An image-based reaction field method for electrostatic interactions in molecular dynamics simulations of aqueous solutions

Yuchun Lin,[1,2] Andrij Baumketner,[1,a),b)] Shaozhong Deng,[2] Zhenli Xu,[2] Donald Jacobs,[1] and Wei Cai[2,a)]

[1]*Department of Physics and Optical Science, University of North Carolina at Charlotte, Charlotte, North Carolina 28223, USA*
[2]*Department of Mathematics and Statistics, University of North Carolina at Charlotte, Charlotte, North Carolina 28223, USA*

In this paper, a new solvation model is proposed for simulations of biomolecules in aqueous solutions that combines the strengths of explicit and implicit solvent representations. Solute molecules are placed in a spherical cavity filled with explicit water, thus providing microscopic detail where it is most needed. Solvent outside of the cavity is modeled as a dielectric continuum whose effect on the solute is treated through the reaction field corrections. With this explicit/implicit model, the electrostatic potential represents a solute molecule in an infinite bath of solvent, thus avoiding unphysical interactions between periodic images of the solute commonly used in the lattice-sum explicit solvent simulations. For improved computational efficiency, our model employs an accurate and efficient multiple-image charge method to compute reaction fields together with the fast multipole method for the direct Coulomb interactions. To minimize the surface effects, periodic boundary conditions are employed for nonelectrostatic interactions. The proposed model is applied to study liquid water. The effect of model parameters, which include the size of the cavity, the number of image charges used to compute reaction field, and the thickness of the buffer layer, is investigated in comparison with the particle-mesh Ewald simulations as a reference. An optimal set of parameters is obtained that allows for a faithful representation of many structural, dielectric, and dynamic properties of the simulated water, while maintaining manageable computational cost. With controlled and adjustable accuracy of the multiple-image charge representation of the reaction field, it is concluded that the employed model achieves convergence with only one image charge in the case of pure water. Future applications to *pKa* calculations, conformational sampling of solvated biomolecules and electrolyte solutions are briefly discussed. *© 2009 American Institute of Physics.*
[doi:10.1063/1.3245232]

I. INTRODUCTION

Due to their long-range nature, Coulomb interactions play a critically important role in computer simulations of chemical and biological systems in condensed phase.[1–6] A diverse class of physical properties and processes studied in computer simulations are affected by the treatment of the long-range electrostatic interactions, including reaction rates, equilibrium structures, electronic spectra, and charge separation and transfer.[7–12] Over the past decades, a wide range of approaches have been introduced to treat electrostatic interactions that differ greatly in their conceptual framework, complexity, and technical detail. The majority of these approaches can be broadly divided into implicit and explicit categories.

In implicit solvation models, the solute is treated in atomic detail while the solvent is modeled as a dielectric continuum. The effect of solvent on the solute is modeled through Poisson or Poisson–Boltzmann (PB) equations.[13–16]

A number of algorithms are available to solve the PB equation numerically, including finite-difference,[17–19] finite-element,[20–23] and boundary-element methods.[24–27] The main motivation behind the development of these models is to reduce the computational cost.[28,15] With water molecules integrated out, implicit solvents present far fewer degrees of freedom in computer simulations than the explicit solvents, thus bringing down the associated computational cost. One drawback is a demanding numerical implementation. Despite much recent progress in developing finite-difference methods[19,29,30] and formulating approximate treatments,[31] the solution of a three-dimensional differential PB equation for a macromolecule of arbitrary shape still presents a challenging computational problem. Another difficulty lies in the approximate representation of discrete molecular solvent as a continuum medium. Obviously, this approximation breaks down on small length scales where the size of the water molecule becomes important. Such a situation is encountered in many implicit solvation models, where there is growing evidence that the common description in which a sharp boundary exists separating a low-dielectric medium of solute and a high-dielectric medium of solvent is not adequate.[32] A more gradual transition between solute and solvent has been

suggested as more appropriate.[33] Another example where the continuum description fails is the formation of a salt bridge between oppositely charged amino acids in proteins that requires participation of one water molecule.[34] It is clear from these examples that it would be advantageous for implicit solvation models to treat some parts of the solvent explicitly, namely, in atomic detail.

Explicit solvents represent entire simulated systems, solute and solvent, in atomic detail, assigning fixed charges to all atoms present in the system. There are a number of ways to compute interactions among charges in explicit solvents, including simple cutoff methods,[35] a variety of lattice-sum methods such as Ewald summation,[36,37] particle-mesh Ewald (PME)[38,39] or particle-particle particle-mesh,[40,41] and fast multipole methods (FMM).[42,43] Of these, the lattice-sum methods are considered to be most accurate.[35] Simulations in explicit solvents are performed under periodic boundary conditions (PBCs) to avoid surface effects. Technically, this means that instead of considering one, central simulation box, an infinite number of identical boxes are distributed throughout physical space on a lattice. Periodicity eliminates the boundary of the box and thus makes it possible to obtain bulk properties of the simulated material. It also induces artifacts resulting from the interactions between particles that are located in different images of the box. For small sizes of box L these interactions may become significant given the long-range nature of the Coulomb forces.[44,45] In the limit of $L \rightarrow \infty$, these artifacts vanish, making increasing the size of the simulation box a good strategy for reducing the periodicity artifacts in lattice-sum simulations. Although large L improve accuracy, they also increase the computational cost of the simulations, negatively affecting their practical use. At present, the periodicity-induced artifacts in explicit solvent simulations are not well understood.[46] In homogeneous media where the size of the constituent molecules is much smaller than the size of the simulation cell, such as pure solvents, they are not significant. Important thermodynamic and dynamic characteristics of liquid water including diffusion coefficient[47] and dielectric constant[48] were seen to be insensitive to the size of the simulation box for sufficiently large boxes. In simulations of solvated macromolecules, on the other hand, the artifacts might be substantial.[49] It was reported recently[44] that increasing L reduces the stability of a short alanine-based α-helical peptide. While the exact mechanism of this destabilization remains unclear, with poor equilibration suggested as one possible reason,[45] the direct interactions between periodic peptide images cannot be ruled out.[49] A distinct advantage of implicit solvents is that they do not require PBCs and thus are not affected by periodicity artifacts.

Considering their complementary strengths, a promising strategy to design better solvation schemes is to combine in one model elements of both implicit and explicit approaches. Receiving much recent attention,[50–52] this research strategy led to the development of numerous explicit/implicit models, also referred to as hybrid. All hybrid models share a common design principle: a central part of the simulated system containing the solute and some solvent is considered in atomic detail while the remaining part is treated as dielectric con-

tinuum. These two parts are separated by a buffer layer in which molecules are treated atomically but they experience forces different from those present in the central part. There are two types of potentials that apply to the charges in the explicit solvent part. First, it is the direct Coulomb potential through which they interact with one another, Φ_S. Second, it is an indirect potential that results from the polarization of the continuum solvent region by the explicit solvent charges and is referred to as reaction field, Φ_{RF}. The total potential inside the explicit solvent region is expressed as $\Phi = \Phi_S + \Phi_{RF}$.

The major difference among hybrid electrostatic solvation models concerns how Φ_{RF} is computed. In one group of models, the geometry of the explicit part is assumed arbitrary and the reaction field is obtained either by solving the PB equation directly[53] or by using an approximate theory for such solutions.[51,54] Both cases involve extensive numerics and thus tend to be computationally expensive. A much larger group of methods[50] takes advantage of the exact solutions that exist for select solute geometries such as a sphere or a plane.[55] Most common in hybrid models[56–58] is the spherical geometry for which a series expansion of the reaction field was developed by Kirkwood.[59,60] Although, in theory, this expansion produces reaction fields of arbitrary accuracy, its convergence is slow near the sphere boundary, making its practical applications problematic. An alternative approach is to employ the method of image charges,[61] where the reaction field appears to be created by one or more fictitious charged particles, or image charges, located outside of the region where the potential is computed. The result is a considerable improvement in accuracy and speed of the reaction field calculations. Friedman[62] was first to apply the image charges method in the context of the solvation problem. He derived an approximation for charges placed in a spherical cavity embedded in a medium with dielectric permittivity ε, which uses one image charge to reconstruct the reaction field with an accuracy of up to $O(1/\varepsilon)$. In the media with high dielectric constant, such as water, this approximation is sufficiently accurate and it becomes exact in the metallic boundary limit $\varepsilon \rightarrow \infty$.

Due to its favorable accuracy-to-cost ratio, the Friedman image approximation has been extensively used previously in implicit and hybrid explicit/implicit solvation models.[37,63] Within the hybrid solvation approach, several groups[64–67] independently proposed models in which solutes are encapsulated in spherical nanodroplets filled with molecular solvent. To keep the solvent molecules inside the droplets, repulsive boundary potentials are applied. How the simulated systems are affected by these potentials is extremely important. As the main purpose of solvation models is to mimic the bulk properties of the solvent, it is desirable that their effect be minimal. This is not always the case in the studies published to date. In all reported simulations on liquid water, for instance, strong surface effects were seen.[64–67] Opinions about how to deal with these effects differ. Some papers find that they can be fully or partially removed[65,67] by carefully adjusting the parameters of the boundary potentials. Others[32,64,66] failed to find such optimal parameter sets, arguing instead that the droplet model is not suitable for the

simulations of bulk properties. Regardless of whether the bulk behavior emerges or not in nanodroplets, the need to adjust parameters becomes a major inconvenience in this model, hampering its use as a general simulation tool.

The discussion above points out two major weaknesses in the current hybrid solvation methods based on the image charges reaction field approach. (I) The Friedman expression for Φ_{RF} is approximate. For systems with moderate ε, its accuracy may not be sufficient. For a general-purpose solvation model, it is highly desirable that the reaction field can be computed with arbitrary accuracy for any ε. (II) The surface effects either are non-negligible or cannot be removed easily. If a solvation method is to be applied on a wide scale and to a diverse class of systems, the surface effects need to be eliminated completely. In this paper we develop a new hybrid explicit/implicit solvation model that overcomes both these difficulties.

Toward the first difficulty, we showed in a number of recent publications[68–70] that better approximations for the reaction field of spherical cavities can be constructed using multiple image charges. These approximations are based on numerical quadratures of the line image charges representation,[71–74] in which the locations of the point image charges are selected on the basis of the Gauss–Radau quadrature points. The number of image charges N_i controls the accuracy of the approximations for varying ε. While for high ε one image charge might be sufficient, media with low ε require a larger number N_i. In a comparative analysis, the multiple-image method is more than an order of magnitude faster than the Kirkwood expansion at the same level of accuracy. When combined with the FMM,[42,43] the multiple image approximation has the potential to calculate electrostatic interactions for N charges inside the spherical cavity in $O(N)$ operations. In addition to pure solvent, our multiple-image method was extended to ionic solutions.[69,70,75]

To address the second difficulty, we apply PBCs for nonelectrostatic interactions, which are the best method to suppress surface effects in computer simulations. The PBCs have been successfully employed in the context of the reaction field calculations previously,[55,58,76–79] so their use in the present work is justified. The spherical geometry of the reaction field cavity dictates the choice of the simulation cell type. We use a truncated octahedron (TO) in our model as it offers one of the best cost per simulated particle ratios among different types of cell geometries. Full description of the proposed solvation model is presented in the first part of the paper, Sec. II.

The second part, Sec. IV, is concerned with the application to pure water. We chose water to validate our model because it is the primary solvent in all biological processes. Our ultimate goal is to investigate problems of biological importance such as protein folding or aggregation. If a solvation model works well for pure water, there is a high likelihood that it will also work for solutes immersed in water. This is why so many previous electrostatics methods focused on this solvent.[55,58,66,77,78,80,81] Our model depends on three main parameters: the size of the spherical cavity, the thickness of the buffer layer, and the number of image charges that are used for the reaction field calculations. The effect of
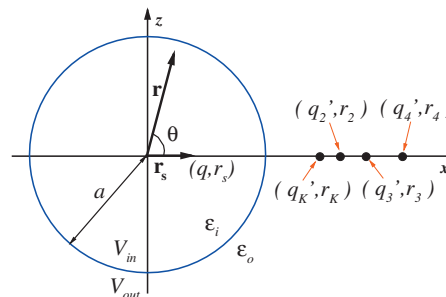

FIG. 1. An illustration of how the multiple-image method is applied to compute reaction field in a spherical cavity V_{in} embedded in the solvent bath V_{out}. The radius of the sphere is a, the dielectric permittivity inside the sphere is ε_i, and that outside of it is ε_o. The polarization of the solvent V_{out} by the source charge q at position $\mathbf{r}_s$ results in the reaction field $\Phi_{RF}(r,\theta)$ that is approximated by the potential created by auxiliary charges, referred to as image charges, q'_K, q'_i, $i \geq 2$, located at positions r_K, r_i, $i \geq 2$, respectively. The total number of image charges N_i controls the accuracy of the approximation, as explained in the main text. The Friedman model (Ref. 62) corresponds to having one image charge q_K located at r_K.

these parameters is systematically investigated for many structural, static, and dynamical properties of water. Based on these investigations, we derive minimal values of parameters that produce pair distribution functions, diffusion coefficient, and dielectric permittivity of the studied water model in very good agreement with the results of the PME calculations. Our conclusions and outlook are presented in Sec. V.

II. THEORY

A. Reaction field in the multiple-image charges approach

Our model for computing reaction fields is based on the spherical geometry of the solute. Consider a local volume Γ of spherical shape and dielectric permittivity ε_i embedded in an infinite solvent of dielectric permittivity ε_o, as shown in Fig. 1. The total electrostatic potential $\Phi(\mathbf{r})$ in this setting satisfies the following PB equation:

$$\varepsilon_i \nabla^2 \Phi(\mathbf{r}) = -\rho_{in}(\mathbf{r}), \quad \mathbf{r} \in V_{in}, \tag{1a}$$

$$[\nabla^2 - \lambda^2]\Phi(\mathbf{r}) = 0, \quad \mathbf{r} \in V_{out}, \tag{1b}$$

where the charge distribution $\rho_{in}(\mathbf{r}) = \Sigma_i q_i \delta(\mathbf{r} - \mathbf{r}_i)$ inside V_{in} contains all explicit charges of the solute and solvent molecules, and λ is the inverse Debye–Hückel screening length. In the pure solvent $\lambda = 0$, while in solutions with nonzero ionic strength $\lambda \neq 0$. On the interface, the continuity of the tangential component of the electric field and the normal component of the displacement field requires that

$$\Phi_{in} = \Phi_{out} \quad \text{and} \quad \varepsilon_i \frac{\partial \Phi_{in}}{\partial \mathbf{n}} = \varepsilon_o \frac{\partial \Phi_{out}}{\partial \mathbf{n}}, \tag{2}$$

where $\mathbf{n}$ is the outward normal of the surface of Γ, and Φ_{in} and Φ_{out} are the values of the potential inside and outside the cavity, respectively.

We will focus our discussion in this section on a single charge q placed at position $\mathbf{r}_s$ inside the cavity, as solution to Eq. (1) for multiple source charges can be obtained using the linear superposition principle. The solution of Eqs. (1) and (2) at position $\mathbf{r}$ inside the cavity V_{in} can be written in terms

of the primary field Φ_S, which results from the source charge, and the reaction field Φ_{RF}, which is created by the source-induced polarization of the solvent in V_{out}. The total potential is expressed as $\Phi = \Phi_S + \Phi_{RF}$. For the case of a spherical cavity with radius a, Φ_{RF} inside the cavity is given by the Kirkwood expansion,[59] which in the pure solvent case $\lambda = 0$ is

$$\Phi_{RF}(r,\theta) = \frac{q}{4\pi\varepsilon_i a} \sum_{n=0}^{\infty} \frac{(\varepsilon_i - \varepsilon_o)(n+1)}{\varepsilon_i n + \varepsilon_o(n+1)} \left(\frac{rr_s}{a^2}\right)^n P_n(\cos\theta),$$

(3)

where P_n are the Legendre polynomials, and θ is the angle between vectors $\mathbf{r}_s$ and $\mathbf{r}$, respectively.

The Kirkwood expansion converges quickly for small values of $r/a \ll 1$ and becomes exact at $r=0$. This property has been exploited in some reaction field-based methods of electrostatic interactions.[76,78] There are problems,[58] however, where Φ_{RF} is computed over the entire cavity V_{in}, requiring a large number of terms in Eq. (3). As a result, slow convergence is a major limitation of the Kirkwood series in practical computer computations. To overcome this limitation, Friedman proposed in 1975 to apply the electrostatics method of image charges to the problem of the dielectric sphere. The method partially sums up the series in Eq. (3), resulting in an approximation that represents Φ_{RF} at an arbitrary $\mathbf{r}$ as generated by a fictitious charged particle located outside of the sphere and referred to as an image charge. The approximation is accurate up to the order of $\varepsilon_i/\varepsilon_o$, which for typical values of $\varepsilon_i = 1$ and $\varepsilon_o = 80$ results in 1% error. The error is set to grow if a system with low dielectric constant ε_o is considered, such as mixtures of solvents or solvents in the supercritical state. Due to its simplicity, the Friedman image method has been widely used in literature.[64–67] Not much work has been done, however, to improve it. We revisited this method recently,[68] in an effort to derive an expression that has an arbitrary accuracy and applies to a wide range of ratios $\varepsilon_i/\varepsilon_o$. Our approach is based on the work of Neumann,[73] where he gave the electrostatic potential in terms of a Kelvin image charge (similar to that for a conductor) plus a line image charge on a ray extending radially from the Kelvin image charge to infinity. We obtained an expression[68] based on Gauss–Radau quadratures that reproduces the reaction field of the line image charge by a set of discrete image charges. Briefly, the derivation is as follows.

First, let $r_K = a^2/r_s$ represent the location of the classic Kelvin image charge and $q_K = \gamma a q/r_s$ be its charge, where $\gamma = (\varepsilon_i - \varepsilon_o)/(\varepsilon_i + \varepsilon_o)$ and r_s is the position of the source charge as shown in Fig. 1. Then, Eq. (3) can be written as

$$\Phi_{RF}(\mathbf{r}) = \frac{q_K}{4\pi\varepsilon_i r_K} \sum_{n=0}^{\infty} \left(\frac{r}{r_K}\right)^n P_n(\cos\theta)$$

$$+ \frac{\delta q}{4\pi\varepsilon_i a} \sum_{n=0}^{\infty} \frac{1}{n+\sigma} \left(\frac{r}{r_K}\right)^n P_n(\cos\theta),$$

(4)

where $\delta = \varepsilon_i(\varepsilon_i - \varepsilon_o)/(\varepsilon_i + \varepsilon_o)^2$ and $\sigma = \varepsilon_o/(\varepsilon_i + \varepsilon_o)$. The first series in Eq. (4) is identified as the potential of the Kelvin image charge, namely,

TABLE I. Gauss–Radau points and weights for $N_i = 1, 2,$ and 3.

N_i	s_m	ω_m
1	−1.000 000 000 000 00	2.000 000 000 000 00
2	−1.000 000 000 000 00	0.500 000 000 000 00
	0.333 333 333 333 33	1.500 000 000 000 00
3	−1.000 000 000 000 00	0.222 222 222 222 22
	−0.289 897 948 556 64	1.024 971 652 376 84
	0.689 897 948 556 64	0.752 806 125 400 93

$$V(q_K, \mathbf{r}_K; \mathbf{r}) = \frac{q_K}{4\pi\varepsilon_i |\mathbf{r} - \mathbf{r}_K|}.$$

(5)

Using the integral identity

$$\frac{1}{n+\sigma} = r_K^{n+\sigma} \int_{r_K}^{\infty} \frac{1}{x^{n+\sigma+1}} dx,$$

(6)

which is valid for all $n \geq 0$ when $\sigma > 0$, we can rewrite the second series in Eq. (4) as

$$\frac{\delta q}{4\pi\varepsilon_i a} \sum_{n=0}^{\infty} \frac{1}{n+\sigma} \left(\frac{r}{r_K}\right)^n P_n(\cos\theta)$$

$$= \int_{r_K}^{\infty} \left[\frac{q'(x)}{4\pi\varepsilon_i x} \sum_{n=0}^{\infty} \left(\frac{r}{x}\right)^n P_n(\cos\theta) \right] dx,$$

(7)

which can now be thought of as the potential created by a line image charge[73,68]

$$q'(x) = \frac{\delta q}{a} \left(\frac{x}{r_K}\right)^{-\sigma}, \quad r_K \leq x.$$

(8)

Therefore, the reaction field inside the cavity $\Phi_{RF}(\mathbf{r})$ is given by the sum

$$\Phi_{RF}(\mathbf{r}) = V(q_K, \mathbf{r}_K; \mathbf{r}) + \int_{r_K}^{\infty} V(q'(x), \mathbf{x}; \mathbf{r}) dx.$$

(9)

Next, to evaluate the integral in Eq. (9), we construct discrete image charges by employing the Gauss–Radau quadrature, yielding

$$\Phi_{RF}(\mathbf{r}) \approx \frac{q_K}{4\pi\varepsilon_i |\mathbf{r} - \mathbf{r}_K|} + \sum_{m=1}^{N_i} \frac{q'_m}{4\pi\varepsilon_i |\mathbf{r} - \mathbf{r}_m|},$$

(10)

where the magnitude and the location of all discrete image charges are defined as

$$q'_m = \frac{\varepsilon_i(\varepsilon_i - \varepsilon_o)}{2\varepsilon_o(\varepsilon_i + \varepsilon_o)} \frac{\omega_m r_m}{a} q, \quad r_m = r_K \left(\frac{2}{1-s_m}\right)^{1+\varepsilon_i/\varepsilon_o}. \quad (11)$$

Here, $\{\omega_m, s_m\}_{m=1}^{N_i}$ are the Gauss–Radau quadrature weights and points.[68] Table I contains the Gauss–Radau quadrature weights and points for $N_i = 1, 2,$ and 3. Furthermore, since $s_1 = -1$ and consequently $r_1 = r_K$, the classical Kelvin image charge q_K and the first discrete image charge q'_1 can be combined, leading to

$$\Phi_{\text{RF}}(\mathbf{r}) \approx \frac{q_K'}{4\pi\varepsilon_i|\mathbf{r} - \mathbf{r}_K|} + \sum_{m=2}^{N_i} \frac{q_m'}{4\pi\varepsilon_i|\mathbf{r} - \mathbf{r}_m|}, \qquad (12)$$

where

$$q_K' = q_K + q_1' = \left(1 + \frac{\omega_1\varepsilon_i}{2\varepsilon_o}\right)q_K.$$

For convenience, in this paper q_K' is referred to as the modified Kelvin image charge and the reaction field method using only q_K' (namely $N_i = 1$ and $\omega_1 = 2$) as the modified Friedman image method.

An illustration of how image charges are constructed for a source charge at position $\mathbf{r}_s$ is shown in Fig. 1. The number of image charges N_i controls the accuracy of the approximation: the greater the number N_i, the better the approximation. Large N_i also add to the computational cost of the problem. Fortunately, the image charges admit treatment by the FMM.[42,43] We showed previously[68] that an $O(N)$ complexity can be achieved in the calculation of electrostatic potential of N charges inside a dielectric sphere with the combination of the FMM and the method of image charges. Our image charge formulation is general and permits computation of Φ_{RF} with arbitrary accuracy at any point $\mathbf{r}$ inside the cavity and for any ratio of the solute and solvent dielectric permittivities $\varepsilon_i/\varepsilon_o$.

B. Integration of the reaction field model with molecular dynamics simulations

1. Role of the buffer layer between explicit and implicit solvents

As Eq. (5) shows, the reaction field diverges if the probe charge is at the boundary $r = a$. To prevent the electrostatic forces from growing too high and breaking the simulation algorithm when particles approach the boundary, droplet solvation models[65,66] employ radii with which reaction fields are computed that are slightly greater than the radii used for the repulsive boundary potential. A similar effect is achieved by the procedure of Alper and Levy[58] who install cutoffs for the forces. The mathematical model that arises from considering two different radii for one spherical cavity is illustrated in Fig. 2(a). The central cavity of a radius R_c, the equivalent of the confining radius, is seen to be separated from the dielectric continuum with permittivity ε_o by an empty buffer layer of thickness τ. The radius used for the reaction field calculations in Eq. (12), which is where the continuum dielectric begins, is $a = R_c + \tau$. As the goal of a hybrid solvation model is to correctly represent bulk solvent, the presence of a vacuum layer is not desirable. For τ on the molecular size scale, the model shown in Fig. 2(a) clearly deviates from the physical system it is meant to represent. An intuitive way to improve this model is to eliminate the vacuum by filling up the buffer layer with water. The resulting scheme is shown in Fig. 2(b). Now, the reaction field is generated by charges with the radius $r < R_c + \tau$ but it acts only on the charges within radius R_c. The advantage is that a more realistic representation of the bulk solvent is achieved for arbitrary buffer thickness τ. For large $\tau \to \infty$, the proper limiting be-
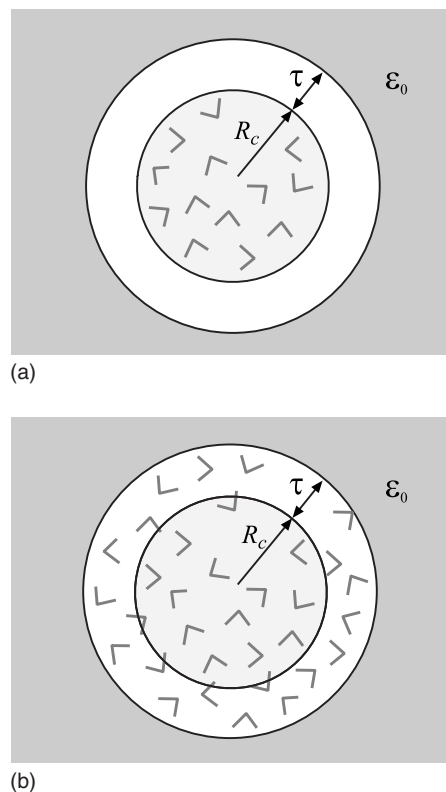

(a)

(b)

FIG. 2. Cartoons illustrating how a buffer layer surrounding a spherical cavity can be introduced. In (a), the reaction field computed at the boundary between explicit and implicit solvents $r = R_c$ diverges. To avoid this singularity, it is common to compute reaction field for particles within R_c using a slightly larger radius $R_c + \tau$. The result is the creation of a vacuum layer of thickness τ between explicit and implicit solvents. For nonzero τ, this mathematical model differs strongly from the underlying physical system. In (b), the buffer layer is filled with solvent; water molecules within radius R_c experience reaction field generated by the molecules within $R_c + \tau$. Even for large τ, the model correctly represents homogeneous solvent.

havior is recovered, where the implicit solvent outside of the spherical cavity of a radius R_c is replaced by the explicit solvent.

Although the buffer layer is originally introduced to remove the mathematical singularity at $r = a$, it has a more general use. The boundary between explicit and implicit solvents is modeled as a sharp transition from ε_i to ε_o. Clearly, this mathematical representation has its limitations, as the discrete jump in dielectric constant is unphysical. One of the limitations is the divergence in the reaction field at $r = a$, but there are others. To find out how the sharp boundary approximation affects electrostatics not only at the boundary but also in its proximity, we tested how forces computed by the image charges method for the model shown in Figs. 2(a) and 2(b) differ from those obtained in an all-atom simulation under PBC. A simulation for liquid water in a large simulation box was performed for this test using the moving-boundary reaction field method.[76] We find that the deviation computed as a function of the distance from the center of the sphere r (data not shown) is negligible for small r but becomes significant when r approaches the boundary a. As expected, the model with the buffer filled with solvent shown in Fig. 2(b) produces much smaller deviations than the model with the vacuum buffer, Fig. 2(a). Importantly, the

decay of the force deviation is gradual and takes over 4–6 Å to fall off appreciably. What this test shows is that the effect of the boundary is not limited to the boundary itself. Strong surface effects, manifested in different electrostatic potential generated by the spherical cavity model and that observed in the bulk molecular solvent, are detected at an appreciable depth into the explicit solvent. It is clear, therefore, that these surface effects are the intrinsic property of the employed model and will persist whenever a sharp transition between two dielectric media is present. It is possible to mitigate the impact of the surface in electrostatic forces by manipulating other interactions, such as boundary repulsive potentials for instance in the droplet models of solvent.[32,64–67] This approach, however, is system specific. It requires different sets of parameters for different sizes of the spherical cavity or different compositions of the solvent. This presents a significant obstacle toward the adoption of the model as a general-purpose simulation tool. In this paper we pursue a different approach. We adopt the buffer layer as a means to reduce the surface effects. As Fig. 2(b) shows, charges inside the radius R_c will not feel the presence of the surface if τ is large enough. We therefore treat τ as an adjustable parameter that needs to be determined phenomenologically. Importantly, this parameter is specific to the type of the explicit-implicit solvent interface and thus needs to be set only once, since the same interface is used in all solvation simulations.

2. Choice of boundary conditions

In the previous section we discussed the treatment of the dielectric boundary. In this section we focus on boundary conditions as they apply to other potentials, primarily dispersion forces between solvent molecules and their surroundings. There are several strategies to treat the explicit-implicit boundary. First, it can be fixed in space, creating a finite-size cavity filled with solvent, or a droplet. Repulsive potentials are applied at the wall of the droplet to prevent the escape of the solvent and solute molecules. The fixed-boundary approach has been implemented in the context of the Kirkwood expansion,[56,57] the Friedman expansion,[32,64–67] and where the reaction field is computed numerically.[82] The boundary potentials play an important role in this approach. In all the reported models, they have to be carefully adjusted so as the properties of the finite simulated system match those of the bulk solvent. However, such adjustments were not successful in all cases.[32,66] Notably, after extensive manipulations of the boundary potential, Wang and Hermans[66] concluded that the droplet model of water cannot reproduce accurately the bulk value of the dielectric constant, obtained independently in lattice-sum simulations. A similar conclusion regarding the dielectric response of a confined solvent is reached by Blaak and Hansen[32] for a model dipolar liquid. As we have already mentioned, the need to adjust parameters is a serious limitation of the droplet models, regardless of whether they lead to the bulk behavior or not.

The second strategy of applying the spherical cavity model is to use a separate sphere for each simulated charge, or a group of charges, resulting in the moving-boundary reaction field method.[76,78,80,83] This method takes advantage of the exact solution for the electric field in Eq. (3) and admits
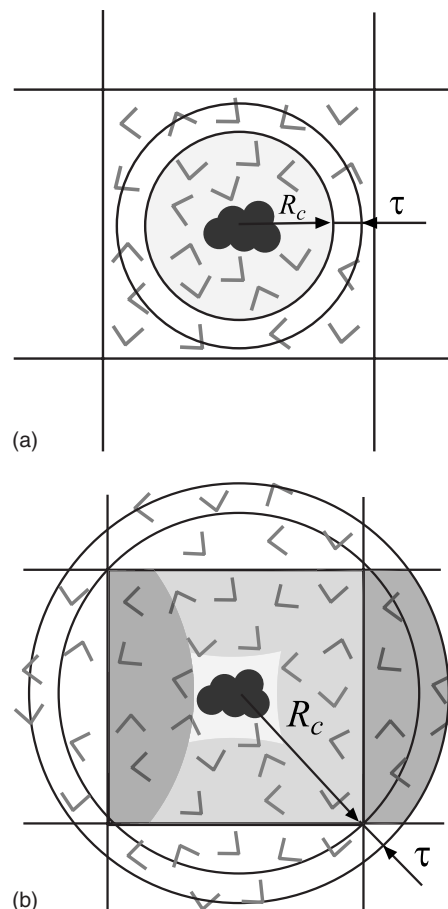

FIG. 3. Cartoon explaining how the spherical cavity can be integrated with the cubic simulation box under PBCs. In (a), the radius of the sphere is smaller than half the length of the simulation box. Charges within radius R_c experience reaction field computed by the method of image charges. Electrostatic interactions for all other particles in the simulation box are computed by a different method, for instance straight cutoff. The result is a coexistence of two different regions of water in one simulation box. Due to different pressures in these regions, neither will resemble bulk water. In (b), the radius R_c is increased to enclose the entire simulation box. The advantage is that all particles are treated by the same electrostatics method. Shaded areas highlight particles within $R_c + \tau$ radius that are not unique but result from periodic imaging from the central simulation box. In both schemes, a solute placed at the center of the sphere avoids electrostatic interactions with its periodic images.

the use of the PBC, which is the best technique currently available to eliminate the surface effects. Unfortunately, the moving boundary implies translational invariance of the simulated system, which consequently requires that the modeled medium be homogeneous. Systems that contain strong inhomogeneities, for example, solutes, are not accurately described by the moving-boundary reaction field method.

The third approach combines the PBC in the simulation cell with the fixed boundary in the spherical cavity.[58] This allows nonhomogeneous systems to be simulated with minimal surface effects and thus presents a significant advance over the other two methods. We pursue the periodic boundary strategy in the present work. We start with the model of Alper and Levy[58] and further improve it so that it can properly reproduce the bulk properties of liquid water. Figures 3(a) and 3(b) show two possible ways in which the spherical cavity can be integrated with the cubic simulation box. First,

the radius R_c can be chosen smaller than half the length of the box $L/2$. Electrostatic forces acting on particles within the sphere are computed using the image charges method (that employs the buffer layer τ), as shown in Fig. 3(a). All other particles need to be treated by some other method, for instance cutoff truncation.[58] No matter how close two electrostatics schemes are, they can never be identical. This results in the coexistence of two different regions of water within one simulation box, each described by a slightly different potential. The coexistence requires that the two regions have identical pressure. It is practically impossible, however, to ensure that different potentials produce the same pressure for a given particle density. Small differences will always remain. The result is that particles in the simulation box will begin to migrate from one region to the other in order to equilibrate the pressure. As a consequence, the density in both regions will deviate from the starting density. Eventually, neither region will properly represent the bulk solvent from which the simulation is started. We verified in simulations that this scenario actually takes place. We used the standard moving-boundary reaction field method to compute electrostatic forces for the particles outside of the spherical cavity. This method is in very good agreement for many structural and dynamic properties of liquid water with more accurate lattice-sum methods.[80] Nevertheless, we were unable to find a set of parameters in the context of the model shown in Fig. 3(a) that would reproduce the bulk properties of water satisfactorily. Our conclusion is that this model is not suitable for high-quality solvation studies.

To avoid the coexistence problem, it is necessary to treat all particles present in the system with one electrostatics method. This can be done by extending the spherical cavity model to include all particles in the simulation box as shown in Fig. 3(b). The radius R_c has to be increased to reach the furthest point in the box from the center of the sphere, which is a corner in the case of a cube. Due to the difference in the geometry between a sphere and a cube, this necessarily means that some particles within R_c radius will not be unique but rather periodically imaged from the nearest copies of the simulation box. The dark-shaded areas in Fig. 3(b) illustrate how two periodic images are created. The light-shaded areas display all imaged parts within the simulation box. The central part is not imaged and thus serves as the "productive" region of the simulation box. Its size is determined by the difference in the distance between the furthest and nearest points in the simulation box to the center. A solute molecule placed in this region will experience no electrostatic interactions with its periodic image copies. If the central region is large enough to also include the nearest solvation shell, the mathematical model shown in Fig. 3(b) will be an accurate representation of a solute embedded in an infinite bath of solvent. This is the major advantage of the electrostatics method proposed in this work over lattice-sum methods in which solute-solute interactions are always present through image copies.

3. Simulation box type

We used the example of a cubic box in the previous section for illustrative purposes. This type of box, however,

FIG. 4. A representation of the model proposed in this work to treat electrostatic interactions of a solute molecule embedded in solvent under PBCs. The TO simulation box Λ is placed inside a spherical cavity with radius R_c. The sphere, together with the buffer layer of depth τ, forms a larger cavity Γ. Charges in Γ generate reaction field that acts on the charges in the simulation box Λ. Shown as hatched area is region III which consists of charges with radius less than $R_c+\tau$ that are nearest periodic images of Λ. Particles in Λ that have images in Γ make up region II. The remaining, nonimaged particles in the simulation box form region I. The size of this region can accommodate solute molecules of a maximum diameter $d=L(\sqrt{3}-(\sqrt{5}/2))-2\tau$, indicated as a broken line.

is not among the best in terms of maximizing the size of the productive region. Box types that have a more spherical shape are desirable. In this work we focus on the truncated octrahedron geometry. The design of our model is explained in Fig. 4. The TO Λ is built from a cube of length L by cutting eight corners at a distance $(1/4\sqrt{2})L$ from the center of its sides. Figure 4 shows that the cross section of the TO in the XZ plane if the cubic box, from which it is created, is centered around the origin. The simulation box has 14 faces, including 8 hexagonal faces, 6 square faces, and 24 corners, all of which are equivalent. The distance from the origin to the square face is $(1/2)L$, to the hexagonal face $(\sqrt{3}/4)L$, and to the corners $R_c=(\sqrt{5}/4)L$. There are 14 nearest neighbors of the central simulation box, each resulting from a translation through an appropriate side of the TO. Particles in these neighboring boxes that are at a distance $R_c+\tau$ or less from the origin form a buffer layer around the central simulation box, which we refer to as region III. Together with the simulation box Λ, the buffer layer forms the local volume of spherical cavity Γ. All charges in Γ contribute to the reaction field, computed using the image charge method. The particles inside the simulation box that give rise to region III through translations are denoted as region II. The remaining particles in Λ, which are not periodically imaged, are labeled as region I. This region is the productive part of the simulation box in which a solute may be solvated without causing any artificial electrostatic solute-solute interactions. The size of this region d can be determined by comparing the points of furthest and nearest separation of particles in Λ from the
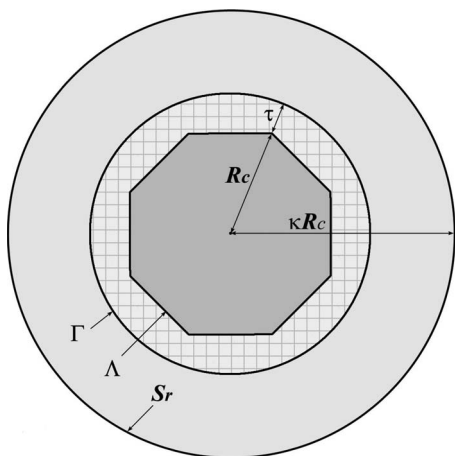
FIG. 5. A representation of the adapted-FMM method with a local expansion. A reference sphere S_r of radius κR_c is centered at the origin with $\kappa > 1$. The evaluation of the field within Λ due to the charges inside S_r is carried out by an adaptive FMM. For all periodic/image charges outside S_r, the field at a point inside Λ generated by these periodic/image charges can be calculated by a local expansion. See the main text for the details.

center of the box. For a TO shown in Fig. 4, $d = L(\sqrt{3} - \sqrt{5}/2) - 2\tau = 0.61L - 2\tau$. The same quantity for a cubic box is $d = L(2 - \sqrt{3}) - 2\tau = 0.27L - 2\tau$. Comparing these two quantities demonstrates the advantage of using the TO geometry. For $L = 45$ Å and $\tau = 5$ Å, for instance, the TO box allows simulations of solute molecules with diameter <17 Å. These same parameters in a cubic box result in $d = 2$ Å, which is meaningless from the molecular size point of view.

4. Fast-multipole method for image charges

To evaluate the electrostatic forces within the simulation box Λ, we use the FMM. In the simplest implementation, such evaluation can be carried out with a single FMM run by including into the FMM box all charges within the box Λ, plus their periodic images in region III and all image charges outside the cavity Γ, with all charges being taken as acting in a homogeneous medium of dielectric permittivity ε_i.[68,69] However, since the image charges outside the cavity Γ are highly nonuniformly distributed, especially for the image charges of those charges close to the center of the box Λ, such a direct application of the FMM is not efficient. Considering that only the force within the simulation box Λ needs to be evaluated, a simple but more efficient way would be to calculate the interaction between the charges in Λ and the periodic/image charges far away from Λ directly by the so-called local expansion.

More specifically, we introduce another reference sphere S_r of radius κR_c centered at the origin with $\kappa > 1$, as shown in Fig. 5. The evaluation of the field within the TO box Λ due to the charges inside this reference sphere is carried out by an adaptive FMM. For all periodic/image charges outside this reference sphere, the potential field at a point $\mathbf{r} = (r, \theta, \phi)$ inside the box Λ generated by these periodic/image charges can be described by a local expansion[70]

$$\Phi(\mathbf{r}) \approx \sum_{j=0}^{p} \sum_{k=-j}^{j} L_j^k Y_j^k(\theta, \phi) r^j, \tag{13}$$

where p is the local expansion order, $Y_j^k(\theta, \phi)$ are the spherical harmonics, and L_j^k are the local expansion coefficients given by

$$L_j^k = \sum_{l=1}^{M} \hat{q}_l \frac{Y_j^{-k}(\alpha_l, \beta_l)}{\rho_l^{j+1}}. \tag{14}$$

Here, $\hat{q}_l, l = 1, 2, \ldots, M$ are the periodic/image charges outside S_r with $(\rho_l, \alpha_l, \beta_l)$ being their locations. Consequently, the force $\mathbf{f}(\mathbf{r}) = (f_x(\mathbf{r}), f_y(\mathbf{r}), f_z(\mathbf{r}))$ exerted on a particle q at $\mathbf{r} = (r, \theta, \phi)$ inside Λ by these periodic/image charges can be calculated by[84]

$$f_x(\mathbf{r}) = -q \frac{\partial}{\partial x} \Phi(\mathbf{r}) = -q \, \mathrm{Re}(H_2 - H_3), \tag{15}$$

$$f_y(\mathbf{r}) = -q \frac{\partial}{\partial y} \Phi(\mathbf{r}) = -q \, \mathrm{Im}(H_2 + H_3), \tag{16}$$

$$f_z(\mathbf{r}) = -q \frac{\partial}{\partial z} \Phi(\mathbf{r}) = -q(H_0 + 2 \, \mathrm{Re}(H_1)), \tag{17}$$

where $\mathrm{Re}(\cdots)$ and $\mathrm{Im}(\cdots)$ represent the real part and the imaginary part of a complex number, respectively, and

$$H_0 = \sum_{j=1}^{p} j L_j^0 P_{j-1}(\cos \theta) r^{j-1},$$

$$H_1 = \sum_{j=1}^{p} \sum_{k=1}^{j-1} (j+k) C_j^k L_j^k e^{ik\phi} P_{j-1}^k(\cos \theta) r^{j-1},$$

$$H_2 = \sum_{j=1}^{p} L_j^0 e^{i\phi} P_{j-1}^1(\cos \theta) r^{j-1}$$
$$+ \sum_{j=1}^{p} \sum_{k=1}^{j} C_j^k L_j^k e^{i(k+1)\phi} P_{j-1}^{k+1}(\cos \theta) r^{j-1},$$

$$H_3 = \sum_{j=1}^{p} C_j^{-1} L_j^1 P_{j-1}^1(\cos \theta) r^{j-1}$$
$$+ \sum_{j=1}^{p} \sum_{k=2}^{j} B_j^k C_j^k L_j^k e^{i(k-1)\phi} P_{j-1}^{k-1}(\cos \theta) r^{j-1},$$

and

$$B_j^k = (j+k)(j+k-1), \quad C_j^k = \sqrt{\frac{(j-k)!}{(j+k)!}}.$$

III. COMPUTATIONAL DETAILS

A. Molecular dynamics simulations

In this work, we use TIP3P (Ref. 85) all-atom model to characterize properties of bulk water. All molecular dynamics simulations and trajectory analysis were performed by an

in-house software, which was specifically written for this project. For the FMM implementation, however, we adopt the free software KIFMM developed by Ying *et al.*[86] using a kernel-independent adaptive FMM. In all simulations, the maximum number of levels in the FMM computation tree and the maximum number of points allowed in a leaf box of the tree were set to 10 and 150, respectively. The parameter that controls the accuracy was set to 4 so an accuracy of better than 10^{-4} in the FMM force evaluation could be achieved. For the local expansion, a sphere of radius $2R_c$ was employed as the reference sphere (namely, $\kappa=2$), and the local expansion order p in Eq. (13) was set to 10. The covalent bond lengths of water molecules were constrained according to the noniterative matrix method.[87,88] The positions and velocities of particles were propagated according to an implementation of Jang and Voth[89] of the velocity Verlet algorithm, coupled with the Nosé–Hoover thermostat. The algorithm labeled VV1 in the original paper was employed. The dielectric constant of the external field was set to 80. The van der Waals interactions were modeled by the Lennard-Jones potential which was truncated at 10 Å. Geometrical centers of molecules were used to measure the distance from one molecule to another. The integration time step was set to 2 fs. The simulations were performed under constant temperature conditions at $T=300$ K and the thermostat's coupling constant was 0.05 ps. The trajectories were recorded at every 0.2 ps for subsequent analysis.

The initial coordinates were generated from an equilibrium 200 ps simulation in *NPT* ensemble, performed by the GROMACS program package.[90,91] A weak coupling to an external heat bath (coupling time constant $\tau_T=0.1$ ps, target temperature $T_0=300$ *K*) and a weak coupling barostat (time constant $\tau_p=0.5$ ps, isothermal compressibility $\beta=4.5 \times 10^{-5}$ atm^{-1}, target pressure $P_0=1$ atm) were applied. Covalent bonds of TIP3P water molecules were constrained using SHAKE procedure with a relative tolerance of 10^{-4}. Once equilibrated, the system was then cut to different TO boxes with desired sizes.

Starting from the equilibrated system, three system sizes containing 408, 1408, and 3425 water molecules in TO boxes were used. The box side lengths of the cubes containing TO boxes were 30, 45, and 60 Å, respectively. Four different buffer layer thicknesses 2, 4, 6, and 8 Å were used to test the buffer layer effects with each system size. To test the dependence on the number of the image charges, simulations with $N_i=1$, 2, and 3 were performed for each box size and thickness of buffer layer. Furthermore, we also evaluate the effect of having no reaction field correction. For convenience, we denote this case as $N_i=0$ in this work. In total, 48 different simulations were carried out in this study. Each of the 48 MD simulations was 2.1 ns long and started with the same initial coordinates and the same initial velocity assignments (i.e., the same seed was used for the random number generator). All trajectories were analyzed between 0.1 and 2.1 ns, leaving the first 0.1 ns as equilibration time allowing the system to adjust to the different parameters.

For reference, a PME simulation was carried out for 2.1 ns on a cubic box of edge length 60 Å and contained 6957 water molecules using the GROMACS program package.[90,91]

TABLE II. Accuracy comparison between image charge method and PME method.

PME	X component	Y component	Z component
	2.43×10^{-3}	3.94×10^{-3}	1.83×10^{-3}
Image	X component	Y component	Z component
$N_i=1$	1.34×10^{-3}	1.17×10^{-3}	1.08×10^{-3}
$N_i=2$	1.29×10^{-4}	8.26×10^{-5}	1.30×10^{-4}
$N_i=3$	1.41×10^{-5}	7.25×10^{-6}	6.94×10^{-6}

Additionally, similar PME simulations were performed for the three octahedral boxes considered by the image charge method.

B. Accuracy of force calculations

In molecular dynamics simulations, forces acting on particles determine the overall evolution of the simulated system. Here we compare the accuracy of electrostatic forces computed in the proposed image-charge method with those in the PME approach. Both methods are approximate and therefore need reference points for comparisons. In the reaction-field method, the accuracy is controlled by the number of images N_i. Starting with some number $N_i^\star$, forces cease to vary with N_i and thus can be treated as converged and used for reference. Our calculations on a single water configuration saved from the simulation with $L=60$ Å and $\tau=6$ Å showed that $N_i^\star \sim 10$.

The accuracy of the PME forces can be evaluated in comparison with the forces of the Ewald summation. The latter are comprised of two parts: direct summation and the summation in the Fourier space. Normally, the force calculation is performed by setting the single parameter controlling the relative contribution of the two parts such that the direct sum is almost exact while the number of wave vectors governing the accuracy of the inverse sum is increased until convergence. This is the algorithm we used in the present work. All calculations were done by the GROMACS package.[90,91] We set the parameter ewald-rtol determining the strength of the direct sum at the cutoff to the suggested value of 10^{-5} while the Fourier spacing parameter was decreased from 0.3 to 0.06 nm, where the convergence of forces was observed. These tests were repeated for Ewald-rtol=10^{-6} and produced identical results.

In our PME calculations, the parameters controlling electrostatic forces were set to their suggested values[92] ewald-rtol=10^{-5} and Fourier spacing=0.12 nm. The resulting accuracy of the forces is analyzed in Table II, which shows relative error averaged over all charges in the simulation box. The results differ slightly for different components of the forces, a consequence of considering one atomic configuration only, but agree on the order of magnitude of the error. All three components exhibit errors of the 10^{-3} magnitude. By comparison, the proposed image-based method is able to compute electrostatic forces with 10^{-3} error for $N_i=1$, 10^{-4} error for $N_i=2$, and 10^{-5} error for $N_i=3$. Based on these data, we conclude that the proposed method in one-image approximation, the modified Friedman model, is

equivalent in accuracy to the typical PME setup. Higher-order approximations, $N_i > 1$, exhibit much better accuracy than is typically used.

IV. RESULTS AND DISCUSSION

To validate our solvation model, we investigate the effect of varying its main geometrical parameters: the size of the simulation box, the thickness of the buffer layer, and the number of image charges in the reaction field calculations on various static and dynamic properties of the simulated water. We compare our results to those of the PME,[39] as a reference electrostatics method.

A. A buffer layer of at least 6 Å is required to yield uniform density

A sharp boundary between explicit and implicit solvents, which is the key element of our model, may induce surface effects, manifested in different properties of water throughout the simulation box. Such surface effects should be avoided in a successful electrostatics model, as the real bulk water is homogeneous. We examined the local particle density across the simulation box as a measure of the homogeneity of the simulated system. We computed the relative density of oxygen atoms along the diagonal in simulation boxes with $L = 30$, 45, and 60 Å, and $\tau = 2$, 4, 6, and 8 Å. Our data reveal strong dependence on the thickness of the buffer layer, as shown in Fig. 6 for $N_i = 2$. Other values considered of N_i yielded similar results.

It is seen that $\tau = 2$ Å produces a noticeably nonuniform density. For all L studied, the density at the edges of the simulation box is about 10% lower than the average density. To compensate for the shortage at the edges, a build-up of molecules is seen at the center of the box. All other τ lead to a uniform density distribution with some statistical variations. A summary of density deviations along the diagonal is presented in Table III.

The deviations decrease by an order of magnitude when τ changes from 2 to 4 Å. It is further reduced twofold for $\tau = 6$ Å and remains unchanged for $\tau = 8$ Å. Our results, therefore, indicate that the lowest thickness of the buffer layer $\tau = 2$ Å is not suitable for simulations of liquid water, for any size of the simulation box L. It is evident that a thin buffer layer is not capable of reducing the surface effects sufficiently. One possible explanation for this, as will become clear in the next section, is that two nearest molecules in bulk water are separated, on average, by 2.8 Å. It therefore may happen, that a buffer with $\tau = 2$ Å does not contain enough molecules to serve as a shield from the implicit solvent boundary. Increasing the size of the buffer to 4 Å reduces the surface effects significantly. The relative density for this τ is seen to fluctuate around 1 throughout the simulation box. Numerically, the convergence occurs at $\tau = 6$ Å as a thicker buffer of 8 Å does not lead to an improvement in the density. Our conclusion is that τ should be at least 6 Å to correctly represent homogeneous liquid water. The relative densities of $\tau = 6$ Å boxes are also in good agreement with PME simulations on similar TO boxes (data not shown). The
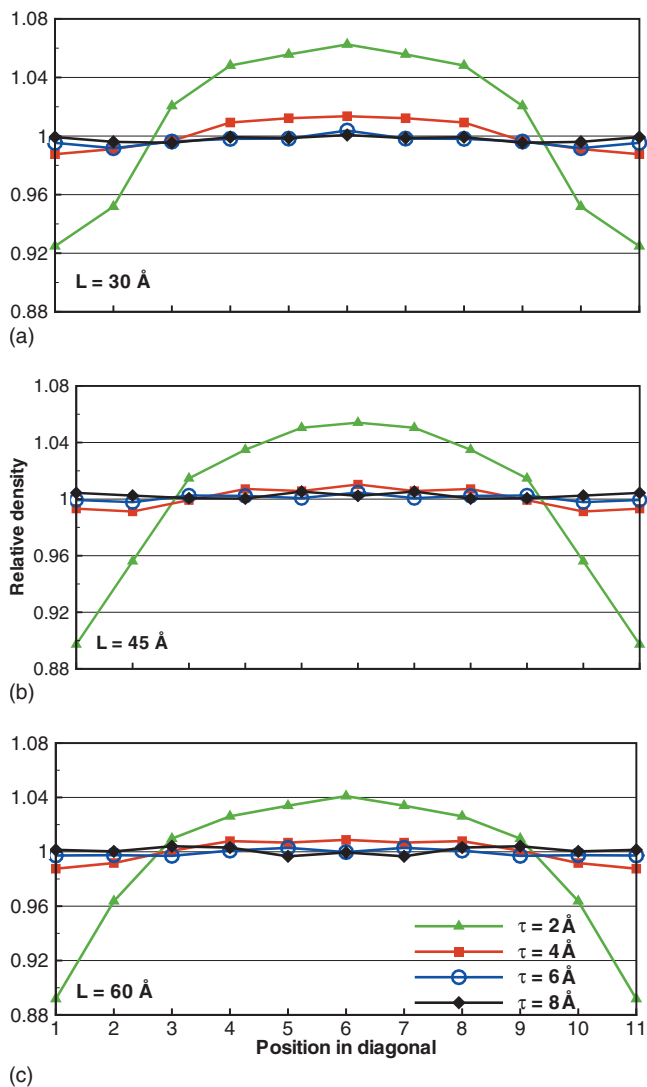
FIG. 6. Relative density along the diagonal of TO boxes with different τ and $N_i = 2$. The thinnest buffer layer $\tau = 2$ Å creates strongly nonuniform density across the simulation box.

size of the simulation box L or the number of image charges N_i, as long as $N_i > 0$, is unimportant for the density.

B. Structural properties are not strongly affected by the geometry parameters of the simulation box

Next, we investigate how L, τ, and N_i affect the structure of the simulated water. The structure is evaluated using $g_{oo}(r)$, the oxygen-oxygen radial distribution function (RDF) computed over the entire simulation box. First, we tested the dependence on the number of image charges that characterizes the accuracy of the reaction field. We find that RDFs converge for $N_i = 1$ for all simulation boxes and buffer layers

TABLE III. Comparison of the standard deviations of relative densities along the diagonal of TO boxes for $N_i = 2$ simulations.

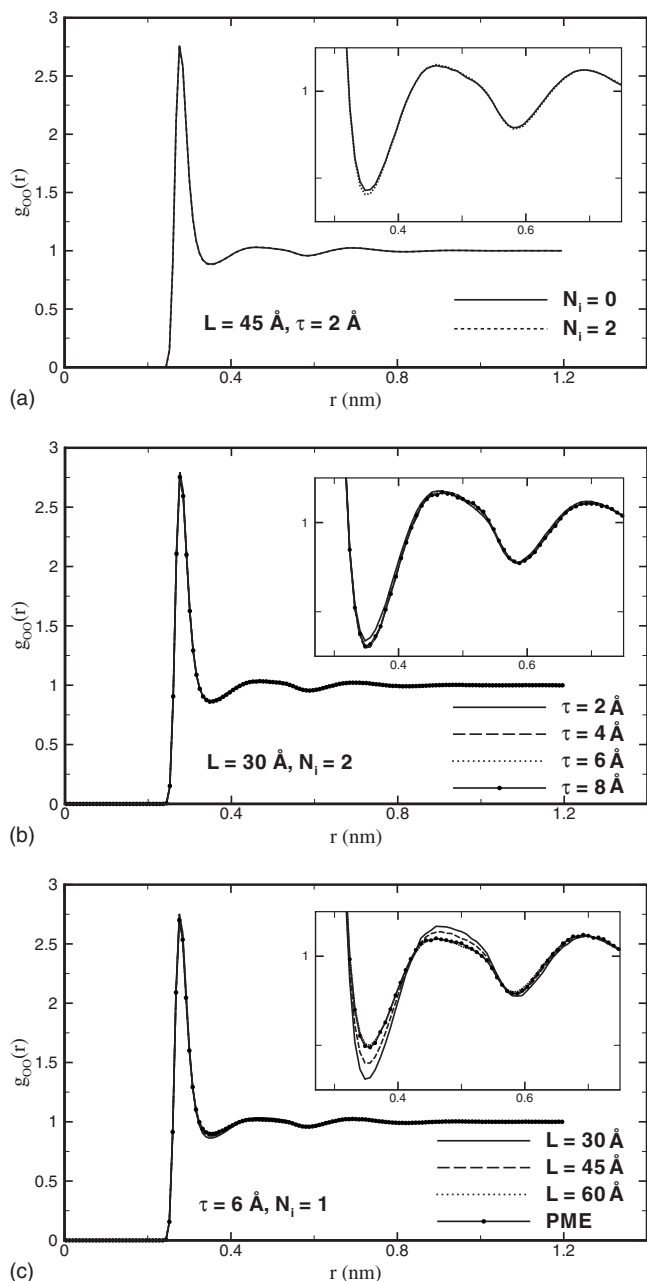|	$\tau=2$ Å	$\tau=4$ Å	$\tau=6$ Å	τ Å
$L=30$ Å	0.056	0.011	0.003	0.002
$L=45$ Å	0.060	0.007	0.002	0.002
$L=60$ Å	0.055	0.009	0.002	0.003

FIG. 7. Oxygen-oxygen RDF computed in this work. (a) Using the multiple-image method with, $N_i=2$, and without, $N_i=0$, reaction field corrections. The effect of the reaction field is small and affects only simulations with $\tau \leq 2$ Å. The inset shows the closeup of the first two minima. (b) Different values of τ at $L=30$ Å and $N_i=2$. Convergence occurs at $\tau=4$ Å. Larger L produces similar results. (c) Different L and $\tau=6$ Å, $N_i=1$. The variation in $g_{OO}(r)$ is small and mostly concentrated on the first few extrema. For increasing L, convergence occurs to the distribution function obtained in PME (Ref. 39) simulations.

studied. The only noticeable difference for $N_i=0$, when there is no reaction field, occurs when the buffer layer is $\tau=2$ Å. In this case, the RDF has a slightly higher first minimum, compared to the converged result with $N_i=2$, as shown in Fig. 7(a) for $L=45$ Å. The difference, however, is very small, leading us to conclude that the presence of the reaction field is not critical for the proper description of the static structure. That the structure of water is not very sensitive to the reaction field corrections has been noted previously.[80]

Second, we examined the dependence of the RDFs on

the thickness of the buffer layer. Increasing values of τ lead to convergence for all sizes of the simulation box considered. For $\tau \geq 4$ Å the changes in the RDFs become unnoticeable. The RDFs computed for $\tau=2$ Å differ slightly from the converged curve. The difference is greatest for the smallest box, $L=30$ Å, where the first minimum in the RDF is overestimated, as shown in Fig. 7(b). For larger simulation boxes, the difference diminishes, becoming hardly noticeable at $L=60$ Å. This is the direct consequence of water molecules located near the surface of the simulation box, which are strongly affected by the surface effects at $\tau=2$ Å, making progressively smaller contributions to the structural function averaged over the entire simulation volume. Data shown in Fig. 7(b) indicate that the buffer layer τ needs to be at least 4 Å in our model in order to reproduce the structure of liquid water correctly.

Finally, we analyzed the dependence on the size of the simulation box L. We find that the RDFs, shown in Fig. 7(c) for $L=30$, 45, and 60 Å, exhibit the largest variation with L among all parameters examined. The most important features of $g_{OO}(r)$ are the locations and magnitudes of the first three density peaks and the first two minima. Only the magnitudes of the first peak and minimum are slightly different for all cases. A close look at $g_{OO}(r)$ using a higher resolution, shown in the inset of Fig. 7(c), demonstrates that the $g_{OO}(r)$ overestimates the first and second maxima and underestimates the first minimum for small L. In all models, the first density peaks lie at 2.8 Å; the magnitude is 2.76 for $L=30$ Å, 2.73 for $L=45$ Å, and 2.70 for $L=60$ Å. For the first density minimum, the position is around 3.5 Å and the magnitude is 0.86 for $L=30$ Å, 0.88 for $L=45$ Å, and 0.90 for $L=60$ Å. Importantly, the RDFs computed for increasing values of L agree progressively better with the RDF obtained in the PME calculations. The latter has the first peak of height 2.70 at $r=2.76$ Å and the first minimum of depth 0.90 at $r=3.5$ Å, both of which do not exhibit variations with the size or type (tested cubic and octahedral types) of the simulation box. These are in excellent agreement with our data for $L=60$ Å, indicating that our method converges to the proper limit for large simulation boxes.

C. Reaction field is critical for the proper description of diffusion

In addition to the structural properties, we also assess how well our model of electrostatics describes dynamical properties of the simulated water. We choose to evaluate the self-diffusion coefficient D, which is a transport coefficient characterizing how quickly equilibrium is established in particle density following a small perturbation. Relating particle flux with the density gradient, D is an important dynamical characteristic of a liquid.[35] The effect of L, τ, and N_i on D was systematically studied in computer simulations. The results of these studies are summarized in Tables IV and V.

The dependence on N_i is not substantial except for $N_i=0$. For $L=45$ Å, the diffusion coefficient drops from 6.81×10^{-9} to 6.19×10^{-9} m^2 s^{-1} when N_i is increased from 0 to 1, and stays at that level for $N_i \geq 1$. A similar trend is seen for $L=60$ Å suggesting that (a) reaction field correction (N_i

TABLE IV. Diffusion coefficients with error estimates (unit of $D = 10^{-9}$ m^2 s^{-1}). Without reaction field correction, our method substantially overestimates the mobility of water molecules.

$N_i = 0$	$\tau = 2$ Å	$\tau = 4$ Å	$\tau = 6$ Å
$L = 45$ Å	6.69($\pm$0.07)	6.81($\pm$0.33)	7.04($\pm$0.27)
$L = 60$ Å	6.37($\pm$0.04)	6.39($\pm$0.10)	6.41($\pm$0.12)

> 0) is essential for the proper description of the diffusion process and (b) accuracy of the reaction field achieved under $N_i = 1$ approximation is adequate for the given system.

The thickness of the buffer layer τ has no noticeable effect on D except for the smallest simulation box $L = 30$ Å. For this system, D changes from 6.38×10^{-9} to 6.16×10^{-9} m^2 s^{-1} when τ is increased from 4 to 8 Å. In larger systems, $L = 45$ and 60 Å, the diffusion coefficient remains within a statistical error for varying τ. This applies to $N_i = 0$ case, where τ as small as 2 Å produces similar results to those at 4 and 6 Å for $L = 60$ Å. Clearly, the dynamics of water molecules are not affected by the buffer layer.

Like the structural function, the diffusion coefficient exhibits its strongest dependence among all geometrical parameters on L. Table V shows that D decreases from 6.38×10^{-9} m^2 s^{-1} at $L = 30$ Å and $\tau = 4$ Å to 6.02×10^{-9} m^2 s^{-1} at $L = 60$ Å and the same value of τ. The trend toward decreasing with L is independent of τ. A larger simulation box $L = 80$ Å produces D statistically indistinguishable from that of $L = 60$ Å. Comparison with the PME simulation, $D = 5.98 \times 10^{-9}$ m^2 s^{-1}, shows that (a) small simulation boxes overestimate the mobility of water molecules and (b) increasing L progressively improves the diffusion constant with the convergence seen at $L = 60$ Å. For this simulation box and larger, our electrostatics method produces excellent agreement with the PME. Simulations in octahedral boxes by PME produced D within statistical errors

TABLE V. Diffusion coefficients with error estimates (unit of $D = 10^{-9}$ m^2 s^{-1}). Noticeable dependence on model parameters is seen only for L.

| $N_i = 1$	$\tau = 4$ Å	$\tau = 6$ Å	$\tau = 8$ Å
$L = 30$ Å	6.40($\pm$0.26)	6.28($\pm$0.11)	6.16($\pm$0.12)
$L = 45$ Å	6.21($\pm$0.08)	6.20($\pm$0.10)	6.16($\pm$0.14)
$L = 60$ Å	6.02($\pm$0.06)	6.02($\pm$0.07)	6.02($\pm$0.04)
$L = 80$ Å	5.98($\pm$0.02)	5.98($\pm$0.02)	5.99($\pm$0.03)
$N_i = 2$	$\tau = 4$ Å	$\tau = 6$ Å	$\tau = 8$ Å
$L = 30$ Å	6.32($\pm$0.12)	6.33($\pm$0.24)	6.23($\pm$0.12)
$L = 45$ Å	6.16($\pm$0.09)	6.16($\pm$0.10)	6.15($\pm$0.08)
$L = 60$ Å	6.02($\pm$0.04)	6.01($\pm$0.05)	6.01($\pm$0.03)
$L = 80$ Å	5.96($\pm$0.02)	5.98($\pm$0.02)	5.98($\pm$0.02)
$N_i = 3$	$\tau = 4$ Å	$\tau = 6$ Å	$\tau = 8$ Å
$L = 30$ Å	6.34($\pm$0.17)	6.29($\pm$0.25)	6.24($\pm$0.15)
$L = 45$ Å	6.18($\pm$0.11)	6.19($\pm$0.10)	6.16($\pm$0.07)
$L = 60$ Å	6.01($\pm$0.03)	6.00($\pm$0.05)	6.03($\pm$0.07)
$L = 80$ Å	5.98($\pm$0.02)	6.00($\pm$0.04)	5.98($\pm$0.03)
PME	5.98($\pm$0.05)(TO box with $L = 60$ Å)		
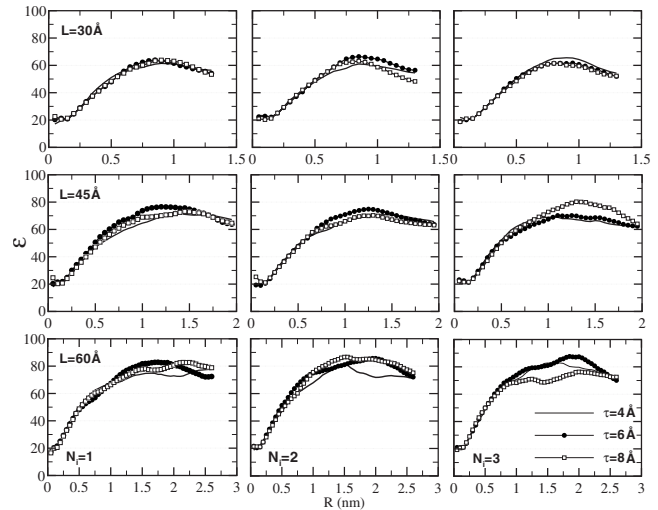
FIG. 8. Dielectric constant $\varepsilon(R)$ of spherical sample of radius R computed using Eq. (A3) for varying parameters L, τ, and N_i. No discernible dependence on τ is observed. For sufficiently large R, $\varepsilon(R)$ is seen to level off. The plateau value in $\varepsilon(R)$ is identified as the static dielectric constant of the material. This dielectric constant is seen to grow from $\sim$60 at $L = 30$ Å to $\sim$80 at $L = 60$ Å.

of each other. Our results show that reaction field correction and sufficiently large simulation box are required to properly describe self-diffusion processes.

D. Dielectric properties require large simulation boxes and reaction field corrections

Dielectric properties are evaluated by computing the dielectric constant ε. Equation (A3) (Appendix) is used, producing $\varepsilon(R)$ as a function of the sample radius R. As ε is a macroscopic property of a liquid material, it is physically meaningful only for relatively large objects. Once a crossover from a finite-size to macroscopic dimensions is achieved, the dielectric constant should be largely independent of the geometry of the investigated object. In the case of $\varepsilon(R)$, a plateau is expected starting at some radius $R^\star$. This is the behavior observed by Ballenegger and Hansen[93] for a droplet of dipolar liquid. As Fig. 8 shows, in which we plot $\varepsilon(R)$ for various parameters L, τ, and N_i, we observe the same behavior in the present work for water. For the smallest simulation box $L = 30$ Å, the $\varepsilon(R)$ curve has a bell-shape character. Apparently, this size of the simulation cell is close to where a crossover between microscopic and macroscopic behaviors occurs so that no clear plateau is developed. For larger $L = 45$ Å and especially 60 Å, the curves visibly flatten out, reaching regimes at $R > R^\star \simeq 1$ nm where $\varepsilon(R)$ is essentially constant. We identify the static dielectric constant of water under the proposed method of electrostatics with the plateau values of $\varepsilon(R)$. We estimate that ε is 60 for $L = 30$ Å, 70 for $L = 45$ Å, and 80 for $L = 60$ Å.

A clear trend is seen toward higher ε with increasing L. For comparison, the dielectric constant we estimated in the PME simulation is 91 ± 10. We find that this value is independent of the type of the simulation box, cubic or octahedral, or the size of the box; we tested all three sizes. Although slightly higher than the value obtained for $L = 60$ Å, our result is certainly within statistical error of the PME cal-
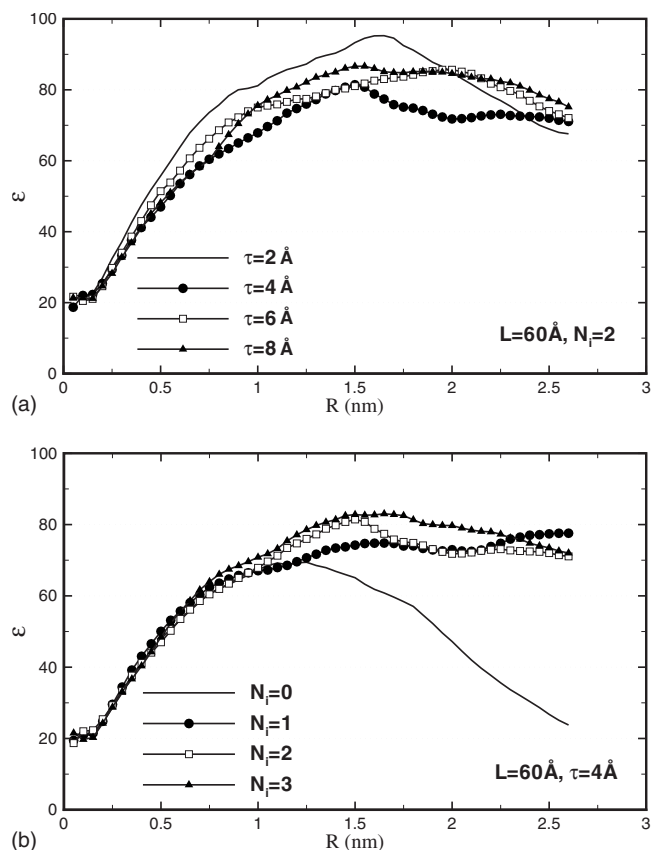
FIG. 9. Dielectric constant $\varepsilon(R)$. (a) For different values of τ. No significant dependence is seen even for the shortest $\tau=2$ Å. (b) For increasing number of image charges N_i. Convergence occurs at $N_i=1$. The reaction field ($N_i \geq 1$) is necessary to produce uniform dielectric properties across the simulation box.

culation. It is clear from these results that ε is a very sensitive function of the simulation box size L. In boxes that are too small, the dielectric response is significantly different from that in the bulk solvent, as has been pointed out previously.[66] This is important as ε describes the strength with which interactions between ions solvated in water are screened. The simulation box $L=30$ Å would lead to a 30% overestimation of such interactions. On the other hand, the $L=60$ Å box is acceptable from the dielectric response perspective.

The dependence of ε on the thickness of the buffer layer τ is weak. As Fig. 9(a) shows, in which $\varepsilon(R)$ is plotted for $L=60$ Å and $N_i=2$, the curve for even the smallest $\tau=2$ Å is not distinguishably different from others. This could be because the statistical noise masks the difference. Dielectric constant is computed by averaging the total dipole moment $\mathbf{M}^2$, which is a collective property and thus requires very long time to converge. Or, the $\tau=2$ Å data may be genuinely not too different from other τ. We showed in the previous sections that the smallest τ causes density perturbations that are substantial but affect predominantly the cell boundaries. In the dielectric constant calculations, the electric dipoles of the water molecules in the boundary layers are not included in the total dipole moment, but rather represent a dielectric medium. Instead of assessing directly the dielectric constant of this medium, it is recursively assigned based on ε. What

exactly happens at the simulation boundary, therefore, has a limited effect on the calculated $\varepsilon(R)$, as long as the behavior of the bulk molecules is unchanged. Which of these two scenarios takes place, excessive noise or weak dependence on τ, cannot be determined on the basis of the available data. For the purpose of estimating ε, all studied τ are equally suitable.

Finally, we examine how ε depends on the accuracy of the reaction field in the central cavity. Figure 9(b) shows $\varepsilon(R)$ computed for N_i varying from 0 to 3 for $L=60$ Å and $\tau=4$ Å. The convergence with the number of image charges occurs at $N_i=1$ as values greater than 1 do not produce appreciably different results. Most interesting is the behavior in the absence of the reaction field $N_i=0$. The curve that corresponds to this state is very different from those computed for $N_i>0$. At small R it grows rapidly with other curves but then instead of leveling off to a plateau, falls back to the small values at $R>R^\star$ reaching $\varepsilon=24$ at $R=2.6$ nm. This behavior leads to two conclusions. First, at $R<R^\star$, the spherical sample is separated by a large layer of water from the vacuum. This layer alone acts as a dielectric continuum, inducing reaction fields and defining ε in the sample. Second, as R grows, more water shells are assigned to the sample and less to the screening layer. The accompanied dramatic change in $\varepsilon(R)$ indicates that the dielectric response in the newly added parts of the sample is different from that in the center of the sample. In other words, the simulation box becomes inhomogeneous in its dielectric properties. The data shown in Fig. 9(b) therefore reveal that the reaction field is essential for maintaining a uniform dielectric response throughout the simulation box. This finding is not surprising as many previous studies emphasized the importance of reaction field corrections.[66]

V. CONCLUSIONS

We introduced in this paper a new method to treat electrostatic interactions in simulations of aqueous solutions. The method is based on a combination of explicit solvent representation for the central region of the system that includes a solute molecule and some solvent, and the continuum representation of distant parts of the solvent. In the resulting mathematical model, the solute molecule does not experience electrostatic interactions with its periodic images. This is a major advantage over the lattice-sum methods, which are considered the most accurate electrostatics treatments currently available. We validated our model using liquid water as a test system. As a result of an extensive investigation, we find that our method is in excellent agreement with the PME simulations for sufficiently large simulation boxes. This is the general conclusion of this work. Systematic studies of the effect of different geometrical parameters, including the simulation box size L, the buffer layer thickness τ, and the number of image charges N_i, lead to the following specific conclusions:

(A) We find that the explicit solvent part in a hybrid simulation model has to be separated from the continuum part by an interfacial buffer layer. The purpose of the layer is to mitigate adverse surface effects induced by

the sharp boundary between two dielectrics. We showed that buffers that are too thin, less than 6 Å, create strong perturbations in the local particle density, and, therefore, should be avoided. In addition to the density, the structure of water and its diffusion constant are adversely affected.

Although the minimal $\tau=6$ Å parameter was derived in this work for the $L=60$ Å simulation box, we expect it to apply to larger boxes as well. The parameter τ controls how much of the total reaction field inside the simulation box is modeled by the buffer layer Φ_B and how much is due to the charge-image particles Φ_i. The accuracy of the latter part is determined by the number of images N_i, with locations near the box boundary requiring larger values of N_i. With fixed N_i, simulations with larger L require larger τ to compute Φ_i with the same accuracy. The accuracy of the buffer reaction field part, on the other hand, is expected to saturate with τ. As our data on the pair-distribution functions indicate, liquid water loses its ordered structure starting at some limiting radius $r_c \sim 6$ Å, beyond which its effect on the particles inside the simulation box is vanishing. Importantly, we find in our calculations that Φ_B is at least 20 times greater than Φ_i near the box edges. We therefore believe that with increasing L and fixed τ, the increasing error in Φ_i will not significantly affect the total reaction field. Our simulations in the 80 Å box with a $\tau=6$ Å buffer layer showing converged values of the diffusion constant D support this conjecture.

(B) In addition to the buffer layer, an appreciable dependence on the size of the simulation is observed. Static, structural, and dynamic properties of water simulated in small boxes, $L<60$ Å, are seen to differ noticeably from their limiting behavior at large L. Especially, strongly affected is the dielectric constant ε, which changes by approximately 30% when the size of the simulation box is increased from 30 to 60 Å. The system-size dependence exposes a weakness in the electrostatics approaches dividing the modeled system, solvent together with solutes, into explicit and implicit parts with sharp boundaries between the two. Clearly, explicit solvent compartments need to be of a certain size to acquire macroscopic characteristics. Otherwise, they are not representative of the bulk behavior. This finding has implications for implicit solvent models, based on either PB equation or its approximations,[13,28] that are applied for solvation energy in biomolecular solutes of varying dimensions. It appears that such models need to incorporate the size dependence explicitly, as the dielectric properties of small solutes, reflected in the parametrization of force fields such as atomic radii used to construct explicit/implicit boundary, will likely be different from those of large solutes.

(C) Reaction field plays a critical role in our computational setup. Partially, it is provided by the buffer layer surrounding the explicit solvent region where the solute molecules are located. Another part comes from the implicit solvent beyond the buffer. The accuracy of this part is controlled by the image charges parameter N_i. We find that neglecting the image charges, $N_i=0$, has serious consequences for the simulated system. Among other things, the reaction field is responsible for the increase in the diffusion coefficient by close to 10%. More importantly, dielectric properties become nonuniform across the simulation box. As the accuracy is improved with $N_i>0$, we are able to conclude that our model achieves convergence at $N_i=1$ in many static and dynamic properties, which corresponds to the modified Friedman approximation. As has been pointed out before, the leading contribution in this approximation is of the order of $O(1/\varepsilon)$. For water whose ε is 80 under normal pressure and temperature, it evidently is sufficiently accurate. Water under other thermodynamic conditions, for instance in the supercritical states, where ε is much less than 80, may require a larger number N_i for an adequate description.

Our method is designed for inhomogeneous systems. It is free of periodic artifacts in electrostatic interactions and, therefore, can be applied in areas where PME is not suitable. These include tests on the effects of periodicity artifacts, simulations of highly charged solutes and pKa determinations of proteins. The latter two applications are most efficient for proteins that are nearly spherical in shape. For strongly nonspherical objects, computations of the reaction field in a different geometry, a cylinder or an ellipsoid for instance, could be better suited. Work is underway in our group to generalize the image-charge method to such geometries. To avoid drifting outside of the main simulation cell, proteins in our model will have to be restrained to lie close to the center of the cell, so that sufficient buffer layer to the implicit boundary is afforded. The restraints will have no consequences for the protein structure or thermodynamics. Another extension concerns generalizations to treat not only point charges but also higher electric moments such as dipoles, quadrupoles, and others. These are important in the context of polarizable force-fields which use higher moments in the computation of the electrostatic energy. We are exploring the feasibility of deriving a fast-converging series, similar to that of the image-charge, for computing the reaction field of a point dipole situated inside a spherical cavity. The classical work of Onsager[94] provides the solution for dipoles placed at the center of the sphere. We focus on the off-center placements in the context of our model.[72]

The proposed model is particularly well suited for computing pKa values in proteins. These are most frequently calculated by solving numerically the PB equation[95] for charged and neutral versions of ionizable residues, to obtain the reaction field. The advantage of our model is that the reaction field experienced by a solute molecule is directly accessible in the simulation. There is no need to perform any additional calculations. Unlike the PB approach, our method includes the effects of local conformational fluctuations and the discrete nature of water molecules. Additionally, there is no ambiguity about where to draw the boundary between the protein and the solvent.

The most time-consuming part in our model is the un-

TABLE VI. Comparison of computational efficiency between PME and the proposed image-charge method with varying parameters. The relative effort of generating 1 ns trajectories is shown for $\tau=6$ Å and $N_i=1$.

	$L=30$ Å	$L=45$ Å	$L=60$ Å	$L=80$ Å
Time (image method)/time (PME)	28	55	58	70

derlying FMM calculations. While the FMM has a better complexity $O(N)$ for N charges than the $O(N \log N)$ complexity of the PME, it is found in this study to underperform the latter. One reason is the numerical prefactors in the scaling laws, which make FMM computationally more efficient only for sufficiently large simulation boxes.[96] Additionally, the FMM has to handle considerably more particles than an equivalent PME calculation as N in our model includes both real charges and their images. Together, these two factors contribute to our model with the FMM version of Ref. [86] running slower than the equivalent PME simulations. The relative slowdown depends on the size of the simulated system, as shown in Table VI. It varies from 30 for the smaller simulation cells, $L=30$ Å, to around 70 for larger systems, $L=80$ Å. It is clear from Table VI that the proposed method is not competitive with PME in speed for homogeneous systems. However, continuing development of the FMM algorithm, especially efficient parallelization, will improve its performance in the future. Additionally, the speed advantage of PME may change in simulations of solvated molecules, which may require larger simulation boxes to achieve the same level of accuracy as the proposed model.

ACKNOWLEDGMENTS

APPENDIX: ANALYSIS OF THE TRAJECTORIES

We concentrated our analysis on a few observables known to be sensitive to the treatment of long-range electrostatics. First, to probe how uniform the simulated systems are, we measured relative particle density ρ_r as a function of the position in the simulation box. Due to symmetry, it suffices to look at the density along a line that connects two opposite vertexes of the TO box. We consider 11 equispaced positions along this line. To evaluate the local density, small spheres of radius $r=5$ Å are drawn around each position of the line. Using each position as origin, the number of water molecules is counted within each small sphere and converted to the particle density ρ_i, which is then converted to the relative one ρ_r by normalizing it with the number density of the whole box.

Second, the effect on the structure was evaluated by the pair-distribution functions. The oxygen-oxygen distribution function, $g_{OO}(r)$, was calculated as[35]
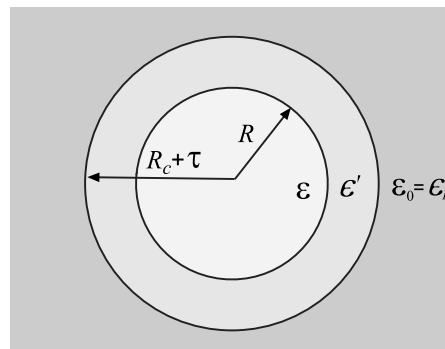

FIG. 10. An illustration for the formula used to compute the dielectric constant of water (Refs. [97]–[99] and [93]). A central sphere of radius R and permittivity ε is separated from the dielectric continuum ϵ_R by a spherical layer with permittivity ϵ'. The radius R is varied between 0 and $(\sqrt{3}/4)L$ so that no periodic images of water molecules are included. For sufficiently large R, $\varepsilon(R)$ reaches a plateau that corresponds to the bulk dielectric constant.

$$g_{OO}(r) = \frac{1}{4\pi\rho r^2 dr}\frac{1}{N}\left\langle \sum_{\alpha}^{N} \sum_{r<r_{\alpha\beta}(\tau)\leq r+dr, \beta\neq\alpha}^{N} 1 \right\rangle_{\tau}, \quad g_{OO}(0)$$
$$= 0, \tag{A1}$$

where N is the total number of molecules, ρ is the number density, dr is the window size (0.8 Å), $r_{\alpha\beta}(\tau)$ is the minimum image oxygen to oxygen distance between molecules α and β at time τ, and $\langle\cdots\rangle_{\tau}$ denotes averaging over all trajectory frames.

Third, dynamical properties were tested by examining self-diffusion of water molecules. Diffusion coefficients were calculated from the mean square displacement (MSD) of all oxygen atoms using the Einstein relation[35]

$$\lim_{t\to\infty}\langle|\mathbf{r}(t'+t)-\mathbf{r}(t')|^2\rangle = 6Dt, \tag{A2}$$

where $\mathbf{r}(t)$ is the position of the oxygen atom of the water molecule at time t, D is the diffusion coefficient, and the brackets denote averaging over all water molecules and time at origins t'. The diffusion coefficient was estimated from the slope of the linear part at long times of MSD versus time plot. The initial part of the line is influenced by inertial effects and is not included in this calculation.

Dielectric properties were evaluated by computing dielectric permittivity constant ε. Formulas for the dielectric constant of liquids encapsulated in spherical cavities were derived previously.[97–99,93] We consider a model devised by Berendsen,[97] in which a central sphere with permittivity ε is enclosed in a spherical layer with permittivity ϵ' and then embedded in dielectric continuum with permittivity ϵ_R, as shown in Fig. 10. The spherical layer generalizes the cavity model to cases where the dielectric constant at the explicit/implicit interface may differ from that on either side of the boundary. The original cavity model is recovered by setting ϵ' to either ε or ϵ_R. The dielectric constant in this model is given by

$$\varepsilon = \frac{1 + \frac{B}{A} 2\epsilon' \left[(2\epsilon_R + \epsilon') + \left(\frac{R}{R_c + \tau}\right)^3 (\epsilon_R - \epsilon') \right]}{1 - \frac{B}{A} \left[(2\epsilon_R + \epsilon') + 2\left(\frac{R}{R_c + \tau}\right)^3 (\epsilon_R - \epsilon') \right]},$$

$$A = (2\epsilon_R + \epsilon')(2\epsilon' + 1) - 2\left(\frac{R}{R_c + \tau}\right)^3 (\epsilon_R - \epsilon')(1 - \epsilon'),$$

$$(A3)$$

where $B = (1/\epsilon_0)(\langle \mathbf{M}^2 \rangle / 3 k_B T V(R))$ describes the fluctuations of the total dipole moment $\mathbf{M}(R)$ of a spherical sample with radius R. Here k_B is the Boltzmann constant, ϵ_0 is the vacuum permittivity, T is the temperature, and $V(R)$ is the volume of the sample. Equation (A3) is reduced to a simpler expression of Ballenegger and Hansen[93] for the dielectric cavity:

$$\varepsilon = \frac{1 + B \frac{2\epsilon_R}{2\epsilon_R + 1}}{1 - B \frac{1}{2\epsilon_R + 1}},$$

$$(A4)$$

if ϵ' is set to ϵ_R. This latter expression can be further reduced to the known Kirkwood–Frölich and the Clausius–Mossotti formulas if ϵ_R is set to ε and 1, respectively. Expression (A3) is derived based on continuum electrostatics[93] and therefore applies at relatively large R. The radius of the sample also has an upper bound, $R_c + \tau$, defined by the geometry of our model. The maximum allowed R cannot be used in our calculations, however, because the sphere then contains periodic images of water molecules. As a linear response approximation, the dielectric constant relies on the quadratic fluctuations of the total dipole moment $\mathbf{M}(R)$, which as Eq. (A3) shows, scales linearly with the sample volume. If periodic images are present in a sample, the linear scaling of $\langle \mathbf{M}^2(R) \rangle$ is violated, thus invalidating the fluctuation formula. We therefore limit the radii in our calculations to $R \leq R_{\max} = (\sqrt{3}/4)L$, ensuring that only one copy of each water molecule is considered. The remaining layer of water $R_{\max} < R < R_c + \tau$ acts as a part of the dielectric continuum. Its dielectric permittivity $\epsilon' = \varepsilon$ is set self-consistently in our calculations using Eq. (A3). We find that only a few iterations are needed between ε and ϵ' to achieve convergence.

We perform two types of calculations in our tests. First, the number of image charges is $N_i > 0$, meaning that the reaction field is turned on. Equation (A3) is then used to compute ε with $\epsilon_R = 80$, which is the value set in the simulations. Second, we evaluate the effect of having no reaction field correction, $N_i = 0$. In this case, our model describes a water droplet in vacuum. The dielectric constant of the medium ϵ_R is accordingly set to 1.

[1] C. Sagui and T. A. Darden, Annu. Rev. Biophys. Biomol. Struct. **28**, 155 (1999).
[2] M. E. Davis and J. A. McCammon, Chem. Rev. (Washington, D.C.) **90**, 509 (1990).
[3] A. Warshel and S. T. Russell, Q. Rev. Biophys. **17**, 283 (1984).
[4] P. Koehl, Curr. Opin. Struct. Biol. **16**, 142 (2006).
[5] W. F. van Gunsteren and J. J. C. Berendsen, Angew. Chem., Int. Ed. Engl. **29**, 992 (1990).
[6] T. Simonson, Rep. Prog. Phys. **66**, 737 (2003).
[7] W. F. van Gunsteren, F. J. Luque, D. Timms, and A. E. Torda, Annu. Rev. Biophys. Biomol. Struct. **23**, 847 (1994).
[8] C. J. Cramer and D. G. Truhlar, Chem. Rev. (Washington, D.C.) **99**, 2161 (1999).
[9] C. Moser, J. Keske, K. Warncke, R. Farid, and P. Dutton, Nature (London) **355**, 796 (1992).
[10] H. Nakamura, Q. Rev. Biophys. **29**, 1 (1996).
[11] R. Zhou, Proteins **53**, 148 (2003).
[12] A. Warshel, P. K. Sharma, M. Kato, Y. Xiang, H. Liu, and M. H. M. Olsson, Chem. Rev. (Washington, D.C.) **106**, 3210 (2006).
[13] B. Roux, *Implicit Solvent Models, Computational Biochemistry and Biophysics* (Marcel Dekker, New York, 2001).
[14] B. Honig and A. Nicholls, Science **268**, 1144 (1995).
[15] N. A. Baker, Curr. Opin. Struct. Biol. **15**, 137 (2005).
[16] J. Wang, C. H. Tan, Y. H. Tan, Q. Lu, and R. Luo, Comm. Comp. Phys **3**, 1010 (2008).
[17] J. Warwicker and H. C. Watson, J. Mol. Biol. **157**, 671 (1982).
[18] M. Holst and F. Saied, J. Comput. Chem. **14**, 105 (1993).
[19] R. Luo, L. David, and M. K. Gilson, J. Comput. Chem. **23**, 1244 (2002).
[20] W. H. Orttung, Ann. N.Y. Acad. Sci. **303**, 22 (1977).
[21] M. J. Holst, N. A. Baker, and F. Wang, J. Comput. Chem. **21**, 1319 (2000).
[22] L. Chen, M. J. Holst, and J. C. Xu, SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **45**, 2298 (2007).
[23] D. Xie and S. Zhou, BIT Numerical Mathematics **47**, 853 (2007).
[24] D. G. Levitt, Biophys. J. **22**, 209 (1978).
[25] Y. N. Vorobjev and H. A. Scheraga, J. Comput. Chem. **18**, 569 (1997).
[26] A. H. Boschitsch, M. O. Fenley, and H. X. Zhou, J. Phys. Chem. B **106**, 2741 (2002).
[27] B. Z. Lu, X. L. Cheng, J. F. Huang, and J. A. McCammon, Proc. Natl. Acad. Sci. U.S.A. **103**, 19314 (2006).
[28] M. Feig and C. L. Brooks III, Curr. Opin. Struct. Biol. **14**, 217 (2004).
[29] Q. Lu and R. Luo, J. Chem. Phys. **119**, 11035 (2003).
[30] W. Im, D. Beglov, and B. Roux, Comput. Phys. Commun. **111**, 59 (1998).
[31] M. Feig, A. Onufriev, M. S. Lee, W. Im, D. A. Case, and C. L. Brooks III, J. Comput. Chem. **25**, 265 (2004).
[32] R. Blaak and J. P. Hansen, J. Chem. Phys. **124**, 144714 (2006).
[33] J. Dai, I. Tsukerman, A. Rubinstein, and S. Sherman, IEEE Trans. Magn. **43**, 1217 (2007).
[34] Z. Yu, M. P. Jacobson, C. S. Rapp, and R. A. Friesner, J. Phys. Chem. B **108**, 6643 (2004).
[35] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, Oxford, 1987).
[36] P. Ewald, Ann. Phys. **369**, 253 (1921).
[37] R. Abagyan and M. Totrov, J. Mol. Biol. **235**, 983 (1994).
[38] T. A. Darden, D. M. York, and L. G. Pedersen, J. Chem. Phys. **98**, 10089 (1993).
[39] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. Pedersen, J. Chem. Phys. **103**, 8577 (1995).
[40] J. Shimada, H. Kaneko, and T. Takada, J. Comput. Chem. **14**, 867 (1993).
[41] B. A. Luty, M. E. Davis, I. G. Tironi, and W. F. van Gunsteren, Mol. Simul. **14**, 11 (1994).
[42] L. Greengard and V. Rokhlin, J. Comput. Phys. **73**, 325 (1987).
[43] L. Greengard, *The Rapid Evaluation of Potential Fields in Particle Systems* (MIT Press, Cambridge, 1988).
[44] W. Weber, P. H. Hunenberger, and J. A. McCammon, J. Phys. Chem. B **104**, 3668 (2000).
[45] M. A. Villarreal and G. G. Montich, J. Biomol. Struct. Dyn. **23**, 135 (2005).
[46] M. A. Kastenholz and P. H. Hünenberger, J. Phys. Chem. B **108**, 774 (2004).
[47] S. Balasubramanian, C. J. Mundy, and M. L. Klein, J. Chem. Phys. **105**, 11190 (1996).
[48] M. Belhadj, H. E. Alper, and R. M. Levy, Chem. Phys. Lett. **179**, 13 (1991).
[49] P. H. Hunenberger and J. A. McCammon, Biophys. Chem. **78**, 69 (1999).
[50] A. Okur and C. Simmerling, Annu. Rep. Comp. Chem **2**, 97 (2006).
[51] M. S. Lee, Jr., F. R. Salsbury, and M. A. Olson, J. Comput. Chem. **25**, 1967 (2004).
[52] M. S. Lee and M. A. Olson, J. Phys. Chem. B **109**, 5223 (2005).
[53] W. Im, S. Bernèche, and B. Roux, J. Chem. Phys. **114**, 2924 (2001).

[54] D. Bashford and D. A. Case, Annu. Rev. Phys. Chem. **51**, 129 (2000).

[55] P. K. Yang, S. H. Liaw, and C. Lim, J. Phys. Chem. B **106**, 2973 (2002).

[56] G. King and A. Warshel, J. Chem. Phys. **91**, 3647 (1989).

[57] D. Beglov and B. Roux, J. Chem. Phys. **100**, 9050 (1994).

[58] H. Alper and R. Levy, J. Chem. Phys. **99**, 9847 (1993).

[59] J. G. Kirkwood, J. Chem. Phys. **2**, 351 (1934).

[60] J. G. Kirkwood, Chem. Rev. (Washington, D.C.) **19**, 275 (1936).

[61] J. C. Maxwell, *A Treatise on Electricity and Magnetism* (Dover, New York, 1954).

[62] H. L. Friedman, Mol. Phys. **29**, 1533 (1975).

[63] J. J. Havranek and P. B. Harbury, Proc. Natl. Acad. Sci. U.S.A. **96**, 11145 (1999).

[64] J. A. C. Rullmann and P. T. V. Duijnen, Mol. Phys. **61**, 293 (1987).

[65] A. Wallqvist, Mol. Simul. **10**, 13 (1993).

[66] L. Wang and J. Hermans, J. Phys. Chem. **99**, 12001 (1995).

[67] G. Petraglio, M. Ceccarelli, and M. Parrinello, J. Chem. Phys. **123**, 044103 (2005).

[68] W. Cai, S. Deng, and D. Jacobs, J. Comput. Phys. **223**, 846 (2007).

[69] S. Deng and W. Cai, Commun. Comput. Phys. **2**, 1007 (2007).

[70] S. Deng and W. Cai, J. Comput. Phys. **227**, 1246 (2007).

[71] A. V. Finkelstein, Mol. Biol. **11**, 811 (1977).

[72] I. V. Lindell, Radio Sci. **27**, 1 (1992).

[73] C. Neumann, *Hydrodynamische untersuchen nebst einem anhang uber die probleme der elektrostatik und der magnetischen induktion* (Teubner, Leipzig, 1883), Vol. 279.

[74] W. T. Norris, IEE Proc.: Sci., Meas. Technol. **142**, 142 (1995).

[75] Z. Xu, S. Deng, and W. Cai, J. Comput. Phys. **228**, 2092 (2009).

[76] J. A. Barker and R. O. Watts, Mol. Phys. **26**, 789 (1973).

[77] G. Mathias, B. Egwolf, M. Nonella, and P. Tavan, J. Chem. Phys. **118**, 10847 (2003).

[78] I. G. Tironi, R. Sperb, P. E. Smith, and W. F. van Gunsteren, J. Chem. Phys. **102**, 5451 (1995).

[79] X. W. Wu and B. R. Brooks, J. Chem. Phys. **122**, 044107 (2005).

[80] P. H. Hunenberger and W. F. van Gunsteren, J. Chem. Phys. **108**, 6117 (1998).

[81] D. van der Spoel, P. J. van Maaren, and H. J. C. Berendsen, J. Chem. Phys. **108**, 10220 (1998).

[82] G. Brancato, A. D. Nola, V. Barone, and A. Amadei, J. Chem. Phys. **122**, 154109 (2005).

[83] J. A. Barker, Mol. Phys. **83**, 1057 (1994).

[84] M. Xiang, S. Deng, and W. Cai, "A sixth-order image approximation to the ionic solvent induced reaction field," J. Sci. Comput. (in press).

[85] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).

[86] L. Ying, G. Biros, and D. Zorin, J. Comput. Phys. **196**, 591 (2004).

[87] M. Yoneya, H. J. C. Berendsen, and K. Hirasawa, Mol. Simul. **13**, 395 (1994).

[88] J. T. Slusher and P. T. Cummings, Mol. Simul. **18**, 213 (1996).

[89] S. Jang and G. A. Voth, Biopolymers **107**, 9514 (1997).

[90] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen, Comput. Phys. Commun. **91**, 43 (1995).

[91] E. Lindahl, B. Hess, and D. van der Spoel, J. Mol. Model. **7**, 306 (2001).

[92] D. van der Spoel, E. Lindahl, B. Hess, A. R. van Buuren, E. Apol, P. J. Meulenhoff, D. P. Tieleman, A. L. T. M. Sijbers, K. A. Feenstra, R. van Drunen, and H. J. C. Berendsen (unpublished).

[93] V. Ballenegger and J. P. Hansen, J. Chem. Phys. **122**, 114711 (2005).

[94] L. Onsager, J. Am. Chem. Soc. **58**, 1486 (1936).

[95] A. Warshel, P. K. Sharma, M. Kato, and W. W. Parson, Biochim. Biophys. Acta. **1764**, 1647 (2006).

[96] D. Frenkel and B. Smit, *Understanding Molecular Simulation* (Academic, London, 2002).

[97] H. J. C. Berendsen, "Molecular dynamics and Monte Carlo calculations on water," CECAM Report No. 29, 1972.

[98] D. J. Adams and I. R. McDonald, Mol. Phys. **32**, 931 (1976).

[99] J. G. Powles, R. F. Fowler, and W. A. B. Evans, Chem. Phys. Lett. **107**, 280 (1984).