3P model water molecules was equilibrated at the state point $\{\rho = 1 \text{ g/cm}^3, T = 300 \text{ K}\}$ for 1 nanosecond of simulation time. One TIP3P water molecule was then replaced by an *ab initio* water molecule (8 valence electrons and three “*ab initio* atoms” or “ions”). The electrons were allowed to interact with the TIP3P model molecules and the H^+ and O^{6+} ions via pseudopotentials fit by the authors and standard Troullier–Martins pseudopotentials,²⁴ respectively. In addi-

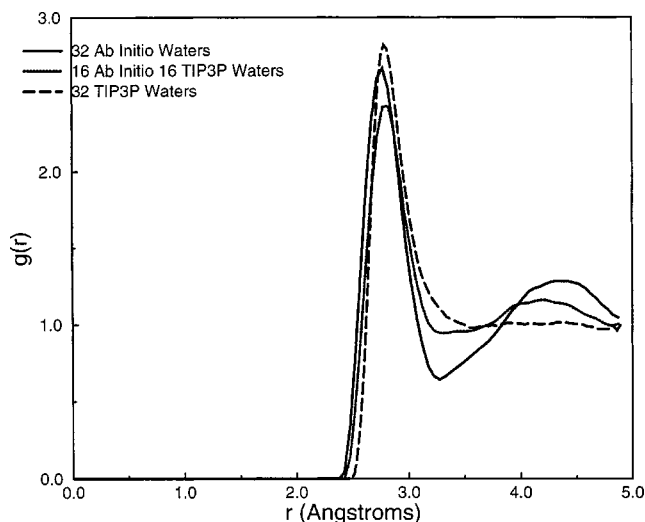


FIG. 1. The oxygen–oxygen radial distribution function for a system consisting of 16 TIP3P water molecules and 16 *ab initio* molecules at the state point $\{\rho=1 \text{ g/cm}^3, T=300 \text{ K}\}$ is presented (dotted line). Comparisons are made to distribution functions generated using fully *ab initio* (solid line) and a fully empirical model (dashed line).

tion, the ions were permitted to interact with the TIP3P molecules via electrostatic and van der Waals forces. The B-LYP exchange–correlation density functional^{18,19} was used in all the calculations, in conjunction with a small cell plane-wave cutoff of $E_{\text{cut}}^{(\text{short})}=100 \text{ Ry}$. The model is described in more detail in Ref. 25 (Fig. 1, for example, shows the radial distribution function of a 50/50 mixture of TIP3P-*ab initio* water molecules). In Table II the convergence of the electronic energy is shown as a function of the large cell plane-wave cutoff and Cardinal B-spline interpolation order. As can be seen from the table, rapid convergence with both parameters is achieved. In addition, the small cell edge, L_s , can be taken to be much less than the large cell edge without loss of accuracy.

Last, an HCA II enzyme solvated in liquid water, is considered. In detail, the system consists of a 260-residue HCA II enzyme (complete with catalytic zinc), solvated by 8,859 waters, for a total of 30,652 atoms subject to periodic boundary conditions. Clearly, a fully *ab initio* treatment of such a large system is not feasible, at present. However, a hybrid model, wherein only the catalytic zinc, the side chains of active site residues, HIS 94, HIS 96, HIS 119, THR 199, GLU 106, and the seven water molecules in the active site are treated using an *ab initio* description, can be studied. Thus, 204 valence electrons of 68 atoms in the active site (see Fig. 2) are treated at an *ab initio* level, while the remainder of the system is treated using the empirical CHARMM22 all-atom force field which includes TIP3P water model.²⁶ Briefly, the electrons are assumed to interact with “*ab initio* atoms” via standard Troullier–Martins pseudopotentials²⁴ and with “empirical atoms” via pseudopotentials³¹ (see also Refs. 27 and 28). The B-LYP, density functional^{18,19} was employed to treat exchange and correlation. *Ab initio* atoms (ion-cores) were permitted to interact with neighboring “empirical atoms” via appropriate bond, bend, torsion, one-four, van der Waals and Coulomb forces.³¹ The parameters were

TABLE II. The total electronic energy of a single *ab initio* water molecule immersed in a bath of TIP3P molecules as a function of large cell plane cutoff and spline interpolation order. The large cell size was fixed by the state point, $L_l=12.43 \text{ \AA}$, on edge. The small cell cutoff was fixed at 100 Ry. The electrostatic division parameter was set to $\alpha=6/L_s$ and $\Delta E_{\text{tot}}=E_{\text{tot}}-E_{\text{tot}}^{(\text{std})}$, where $E_{\text{tot}}^{(\text{std})}=-20.28767$ is the result of a standard calculation with $L_s=L_l=12.43 \text{ \AA}$.

L_s (\AA)	$E_{\text{cut}}^{(\text{long})}$ (Rydberg)	m	E_{tot} (Hartree)	ΔE_{tot} (Kelvin)	
6	4	4	-20.279 80	2480	
		6	-20.281 27	2020	
	8	8	-20.281 33	2002	
		4	-20.281 08	2080	
	6	8	6	-20.281 33	2000
			8	-20.281 34	2000
8	4	4	-20.286 84	260	
		6	-20.287 18	150	
	8	8	-20.287 18	150	
		4	-20.287 11	175	
	6	8	6	-20.287 18	150
			8	-20.287 18	150
9	4	4	-20.287 73	-20	
		6	-20.287 90	-70	
	8	8	-20.287 90	-70	
		4	-20.287 87	-60	
	6	8	6	-20.287 90	-70
			8	-20.287 90	-70

obtained by enforcing good agreement between mixed models, fully empirical models and fully *ab initio* models of relevant fragments. For example, the minimum energy geometry of hybrid model $\text{CH}_3\text{CO}-(\text{HIS})-\text{NHCH}_3$ deviates at most ≈ 3 degrees in the bend angles and $\approx 0.02 \text{ \AA}$ in the bond

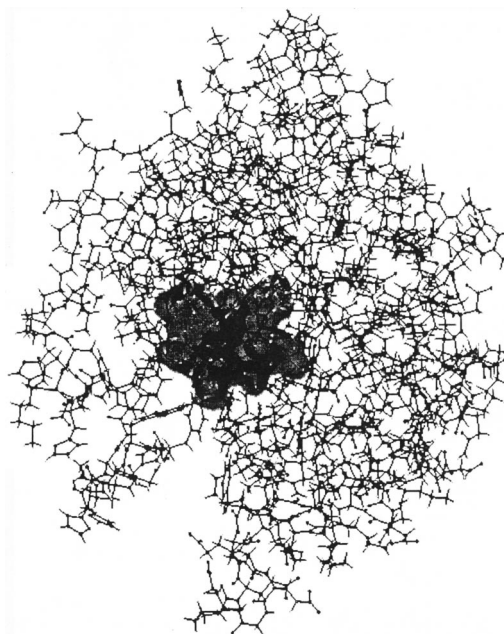


FIG. 2. Human carbonic anhydrase treated using the mixed *ab initio*/empirical force-field-based approach described in the text. The full enzyme wherein the wire frame represents atomic sites and the blue cloud represents the electron density of the valence electrons associated with “*ab initio* atoms.”

TABLE III. The total electronic energy of the active site of HCA II immersed in a bath of TIP3P molecules and CHARMM22 model amino acid residues as a function of large cell plane cutoff and spline interpolation order. The large cell size is fixed by the state point, 66.7 Å, on edge. The small cell size was fixed at 19 Å on edge and the small cell cutoff was fixed at 70 Ry. The electrostatic division parameter was set to be $\alpha=7/L_s$, and the accuracy measure is defined to be $\Delta E_{\text{tot}}=E_{\text{tot}}(E_{\text{cut}}^{(\text{long})}, m) - E_{\text{tot}}(4,8)$.

$E_{\text{cut}}^{(\text{long})}$ (Rydberg)	m	E_{tot} (Hartree)	ΔE_{tot} (Kelvin)
1	4	-548.549 312	1022.0
	6	-548.552 461	28.0
	8	-548.552 545	1.5
2	4	-548.551 775	244.0
	6	-548.552 540	2.9
	8	-548.552 549	0.1
4	4	-548.552 373	55.7
	8	-548.552 548	0.3
	8	-548.552 549	0.0

lengths from the standards (CHARMM and fully *ab initio* treatments as appropriate). More details and a complete analysis of the electronic structure of the fragments is provided in Ref. 25.

The HCA II/water system described above was prepared by taking the crystallographic configuration of the enzyme (PDB identification label, "1RAY")²⁹ and immersing it in TIP3P water. Next, a 1 ns constant temperature molecular dynamics calculation was performed using a fully empirical treatment.²⁶ This was followed by a 1 ns constant pressure molecule dynamics calculation. At this point, the hybrid model was introduced. In Table III, the convergence of the electronic energy for a representative configuration taken from the simulation of the hybrid model, is shown versus the large cell, plane-wave cutoff and the Cardinal B-spline interpolation order. As in previous examples, accurate energies are obtained for low spline orders and plane-wave cutoffs. The parallel efficiency of the algorithm is also good (see Table IV). The results, therefore, indicate that large complex systems can be studied efficiently using the new methodology. Further studies of the physical and chemical properties of the hybrid model HCA II/water system are presented in Ref. 30.

TABLE IV. The parallel efficiency, as defined by the CPU time for 8 processor run divided by the CPU time for a P₁ processor run, obtained in the hybrid model simulation study of the HCA II enzyme solvated in TIP3P water. Superlinear increases in efficiency are observed as the problem begins to fit in the cache at larger number of processors.

Processors	Efficiency
8	1.0
10	1.2
14	1.8
16	2.0
20	2.6
24	3.2

IV. CONCLUSION

In summary, a new method has been designed that greatly increases the efficiency of mixed *ab initio*/empirical model descriptions of chemical systems treated using GGA-DFT electronic structure theory and a plane-wave basis set. In particular, the important case, that the chemically active components are both localized in a small region of space and embedded in a large cell containing a bath of chemically inert constituents, is considered. A Cardinal B-spline-based formalism was used to derive smooth, differentiable, and rapidly convergent expressions for the terms in the electronic energy that contains both long- and short-range length contributions. The new formalism, thus, permits the short range electron-electron and electron-atom interactions, originating from the small region of space where the electrons are localized, to be treated using a large plane-wave basis and the long-range interactions of the electrons with the remainder of the system to be treated using a small basis. The methodology is, therefore, computationally efficient and scales as $N \log N$ at fixed size of the chemically active subsystem. Also, it allows clusters, wires, surfaces and solids/liquids to be treated without loss of generality, in marked contrast to existing plane-wave based "mixed" methods. The new technique was validated on analytically solvable model as well as realistic systems including a mixed *ab initio*/empirical model of liquid water and the human carbonic anhydrase II enzyme, solvated in a bath of empirical model water molecules. Therefore, it is now possible to perform plane-wave-based *ab initio* molecular dynamics and electronic structure calculations of large complex systems that contain both chemically active and chemically quiescent components, efficiently and accurately.

ACKNOWLEDGMENTS

This research was supported by PRF 32139-AC and NSF CHE-965015 (G.J.M.), PRF 33256-G, Research Corporation RI0218, NYU Whitehead Fellowship in Biomedical and Biological Science, NSF CHE-9875824 (M.E.T.), NSF CHE 9623017 (D.A.Y.) and EIA-0081307 (G.J.M. and M.E.T.). The authors thank Professor Ursula Röthlisburger for providing some of the pseudopotentials used in the work.

¹A. Warshel and M. Levitt, J. Mol. Biol. **103**, 227 (1976).

²K. P. Eurenius, D. C. Chatfield, B. R. Brooks, and M. Hodoseck, Int. J. Quantum Chem. **60**, 1189 (1996).

³P. D. Lyne, M. Hodoseck, and M. Karplus, J. Phys. Chem. A **103**, 3462 (1999).

⁴M. Eichinger, P. Tavan, J. Hutter, and M. Parrinello J. Chem. Phys. **110**, 10452 (1999).

⁵Y. Zhang, T. Lee, and W. Yang, J. Chem. Phys. **110**, 46 (1999).

⁶Y. Zhang, T. Lee, and W. Yang, J. Chem. Phys. **112**, 3483 (2000).

⁷R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).

⁸G. Galli and M. Parrinello, Comput. Mater. Sci. **3**, 283 (1991).

⁹M. E. Tuckerman, D. Marx, M. L. Klein, and M. Parrinello, Science **275**, 817 (1997).

¹⁰M. Diraison, M. E. Tuckerman, and G. J. Martyna, J. Chem. Phys. **111**, 1096 (1999).

¹¹G. J. Martyna and M. E. Tuckerman, J. Chem. Phys. **110**, 2810 (1999).

¹²P. Minary, M. E. Tuckerman, K. A. Pihakari, and G. J. Martyna, (in preparation).

¹³M. Eichinger, P. Tavan, J. Huuter, and M. Parrinello, J. Chem. Phys. **110**, 10452 (1999).

- ¹⁴U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, *J. Chem. Phys.* **103**, 8577 (1995).
- ¹⁵W. Kohn and L. J. Sham, *Phys. Rev.* **140**, A1133 (1965).
- ¹⁶P. Hohenberg and W. Kohn, *Phys. Rev.* **136**, B86 (1964).
- ¹⁷*Theory of Inhomogeneous Electron Gas*, edited by S. Lundqvist and N. H. March (Plenum, New York, 1983).
- ¹⁸A. D. Becke, *Phys. Rev. A* **38**, 3098 (1988).
- ¹⁹C. Lee, W. Yang, and R. G. Parr, *Phys. Rev. B* **37**, 785 (1988).
- ²⁰G. Bachelet, D. Hamann, and M. Schluter, *Phys. Rev. B* **26**, 4199 (1982).
- ²¹D. K. Remler and P. A. Madden, *Mol. Phys.* **70**, 921 (1990).
- ²²C. K. Chu, *An Introduction to Wavelets* (Academic, Boston, 1992).
- ²³I. J. Schoenberg, *Cardinal Spline Interpolation* (Society for Industrial and Applied Math, Philadelphia, 1973).
- ²⁴N. Troullier and J. L. Martins, *Phys. Rev. B* **43**, 1993 (1991).
- ²⁵D. A. Yarne, M. E. Tuckerman, M. L. Klein, and G. J. Martyna (in preparation).
- ²⁶A. MacKerell, Jr., D. Bashford, M. Bellott *et al.* *J. Phys. Chem. B* **102**, 3586 (1998).
- ²⁷Y. Zhang, T. Lee, and W. Yang, *J. Chem. Phys.* **110**, 46 (1999).
- ²⁸J. Gao, P. Amara, C. Alhambra, and M. J. Field, *J. Phys. Chem. A* **102**, 4714 (1998).
- ²⁹B. M. Jonsson, K. Hakansson, and A. Liljas, *FEBS Lett.* **322**, 186 (1993).
- ³⁰D. A. Yarne, U. Röthlisburger, M. E. Tuckerman, M. L. Klein, and G. J. Martyna (in preparation).
- ³¹U. Röthlisburger (private communication).