JCTC Journal of Chemical Theory and Computation

Article

Cite This: J. Chem. Theory Comput. 2019, 15, 6213−6224

pubs.acs.org/JCTC

Molecular Dynamics Using Nonvariational Polarizable Force Fields: Theory, Periodic Boundary Conditions Implementation, and Application to the Bond Capacity Model

Pier Paolo Poier,[†] Louis Lagardère,[‡,§] Jean-Philip Piquemal,[∥,⊥,#] and Frank Jensen*[,†]

[†]Department of Chemistry, Aarhus University, Langelandsgade 140, DK-8000 Aarhus, Denmark

[‡]Sorbonne Université, Institut Parisien de Chimie Physique et Théorique, 75005, Paris, France

[§]Sorbonne Université, Institut des Sciences du Calcul et des Données, 75005, Paris, France

[∥]Sorbonne Université, Laboratoire de Chimie Théorique, 75005, Paris, France

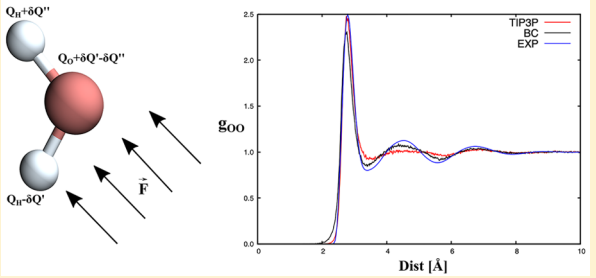
[⊥]Sorbonne Université, Institut Universitaire de France, 75005, Paris, France

[#]University of Texas, Department of Biomedical Engineering, Austin, Texas, United States

Supporting Information

ABSTRACT: We extend the framework for polarizable force fields to include the case where the electrostatic multipoles are not determined by a variational minimization of the electrostatic energy. Such models formally require that the polarization response is calculated for all possible geometrical perturbations in order to obtain the energy gradient required for performing molecular dynamics simulations. By making use of a Lagrange formalism, however, this computationally demanding task can be replaced by solving a single equation similar to that for determining the electrostatic variables themselves. Using the recently proposed bond capacity model that describes molecular polarization at the charge-only level, we show that the energy gradient for nonvariational energy models with periodic boundary conditions can be calculated with a computational effort similar to that for variational polarization models. The possibility of separating the equation for calculating the electrostatic variables from the energy expression depending on these variables without a large computational penalty provides flexibility in the design of new force fields.

$$\mathscr{L}(\mathbf{R},\mathbf{Q},\boldsymbol{\lambda}) = \frac{1}{2}\mathbf{Q}^{\dagger}\mathbf{J}\mathbf{Q} + \boldsymbol{\lambda}^{\dagger}\left[(\mathbf{I}+\mathbf{C}\tilde{\mathbf{J}})\mathbf{Q} + \mathbf{C}\chi - \mathbf{Q}_0\right]$$



INTRODUCTION

Computer simulations of biomolecular systems are a valuable tool for understanding the biological mechanism for many diseases and aiding in the search for new potential drug molecules.[1,2] Given the size of these systems, only simulations based on treating the nuclear motions by classical mechanics is feasible. Exploring the dynamics of liquids and proteins by solving Newton's second equation by a sequence of small finite time steps was pioneered by Rahman, Verlet, and Karplus and Weaver,[3−5] respectively, and molecular dynamics (MD) simulations have become routine methods implemented in several high-performing computer programs. The maximum time step is given by the fastest inherent motion of the particles, being on the order of 1 fs for atomistic resolution. Simulations corresponding to following the atomic motions for nanoseconds thus requires millions of force calculations, and this necessitates that each force calculation can be performed in a few seconds of real time in order to be feasible in a time frame of a few months. This in turn demands that the underlying energy function describing how the molecular energy depends on the atomic positions is computationally very efficient. Ideally, the energy function should be able to account for the electronic structure of the system, but a quantum description of the electronic degrees of freedom is prohibitive from a computational point of view. The vast majority of the simulations of biomolecular systems thus employ parametrized energy functions, usually denoted as force fields, with AMBER,[6] CHARMM,[7] OPLS,[8] and GROMOS,[9] being widely employed examples.

A force field is a parametrized energy function designed to capture the important parts of the inherent quantum nature of the electrons that provide molecular bonding and nonbonding interactions. The design has traditionally been by proposing a mathematical functional form and subsequently fitting the associated parameters to results from either experiments or higher level calculations. The short-range quantum effects resulting in bond formation are parametrized in terms of bond stretching and angle bending functions, typically chosen as low-order Taylor expansions. The long-range nonbonding

Received: July 19, 2019
Published: September 26, 2019

DOI: 10.1021/acs.jctc.9b00721
J. Chem. Theory Comput. 2019, 15, 6213−6224

interaction is parametrized as the sum of a nonpolar dispersion and a polar electrostatic contribution. When the dispersion is combined with a repulsive term arising from the electron−electron exchange interaction, this results in a van der Waals type mathematical function, of which the Lennard-Jones is a popular choice. The molecular conformational degrees of freedom are partly described by the nonbonded interactions, but an explicit Fourier type torsional energy function is usually required to obtain satisfactory results. Construction of a force field is a major undertaking, often proceeding by an initial parametrization for a limited class of molecules and subsequently extended to cover a wider selection of systems. All of the commonly employed force fields have been developed and refined over several decades and thus exist in several different versions. With the continuing increase in computational resources and the use of specialized hardware,[10] the simulation of realistic sized systems in the micro- or millisecond time frame is becoming feasible.[11] Although the results are still scattered, it appears that the accuracy of the underlying force field in some cases is the limiting factor.[12−15] There is thus an increasing interest in designing new force fields that can provide higher accuracy, and ideally it would be desirable to have a systematic sequence of force fields capable of providing increasingly accurate results.

The main limitation of production type force fields is the use of fixed partial atomic charges for describing the electrostatic interaction, which neglects the effect of the electron distribution being perturbed by changes in the electric field due to changes in the molecular geometry. This polarization effect has been modeled in a force field environment by three main approaches: the fluctuating charge (FQ) model,[16] where the atomic charges can respond to changes in the electrostatic potential, the induced dipole model,[17] where an atomic polarizability tensor generates an atomic dipole moment in response to the electric field, and the Drude oscillator model,[18] where the induced atomic dipole moment is generated by an additional charged virtual particle connected to the nucleus by a spring. The induced dipole and Drude oscillator models are closely related, the main difference being that the latter can be implemented using only charge−charge interactions, and a method for transferring parameter between the two models has been published.[19] The Drude oscillator can thus be considered as a variation of the induced dipole model with the same underlying approximations. Versions of the AMBER[20] and CHARMM[21] force fields incorporating polarization models have been published, where a polarization model is added to existing nonpolarizable versions, while more modern force fields such as AMOEBA[22] and SIBFA[23] have been designed with polarization included from the onset. We note in passing that all existing force fields contain a very large number of parameters that have been carefully tuned over time, and inadequacies in the mathematical functions may in the process have been compensated for by parameter redundancies.[24] Designing new force fields must, in our opinion, address the parameter redundancy issue, for example by assigning parameters primarily from electronic structure calculations, rather than from fitting to experimental results.

Assuming that polarization can be described by a monotomic convergent multipolar expansion, one would anticipate charge polarization to be more important than dipole polarization, and Mei et al. have shown that this indeed is the case for typical organic molecules with both localized and delocalized bonding.[25] It is thus of interest to investigate the limitations of a charge-only polarization model before including higher-order (dipole, quadrupole) polarization terms. The FQ model,[16] or its alternative formulation in terms of split-charge parameters,[26,27] is the dominant charge-polarization model, but it displays nonphysical phenomena such as charge-transfer over infinite distance[28] and nonlinear scaling of the polarizability with system size.[29,30] Charge-flow terms also arise when molecular polarizabilities are decomposed into atomic multipole contributions,[31−33] and these are closely related to charge response kernel methods.[34−36] The latter has been used in simulations using the energy expression discussed in the Supporting Information.[37,38]

We have in recent work proposed a charge-only polarization model that does not have the FQ artifacts and that is capable of reproducing molecular anisotropic polarization to within ∼10% using only a limited set of parameters derived directly from electronic structure calculations.[39] This Bond Capacity (BC) model has the potential to be a key building block for the design of new force fields with improved accuracy. However, it differs from all other existing polarization models by not calculating the charge variables by a variational minimization of the electrostatic energy expression, and this leads to computational challenges that we address in the current work. We show how the energy gradient required for performing MD simulations can be calculated with little additional work compared to energy minimization models and present results from an implementation using periodic boundary conditions. The methodology is completely general and thus opens for the design of polarization models using nonvariational energy expressions. We will in the following use the terms variational and nonvariational in the same definition as in electronic structure theory, where variational implies that the model variables are defined by an equation obtained by making the corresponding energy stationary with respect to the variables. The Hartree−Fock model is thus variational, since the molecular orbital coefficients are obtained from equations derived by minimizing the corresponding energy, but the coupled cluster model is nonvariational, as the equations defining the cluster amplitudes are not derived by minimizing the energy but rather from a projection against a reference state. The applicability of these models for modeling purposes similarly requires the capability of calculating energy gradients in a computationally efficient fashion.

THEORY

Calculating the molecular gradient for the BC model is a special case of the more general problem of calculating the gradient for nonvariational energy functions. We will in the first part of this section present a general framework for calculating energy gradients of variational and nonvariational coupled multipoles. The classical analogue of the Hellmann−Feynmann theorem[40] leads to a simplification for the former case, while a Lagrange technique can be used to obtain a similar simplification for the latter case. The second part of this section derives an explicit expression for the energy gradient of the BC model, and the last part discusses the implementation with Periodic Boundary Conditions in the Tinker-HP package.[41]

Lagrangian formulation of the Electrostatic Energy for Nonvariational Multipoles. We define an N sets of field dependent (induced) point multipoles $\mathbf{M}_1,...,\mathbf{M}_i,...,\mathbf{M}_N$ centered at position $\mathbf{R}_1,...,\mathbf{R}_i,...,\mathbf{R}_N$, each containing M multipoles where Q represents the charge, $\boldsymbol{\mu}$ the dipole moment, etc. $\mathbf{M}$

DOI: 10.1021/acs.jctc.9b00721
J. Chem. Theory Comput. 2019, 15, 6213−6224

represents the one-dimensional array of dimension $(M{\cdot}N)$ gathering all the multipoles in the system, eq 1.

$$\mathbf{M}^{\dagger} = [Q_1, \mu_1^x, \mu_1^y, \mu_1^z, ..., Q_i, \mu_i^x, \mu_i^y, \mu_i^z, ..., Q_N, \mu_N^x, \mu_N^y, \mu_N^z, ...] \quad (1)$$

The interaction energy of the electrostatic variables can be written as a general second order functional shown in eq 2.

$$\mathcal{E}_{el} = \mathcal{E}_0 + \mathbf{k}^{\dagger}\mathbf{M} + \frac{1}{2}\mathbf{M}^{\dagger}\mathbf{T}\mathbf{M} \quad (2)$$

The interaction energy of the permanent (nonpolarizable) multipoles is denoted with $\mathcal{E}_0$, while the linear $\mathbf{k}^{\dagger}\mathbf{M}$ term includes interaction between permanent and induced multipole moments. The second order term is the interaction between induced multipole moments and is defined by the multipolar operator $\hat{T}$, with the $M \times M$ matrix representation of the off-diagonal block $\mathbf{T}_{ij}$ shown in eq 3, where $f(R_{ij})$ represents a general damping function that screens the electrostatic interactions at a short distance.[42,43]

$$\mathbf{T}_{ij} = -\hat{T}\left(\frac{f(R_{ij})}{R_{ij}}\right) = -\begin{bmatrix} 1 & \nabla_x & \nabla_y & \nabla_z & ... \\ \nabla_x & \nabla_{xx} & \nabla_{xy} & \nabla_{xz} & ... \\ \nabla_y & \nabla_{yx} & \nabla_{yy} & \nabla_{yz} & ... \\ \nabla_z & \nabla_{zx} & \nabla_{zy} & \nabla_{zz} & ... \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}\left(\frac{f(R_{ij})}{R_{ij}}\right) \quad (3)$$

The diagonal blocks $\mathbf{T}_{ii}$ consist of a self-interaction term J_i for the charge−charge entry, an inverse polarizability $\boldsymbol{\alpha}_i^{-1}$ for the dipole−dipole entries, and analogous terms for higher order multipoles, eq 4.

$$\mathbf{T}_{ii} = \begin{bmatrix} J_i & 0 & ... \\ 0 & \boldsymbol{\alpha}_i^{-1} & ... \\ \vdots & \vdots & \ddots \end{bmatrix} \quad (4)$$

Table 1 shows the connection between the $\mathbf{M}$, $\mathcal{E}_0$, $\mathbf{k}$, and $\mathbf{T}$ terms in eq 2 and the variables in the FQ, Drude oscillator, and induced dipole, as well as the recently proposed BC models.

The FQ is a charge-only model, and $\mathbf{M}$ is in this case a set of polarizable (fluctuating) charges. The $\mathcal{E}_0$ term is null as there are no permanent (nonpolarizable) charges. The Drude oscillator and induced dipole models introduce polarization at the dipole level, where the induced dipoles in the Drude model arise from a pair of point charges at the nucleus and at a virtual site with a field-dependent position, while those in the induced dipole model are generated by atomic polarizabilities. The $\mathbf{k}$ term in the FQ model is a set of atomic electronegativity parameters, while $\mathbf{k}$ in the Drude and induced dipole models represents the electric field arising from the permanent multipoles. In the CHARMM implementation of the Drude model,[21] the permanent multipoles are the nuclear centered charges, while in the AMOEBA implementetion of the induced dipole model,[22] the permanent multipoles include atomic charges, dipoles, and quadrupoles. The charge self-interaction parameters in the FQ model are usually denoted atomic hardnesses.

Polarization models can be derived from different starting points, but the end result can be described by the general form

Table 1. Key Quantities for the FQ, Drude Oscillator, Induced Dipole, and BC Models[a]

term	FQ	Drude	induced dipole	BC
$\mathbf{M}$	charges ($\mathbf{Q}$)	dipoles ($\boldsymbol{\mu}$)[b]	dipoles ($\boldsymbol{\mu}$)	charges ($\mathbf{Q}$)
$\mathcal{E}_0$	none	inter. of perm. multip.[b]	inter. of perm. multip.[c]	none
$\mathbf{k}$	elneg. param. ($\boldsymbol{\chi}$)	field from nucl. charges	field from perm. multip.	none
$\mathbf{T}$	$\mathbf{T}_{ii} = J_i$	$\mathbf{T}_{ii} = \boldsymbol{\alpha}_i^{-1}$	$\mathbf{T}_{ii} = \boldsymbol{\alpha}_i^{-1}$	$\mathbf{T}_{ii} = 0$
$\mathbf{A}$	$\mathbf{T}$	$\mathbf{T}$	$\mathbf{T}$	$(\mathbf{I} + \tilde{\mathbf{C}\mathbf{J}})$
$\mathbf{b}$	$-\mathbf{k}$	$-\mathbf{k}$	$-\mathbf{k}$	$-\mathbf{C}\boldsymbol{\chi} + \mathbf{Q}_0$

[a]The $\mathbf{A}$ and $\mathbf{b}$ terms are introduced later in this subsection, and the BC model is reviewed in the next section. [b]Denotes that the induced dipoles in the Drude model arise from point charges on the nucleus and on a field-sensitive virtual particle. In the CHARMM implementation of the Drude model, the static multipoles are represented by the nuclear centered fixed charges. [c]Denotes that in the AMOEBA implementation of the induced dipole, permanent multipoles include charges, dipoles, and quadrupoles.

in eq 2. All existing polarization models furthermore determine the polarizable multipoles by making the electrostatic energy in eq 2 stationary, as shown in eq 5.

$$\frac{\partial \mathcal{E}_{el}}{\partial \mathbf{M}} = \mathbf{k} + \mathbf{T}\mathbf{M} = 0$$
$$\mathbf{M} = -\mathbf{T}^{-1}\mathbf{k} \quad (5)$$

The energy gradient required for performing MD simulations is obtained by differentiating eq 2 with respect to the $\sigma = \{x,y,z\}$ component of vector $\mathbf{R}_i$ by applying the chain rule, eq 6.

$$\frac{d\mathcal{E}_{el}}{d\sigma_i} = \frac{\partial \mathcal{E}_0}{\partial \sigma_i} + \left(\frac{\partial \mathbf{k}}{\partial \sigma_i}\mathbf{M} + \frac{1}{2}\mathbf{M}^{\dagger}\frac{\partial \mathbf{T}}{\partial \sigma_i}\mathbf{M}\right) + (\mathbf{k} + \mathbf{T}\mathbf{M})\frac{\partial \mathbf{M}}{\partial \sigma_i} \quad (6)$$

The second term on the right-hand side of eq 6 only requires explicit differentiation of $\mathbf{k}$ and $\mathbf{T}$ with respect to Cartesian components, but the third term requires the calculation of the multipolar response to the geometric perturbation. The latter formally requires solving $3N$ linear systems, corresponding to $3N$ geometrical perturbations, each having dimension $M{\cdot}N$ (total number of multipoles in the system). For any realistically sized system, this is prohibitively expensive computationally. If the multipoles are derived from a variational minimization of the energy, however, the multipolar response term vanishes, as seen from eq 5, yielding eq 7.

$$\frac{d\mathcal{E}_{el}}{d\sigma_i} = \frac{\partial \mathcal{E}_0}{\partial \sigma_i} + \left(\frac{\partial \mathbf{k}}{\partial \sigma_i}\mathbf{M} + \frac{1}{2}\mathbf{M}^{\dagger}\frac{\partial \mathbf{T}}{\partial \sigma_i}\mathbf{M}\right) \quad (7)$$

This is in analogy to the Hellmann−Feynmann theorem[40] in quantum mechanics, where the calculation of energy gradients for variational wave functions does not require calculating the wave function response.

We now consider the case where the coupled N sets of multipoles $\mathbf{M}$ are not determined by minimizing the energy (eq 5) but rather from solving a linear system of equations defined in eq 8.

$$\mathbf{A}\mathbf{M} = \mathbf{b} \quad (8)$$

The $\mathbf{A}$ matrix here represents the interaction kernel for the specific nonvariational model. The classical Hellmann−

DOI: 10.1021/acs.jctc.9b00721
J. Chem. Theory Comput. 2019, 15, 6213−6224

Feynmann theorem cannot be applied in this case, and calculation of the energy gradient thus formally requires the calculation of the multipolar response term in eq 6. A very similar problem is encountered in electronic structure theory when evaluating linear-response properties for nonvariational wave functions where wave function response terms are needed. As first shown by Handy and Schaefer,[44] however, the calculation of the response terms can be avoided by making use of the Lagrange method of undetermined multipliers,[45] and we exploit the same idea here. The starting point is to define a Lagrangian by augmenting the electrostatic energy in eq 2 with the condition used to compute $\mathbf{M}$ (eq 8), premultiplied by a vector of undetermined multipliers $\boldsymbol{\lambda}$, eq 9.

$$\mathcal{L}(\mathbf{R}, \mathbf{M}, \boldsymbol{\lambda}) = \mathcal{E}_0 + \mathbf{k}^\dagger\mathbf{M} + \frac{1}{2}\mathbf{M}^\dagger\mathbf{T}\mathbf{M} + \boldsymbol{\lambda}^\dagger(\mathbf{A}\mathbf{M} - \mathbf{b}) \tag{9}$$

We note that the function value of the Lagragian in eq 9 is identical to the energy in eq 2, once the multipoles are determined from eq 8. The Lagrange multipliers $\boldsymbol{\lambda}$ are defined by the condition that the Lagrange function is stationary with respect to variations of the multipoles $\mathbf{M}$, eq 10, where the symmetry of $\mathbf{T}$ has been employed.

$$\frac{\partial\mathcal{L}(\mathbf{R}, \mathbf{M}, \boldsymbol{\lambda})}{\partial\mathbf{M}} = \mathbf{A}^\dagger\boldsymbol{\lambda} + \mathbf{k} + \mathbf{T}\mathbf{M} = 0$$
$$\boldsymbol{\lambda} = -(\mathbf{A}^\dagger)^{-1}(\mathbf{k} + \mathbf{T}\mathbf{M}) \tag{10}$$

We note that the stationary condition in eq 10 does not make the model variational, as it is used to define the Lagrange multipliers $\boldsymbol{\lambda}$, not the multipoles $\mathbf{M}$. The energy gradient can now be replaced by the derivative of the Lagrange function in eq 9, yielding eq 11.

$$\begin{aligned}\frac{d\mathcal{E}_{el}}{d\sigma_i} &= \frac{d\mathcal{L}(\mathbf{R}, \mathbf{M}, \boldsymbol{\lambda})}{d\sigma_i} \\ &= \frac{\partial\mathcal{E}_0}{\partial\sigma_i} + \left(\frac{\partial\mathbf{k}}{\partial\sigma_i}\mathbf{M} + \frac{1}{2}\mathbf{M}^\dagger\frac{\partial\mathbf{T}}{\partial\sigma_i}\mathbf{M}\right) \\ &\quad + (\mathbf{A}^\dagger\boldsymbol{\lambda} + \mathbf{k} + \mathbf{T}\mathbf{M})^\dagger\frac{\partial\mathbf{M}}{\partial\sigma_i} + \boldsymbol{\lambda}^\dagger\left(\frac{\partial\mathbf{A}}{\partial\sigma_i}\mathbf{M} - \frac{\partial\mathbf{b}}{\partial\sigma_i}\right)\end{aligned} \tag{11}$$

The three terms involving the multipolar response disappear due to eq 10, yielding the expression for the gradient in eq 12.

$$\frac{\partial\mathcal{L}}{\partial\sigma_i} = \frac{\partial\mathcal{E}_0}{\partial\sigma_i} + \frac{\partial\mathbf{k}}{\partial\sigma_i}\mathbf{M} + \frac{1}{2}\mathbf{M}^\dagger\frac{\partial\mathbf{T}}{\partial\sigma_i}\mathbf{M} + \boldsymbol{\lambda}^\dagger\left(\frac{\partial\mathbf{A}}{\partial\sigma_i}\mathbf{M} - \frac{\partial\mathbf{b}}{\partial\sigma_i}\right) \tag{12}$$

The expression for the energy gradient of a set of nonvariational coupled multipoles in eq 12 is completely general. Compared to the energy gradient in eq 7, the gradient in eq 12 contains additional terms corresponding to the product of the Lagrange multipliers with explicit geometry derivatives of the kernels of the equation for calculating the multipoles. The important point is that the Lagrange multipliers are independent of the geometry perturbation and can be determined by solving a single set of linear eqs (eq 10), rather than $3N$ sets of equations required for determining the multipole response, implying a computational savings of a factor of $\sim 3N$. The computational cost of solving eq 10 is comparable to that for solving eq 5, making the calculation of the gradient for nonvariational energy expressions feasible for any systems amendable to calculations using variational energy functions. We will in the next section exemplify the procedure and computational implementation of the gradient in eq 12 for the recently proposed BC model, which is a charge-only polarization model using nonvariational charges for modeling the electrostatic energy.

Application of the Lagrangian Formulation of the Electrostatic Energy to the Bond Capacity Model. The FQ, induced dipole, and Drude oscillator models all employ the variational energy criterion in eq 5 to determine the electrostatic variables, in terms of charge flow or induced dipole moments, respectively. The total charge is a conserved quantity, but the variational condition in eq 5 does not guarantee that. Charge conservation for the whole system can be fulfilled in the FQ model by introducing a Lagrange multiplier, but this leads to a global coupling of all atoms, regardless of distance, and results in metallic conductivity and artificial long-range charge transfer.[39] Charge conservation in predefined fragments can be enforced by introducing additional Lagrange multipliers, but this prohibits genuine charge transfer between fragments. It furthermore leads to a very large number of constraint parameters, for example, when performing simulations with explicit water, as each water molecule is a separate fragment. The unphysical charge transfer problem is also present in induced charge-dipole polarization models,[46] corresponding to combining the FQ and induced dipole models[47] and representing atomic charges by a continuous distribution. Applications of this method introduce Lagrange multipliers to ensure charge conservation within molecular fragments or employ an *ad hoc* two-parameter exponential damping factor to provide the correct long-range charge transfer behavior.[48,49] The polarization response in the induced dipole and Drude models can be obtained without additional constraints, as there is no bound on the sum of induced dipole moments.

The proposed BC model is in analogy to FQ a charge-only polarization model, but charge conservation is guaranteed by virtue of the model itself without the need for Lagrange multipliers and with guaranteed integer charge fragments at large separation, while still allowing charge transfer to occur at contact distances.[39] The key feature of the BC model is that the charge transferred between a pair of atoms is proportional to the product of a bond capacity and an electrostatic potential difference between the atom pair, and the distance dependence of the bond capacity prohibits long-range charge transfer and leads to a linear scaling of polarizability with system size, in contrast to the FQ model, as discussed in more detail in ref 39. The BC model is capable of smoothly connecting the covalent bonded and dissociation limits, and this is a valuable property for designing reactive force field models.[50] The BC model introduces additional flexibility by separating the calculation of the atomic charges from the equation for calculating the electrostatic energy resulting from interaction of the charges, and this implies a nonvariational energy with respect to the atomic charges. For simplicity, the BC model is in the following described without periodic boundary conditions (PBC), while the PBC implementation is discussed in the next section.

As derived in detail in ref 39, the full set of atomic charges $\mathbf{Q}$ is obtained by solving eq 13, which is the model specific version of eq 8.

DOI: 10.1021/acs.jctc.9b00721
J. Chem. Theory Comput. 2019, 15, 6213−6224

$$(\mathbf{I} + \mathbf{C}\tilde{\mathbf{J}})\mathbf{Q} = [-\mathbf{C}\chi + \mathbf{Q}_0] \tag{13}$$

The $\tilde{\mathbf{J}}$ matrix describes a screened Coulomb charge−charge interaction and is the charge-only analogue of $\mathbf{T}$ in eq 3, with the modification that the diagonal elements $\tilde{J}_i$ describe charge self-interaction and are equivalent to the atomic hardness parameters in the FQ model. The vectors χ and $\mathbf{Q}_0$ represent atomic electronegativity parameters and reference atomic charges for describing non-neutral systems, respectively. The $\mathbf{C}$ matrix is analogous to the Maxwell capacitance matrix[39,51] for electrical systems and is defined in eq 14.

$$C_{ij} = \begin{cases} \sum_{k\neq j}^{N} \xi_{ik}^0 g(R_{ik}) & i = j \\ -\xi_{ij}^0 g(R_{ij}) & i \neq j \end{cases} \tag{14}$$

The ξ_{ij}^0 represents the bond capacity parameter at the bond equilibrium distance for the ij atom pair, and $g(R_{ij})$ is an attenuation function taking a value of 1 at equilibrium distance and decaying to 0 for large interatomic distances R_{ij}. The $g(R_{ij})$ attenuation function ensures that charge transfer can only occur between atoms that are within wave function overlap, and the structure of the $\mathbf{C}$ matrix ensures that the total charge is constrained to the sum of the elements in the $\mathbf{Q}_0$ vector.[39] The $\boldsymbol{\xi}^0$ parameters can be obtained from calculated molecular polarizabilities, while the χ parameters can be determined subsequently from calculated molecular electrostatic potentials. We show in the Supporting Information that the set of BC charges can be derived by minimizing an energy expression corresponding to a set of coupled capacitors. This capacitor energy expression, however, differs from the physically motivated Coulomb energy expression in eq 15, which is the counterpart of eq 2.

$$\mathcal{E}_{\mathrm{el}} = \frac{1}{2}\mathbf{Q}^{\dagger}\mathbf{J}\mathbf{Q} \tag{15}$$

Note that $\mathbf{J}$ differs from $\tilde{\mathbf{J}}$ by excluding self-interaction diagonal elements, that is, self-interaction is included when calculating the charges, but not when calculating the interaction energy. The electrostatic energy in eq 15 is not stationary with respect to the BC charges, and the energy gradient therefore contains nonzero charge-response terms. Rewritting eq 6 for the specific case of nonvariational BC charges gives eq 16.

$$\frac{d\mathcal{E}_{\mathrm{el}}}{d\sigma_i} = \frac{\partial \mathcal{E}_{\mathrm{el}}}{\partial \sigma_i} + \frac{\partial \mathcal{E}_{\mathrm{el}}}{\partial \mathbf{Q}}\frac{\partial \mathbf{Q}}{\partial \sigma_i} = \frac{1}{2}\mathbf{Q}^{\dagger}\frac{\partial \mathbf{J}}{\partial \sigma_i}\mathbf{Q} + \frac{\partial \mathcal{E}_{\mathrm{el}}}{\partial \mathbf{Q}}\frac{\partial \mathbf{Q}}{\partial \sigma_i} \tag{16}$$

As outlined in the previous section, the charge-response term can be avoided by the Lagrange technique, providing the general gradient expression in eq 12. For the specific case of the BC model, the $\mathbf{A}$ and $\mathbf{b}$ elements are identified from eq 13 and given in Table 1 together with other quantities. The interaction kernel $\mathbf{A}$ in the BC model is a nonsymmetric matrix as a direct consequence of the properties of Maxwell's capacitance matrix $\mathbf{C}$. The equivalence of eqs 9, 10, and 12 for the specific case of the BC model are shown in eqs 17, 18, and 19, respectively.

$$\mathcal{L}(\mathbf{R}, \mathbf{Q}, \lambda) = \frac{1}{2}\mathbf{Q}^{\dagger}\mathbf{J}\mathbf{Q} + \lambda^{\dagger}[(\mathbf{I} + \mathbf{C}\tilde{\mathbf{J}})\mathbf{Q} + \mathbf{C}\chi - \mathbf{Q}_0] \tag{17}$$

$$(\mathbf{I} + \tilde{\mathbf{J}}\mathbf{C})\lambda = -\mathbf{J}\mathbf{Q} \tag{18}$$

$$\frac{\partial \mathcal{L}(\mathbf{R}, \mathbf{Q}, \lambda)}{\partial \sigma_i} = \frac{1}{2}\mathbf{Q}^{\dagger}\frac{\partial \mathbf{J}}{\partial \sigma_i}\mathbf{Q} + \lambda^{\dagger}\left[\frac{\partial \mathbf{C}}{\partial \sigma_i}(\tilde{\mathbf{J}}\mathbf{Q} + \chi) + \mathbf{C}\frac{\partial \tilde{\mathbf{J}}}{\partial \sigma_i}\mathbf{Q}\right] \tag{19}$$

Implementation of the Bond Capacity Model for Molecular Dynamics with Periodic Boundary Conditions. An accurate description of the electrostatic energy is of vital importance for simulating polar systems where electrostatic interactions often play a dominating role. The electrostatic interactions are long range, and this can be problematic when the size of the simulation box is comparable to the range of such interactions. The use of PBC allows for taking into account the interactions with the periodic images of the simulation cell. However, a simple distance based truncation of the Coulomb interactions can lead to nonphysical artifacts,[52,53] and all periodic images must be taken into account. This can be done by the Ewald summation method that is reviewed in the Supporting Information, together with the Smooth Particle Mesh Ewald (SPME) reformulation[54] adopted in this work. Using these techniques for implementing the BC model in the Tinker-HP package requires, in analogy with other polarization models, that the equations requiring a formal inversion of a large dimension matrix are reformulated as iterative procedures using the potential $\phi_i(\mathbf{Q})$ from all (other) charges as the fundamental quantity. For solving the equation for the set of Lagrange multipliers λ, we in addition need to introduce a generalized potential $\phi_i(\boldsymbol{\omega})$, which is the potential at $\mathbf{R}_i$ due to a set of generalized charges $\boldsymbol{\omega}$ and all their periodic images. For a detailed discussion of the procedure employed to compute $\phi_i(\boldsymbol{\omega})$ with the SPME, we again refer to the Supporting Information material. The $\tilde{\phi}(\boldsymbol{\omega})$ similarly represents the generalized potential $\phi(\boldsymbol{\omega})$ where the self-interaction terms have been included via the J_i parameters, eq 20.

$$\tilde{\phi}_i(\boldsymbol{\omega}) = \phi_i(\boldsymbol{\omega}) + J_i\omega_i \tag{20}$$

The BC energy gradient needed for performing MD simulations requires that eqs 13, 18, and 19 are solved at each time step. The direct matrix inversion for calculating the set of charges $\mathbf{Q}$ and multipliers λ is unfeasible for large systems and must be replaced by an iterative procedure. The defining equation for $\mathbf{Q}$ (eq 13) can be rearranged into eq 21, where the $\tilde{\mathbf{J}}\mathbf{Q}$ term is replaced with the generalized potential in eq 20.

$$\mathbf{Q} = -\mathbf{C}(\chi + \tilde{\mathbf{J}}\mathbf{Q}) + \mathbf{Q}_0 \Rightarrow \mathbf{Q} = -\mathbf{C}(\chi + \tilde{\phi}(\mathbf{Q})) + \mathbf{Q}_0 \tag{21}$$

The equality in eq 21 is satisfied when $\mathbf{Q}$ is obtained from eq 13, but it can be recast into an iterative form as shown in eq 22.

$$\mathbf{Q}^{[n+1]} = -\mathbf{C}(\chi + \tilde{\phi}(\mathbf{Q}^{[n]})) + \mathbf{Q}_0 \tag{22}$$

Here, $\tilde{\phi}(\mathbf{Q}^{[n]})$ represents the electrostatic potential calculated with the SPME method from the charges at the nth iteration and augmented with the self-interaction terms.

The equation defining the multipliers (eq 18) can similarly be rearranged to give eq 23 where $\tilde{\phi}(\mathbf{C}\lambda)$ is the generalized potential in eq 20 with $\boldsymbol{\omega} = \mathbf{C}\lambda$.

$$\lambda = -\tilde{\mathbf{J}}\mathbf{C}\lambda - \mathbf{J}\mathbf{Q} \Rightarrow \lambda = -\tilde{\phi}(\mathbf{C}\lambda) - \phi(\mathbf{Q}) \tag{23}$$

DOI: 10.1021/acs.jctc.9b00721
J. Chem. Theory Comput. 2019, 15, 6213−6224

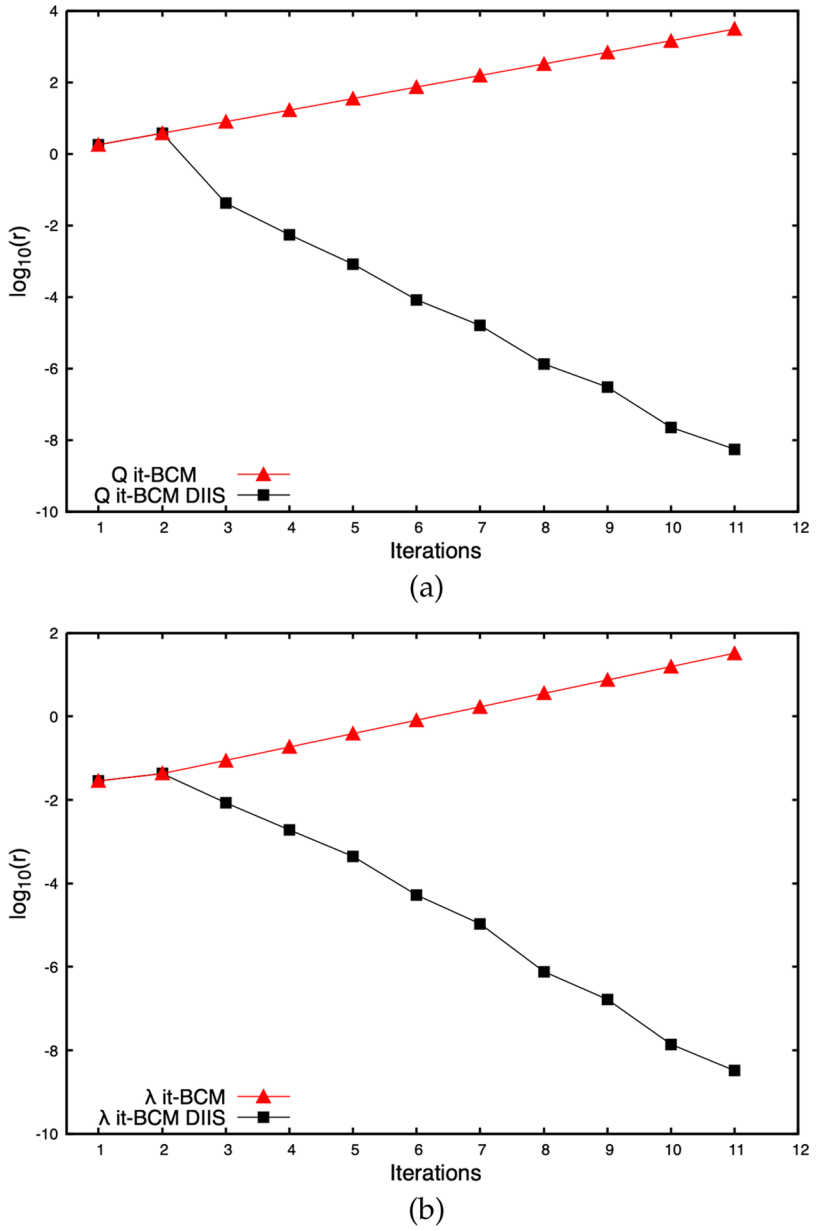


Figure 1. (a) Iterative solution of eq 22 (charges $\mathbf{Q}$) with and without DIIS extrapolation. (b) Iterative solution of eq 24 (Lagrange multipliers $\boldsymbol{\lambda}$). The x-axis shows the number of iterations while the y-axis shows the logarithm of the residual's norm $r = \frac{1}{\sqrt{N}} \|\boldsymbol{\tau}^{[n+1]} - \boldsymbol{\tau}^{[n]}\|_l{}^2$, where $\boldsymbol{\tau}$ represents either $\mathbf{Q}$ or $\boldsymbol{\lambda}$. The starting guess has in both cases been taken as the null vector $\mathbf{0}$.

The equality is satisfied for a set of $\boldsymbol{\lambda}$ obtained via eq 18 and can be turned into the iterative form shown in eq 24.

$$\boldsymbol{\lambda}^{[n+1]} = -\tilde{\phi}(\mathbf{C}\boldsymbol{\lambda}^{[n]}) - \phi(\mathbf{Q}) \tag{24}$$

The term $\phi(\mathbf{Q})$ is the potential due to the set of converged charges $\mathbf{Q}$ (and all their periodic images), and this implies that the solution of eq 24 requires a prior solution of eq 22. It is possible that more efficient algorithms can be derived that solve eqs 22 and 24 simultaneously. As mentioned above, the $\mathbf{A}$ matrix is nonsymmetric, and this excludes the possibility of, for example, employing the conjugate gradient family of methods for this specific problem. The solution of both eqs 22 and 24 is in the present implementation coupled with the DIIS extrapolation procedure,[55] and the initial guesses $\mathbf{Q}^{[0]}$ and $\boldsymbol{\lambda}^{[0]}$ can either be taken as the converged sets from the previous time step or extrapolated with the Kolafa predictor algorithm.[56] Eqs 22 and 24 represent iterative procedures for the formal inversion of the $\mathbf{A}$ and $\mathbf{A}^{\dagger}$ matrices, respectively, and numerical analyses regarding the convergence are reported in the next section.

The gradient of the Lagrange function can now be written in terms of the SPME generalized potentials $\phi(\mathbf{Q})$ and $\tilde{\phi}(\mathbf{Q})$ generated by the converged set of charges and their periodic images, eq 25.

$$\frac{\partial \mathcal{L}(\mathbf{R}, \mathbf{Q}, \boldsymbol{\lambda})}{\partial \sigma_i} = \frac{1}{2}\mathbf{Q}^{\dagger}\frac{\partial \phi(\mathbf{Q})}{\partial \sigma_i} + \boldsymbol{\lambda}^{\dagger}\left[\frac{\partial \mathbf{C}}{\partial \sigma_i}(\tilde{\phi}(\mathbf{Q}) + \chi) + \mathbf{C}\frac{\partial \tilde{\phi}(\mathbf{Q})}{\partial \sigma_i}\right] \tag{25}$$

DOI: 10.1021/acs.jctc.9b00721
J. Chem. Theory Comput. 2019, 15, 6213−6224

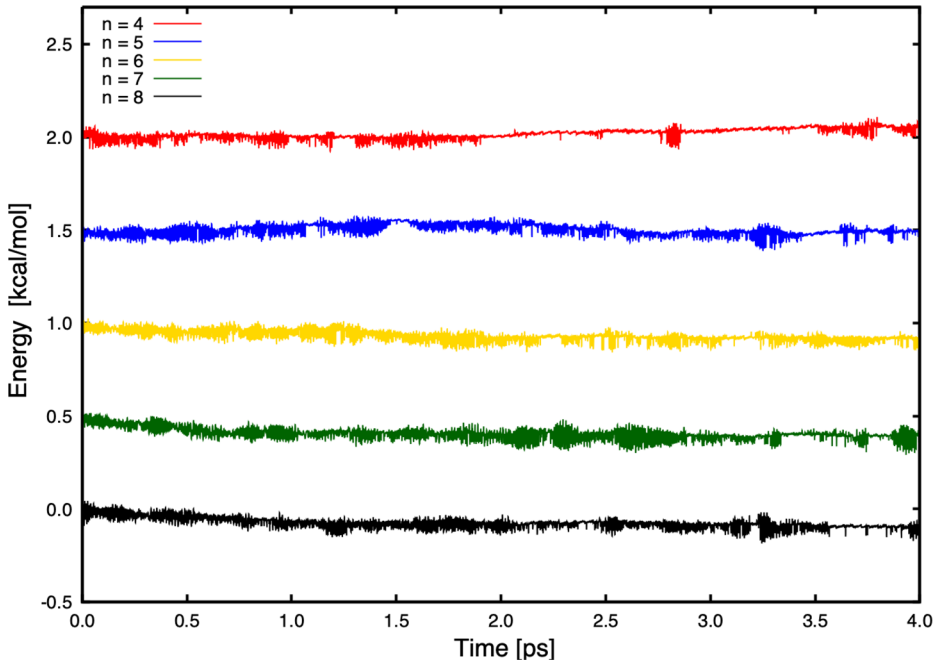


Figure 2. Energy conservation for MD simulations in the microcanonical ensemble of $[H_2O]_{500}$ as a function of the common ($\mathbf{Q}$ and $\boldsymbol{\lambda}$) convergence threshold parameter $\gamma = 10^{-n}$. The energy scale is offset by 0.5 kcal/mol for increasing values of n.

NUMERICAL RESULTS

All calculations have been performed with a 24.66 Å cubic box containing 500 water molecules ($[H_2O]_{500}$) except for the tests of the parallelization performance, which have been performed with a box of side lengths 49.32 Å containing 4000 water molecules. The bonding and Lennard-Jones parameters are taken from the TIP3P water model[57] and the electrostatics have been parametrized to reproduce TIP3P atomic charges. This choice of parameters allows a test of the correct implementation of the BC model by comparing to TIP3P results, but we emphasize that the BC model has the potential for improved accuracy by a refined parametrization, but this is beyond the scope of the present work. Fifth-order B-spline functions have been employed for the SPME together with a charge grid of size 30 × 30 × 30 with the Ewald parameter being $\beta = 0.42$, and Ewald and van der Waals cutoffs are set to 9 Å.

The iterative calculation of $\mathbf{Q}$ and $\boldsymbol{\lambda}$ using eqs 22 and 24 must be sufficiently converged to ensure stable MD simulations. We define $\boldsymbol{\tau}^{[n]}$ as the vector representing either $\mathbf{Q}$ or $\boldsymbol{\lambda}$ at iteration n and $\mathcal{F}(\boldsymbol{\tau}^{[n]})$ the function corresponding to the right-hand sides of eqs 22 and 24. Denoting the convergence threshold γ and the DIIS extrapolation procedure $\mathcal{G}$, the structure of the solver is shown schematically below.

while $(r \geq \gamma)$ *and* $(iter \leq maxiter)$ **do**
  $\boldsymbol{\tau}_{tmp}^{[n+1]} = \mathcal{F}(\boldsymbol{\tau}^{[n]})$;
  if $(use\ DIIS)$ **then**
    $\boldsymbol{\tau}^{[n+1]} = \mathcal{G}(\boldsymbol{\tau}_{tmp}^{[n+1]}, \boldsymbol{\tau}^{[n]}, \boldsymbol{\tau}^{[n-1]}, \dots)$;
  else
    $\boldsymbol{\tau}^{[n+1]} = \boldsymbol{\tau}_{tmp}^{[n+1]}$;
  end
  $r = \frac{1}{\sqrt{N}} \left\| \boldsymbol{\tau}^{[n+1]} - \boldsymbol{\tau}^{[n]} \right\|_{l^2}$
  $iter = iter + 1$
end

Figure 1 shows the numerical convergence of the iterative BC equations based on the above scheme. The simple iterative solution of the BC equations shows divergence for both eq 22 and eq 24, while the use of the DIIS extrapolation algorithm ensures convergence in both cases. Similar results have been found in prior work where the convergence of the polarization equations for the induced dipole model was analyzed for different iterative methods.[58] Eqs 22 and 24 can be seen as fixed point iterative procedures for inverting $(\mathbf{I} + \mathbf{C}\tilde{\mathbf{J}})$ and $(\mathbf{I} + \tilde{\mathbf{J}}\mathbf{C})$, respectively, and convergence is only guaranteed for strictly diagonally dominant matrices. The DIIS algorithm can be considered as an extrapolation method when applied to solving nonlinear equations, but when used in connection with linear equations, as in the present case, it becomes equivalent to the Generalized Minimal Residual method (GMRES). The GMRES belongs to the Krylov subspace class of methods, and it has guaranteed convergence properties, even when the matrix is not diagonally dominant.[59]

The atomic trajectories in an MD simulation are obtained by solving Newton's equation of motion and in practice are integrated by a variety of algorithms where the time variable is discretized into finite time steps. The size of the time step is the main parameter for controlling the accuracy of the integration, and it can operationally be probed by monitoring

DOI: 10.1021/acs.jctc.9b00721
J. Chem. Theory Comput. 2019, 15, 6213−6224

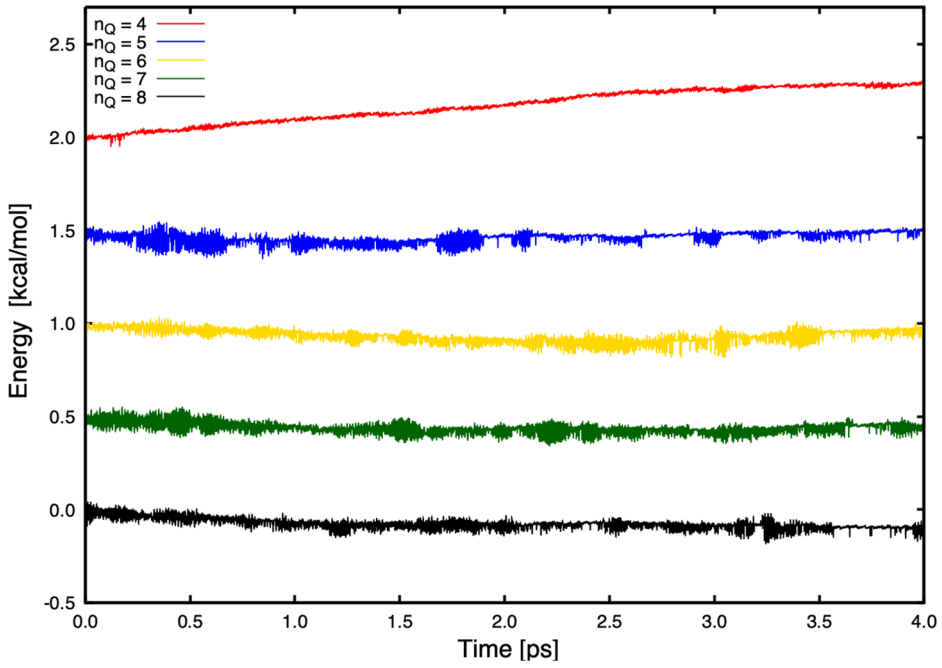


Figure 3. Energy conservation for MD simulations in the microcanonical ensemble of $[H_2O]_{500}$ as a function of the **Q** convergence threshold parameter $\gamma = 10^{-n}$, with the λ convergence threshold = 10^{-8}. The energy scale is offset by 0.5 kcal/mol for increasing values of n.

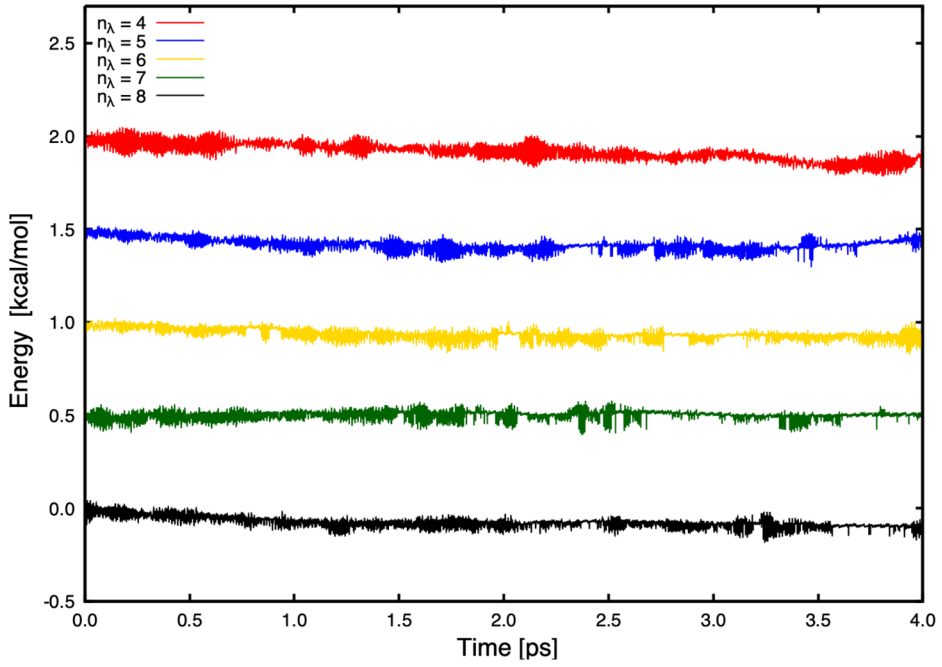


Figure 4. Energy conservation for MD simulations in the microcanonical ensemble of $[H_2O]_{500}$ as a function of the λ convergence threshold parameter $\gamma = 10^{-n}$, with the **Q** convergence threshold = 10^{-8}. The energy scale is offset by 0.5 kcal/mol for increasing values of n.

how closely the total energy is conserved when simulating a microcanonical ensemble where the total energy $\mathcal{E}$, the volume V, and the number of particles N are conserved quantities. Another possible source of energy drift when employing polarizable force fields is the incomplete convergence of the self-consistent equations used in calculating the potential and forces in each time step. It is therefore important to investigate the effects of the convergence threshold γ on the dynamics in the implemented BC model, for both the charges and multipliers. Figure 2 shows the time dependence of the total energy as a function of the common convergence parameter γ using a time step of 0.25 fs. Convergence to 10^{-4} leads to a small energy drift, but for γ values in the range 10^{-5} to 10^{-8} the energy is conserved within the numerical noise. Figures 3 and 4 show corresponding plots where the sensitivity toward the convergence threshold of either **Q** or λ is probed. Figure 3 shows that a γ value of 10^{-4} for the convergence of **Q** leads to an energy drift compared to smaller γ values. Note that the energy drift is larger than in Figure 2, which indicates that some error compensation with respect to the λ convergence takes place. Figure 4 shows that the γ convergence threshold for λ is less sensitive than for **Q**, which agrees with the

DOI: 10.1021/acs.jctc.9b00721
J. Chem. Theory Comput. 2019, 15, 6213−6224

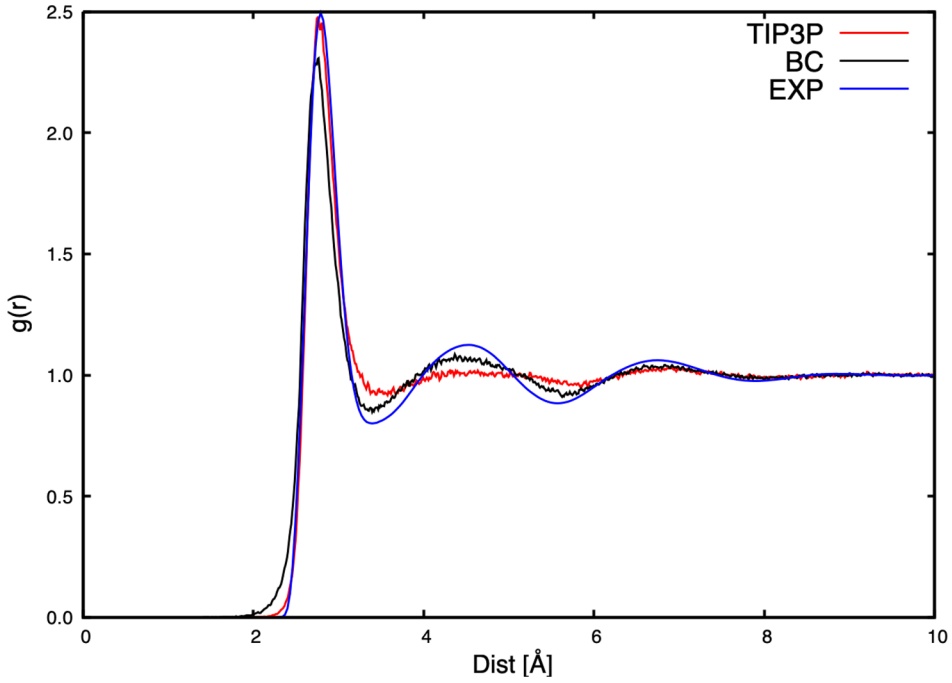


Figure 5. Radial distribution function for the oxygen−oxygen pair, calculated from a 0.5 ns trajectory of the $[H_2O]_{500}$ system in the isothermal−isobaric ensemble at 300 K and 1 atm. The experimental data have been extracted from ref 61 and refer to 298 K.

observation that the contribution from the Lagrange terms in eq 19 are smaller than the contribution from the $\mathbf{Q}$ term. A common γ value of 10^{-6} would appear sufficient to ensure that the limiting error is due to the size of the time step, and this convergence can, according to Figure 1, be achieved in typically eight iterations if starting from a null vector for both $\mathbf{Q}$ or $\boldsymbol{\lambda}$. In a simulation, however, the starting guess can be taken as the values from the previous time step, or extrapolated using information from several previous time steps. Simply using the converged solutions from the previous time step typically reduces the required number of iterations to 7. We have also tested the effect of employing Kolafa's predictor[56] using information from the previous six time steps, but this led to only marginal improvements of a few percent in the total number of iterations over a whole simulation. This is slightly surprising, as a factor of 2 improvement has been found in other work,[60] and implies that alternative predictor formulas should be investigated.

Figure 5 shows the oxygen−oxygen radial distribution function for the current implementation of the BC model compared to the corresponding TIP3P and experimental results.[61] The overall similarity of the BC and TIP3P curves serves as a validation of the implementation, with differences due to the inclusion of polarization in the BC model not present in the TIP3P results. We note that including polarization by the BC model provides a slightly better agreement with the experimental results for the second solvation shell, and specific parametrization may provide further improvements. The computational requirements for a simulation of the $[H_2O]_{500}$ system with the polarizable BC model is with the present implementation roughly a factor of 2 slower than for a corresponding simulation using the polarizable AMOEBA model. The AMOEBA model employs $3N$ electrostatic variables (x, y, z dipole components) that must be calculated in each time step, while the BC model has $2N$ variables ($\mathbf{Q}$, $\boldsymbol{\lambda}$) and should thus inherently be competitive. The better performance of AMOEBA arises from the reduction in the number of iterations for solving for the $3N$ electrostatic variables by the Kolafa extrapolation, while this extrapolation leads to essentially no improvement for the BC case. This suggests that there likely is room for computational improvements of the present implementation, both in the algorithm for solving the key equations of the BC model and for designing suitable extrapolation formulas.

The feasibility of performing MD simulations on realistically sized systems requires that the computational implementation shows good scalability over the number of computational cores. The implementation of the BC model follows the 3D spatial decomposition parallel framework of Tinker-HP[41] with the cubic simulation box being split into smaller boxes, each assigned to an MPI process (that may contain several cores in a hybrid MPI/OpenMP context). Each process calculates in each time step the interactions between atoms whose midpoint lies in its box using the midpoint rule.[62] The 3D grid used for the reciprocal space is similarly split into 2D pencils, each of which being assigned to a process that does the required calculations by placing the charges on the grid, followed by a Fast Fourier Transform and extracting the potential from the grid.

The solution of eqs 22 and 24 requires the calculation of the electrostatic potential due to the charges (or generalized charges) in each iteration. The direct part of these potentials requires that each process has to communicate the values of its (local) charges (or generalized charges) to the processes assigned to neighboring regions of space. Due to the use of the symmetry, these interactions are calculated only once for each atom pair, and each process must therefore do a symmetrical communication of the obtained potentials to the neighboring process afterward. Because there may be discrepancies between the real space 3D decomposition and the reciprocal space 2D decomposition, additional communication may be required before each iteration to compute the reciprocal part of the

DOI: 10.1021/acs.jctc.9b00721
J. Chem. Theory Comput. 2019, 15, 6213−6224

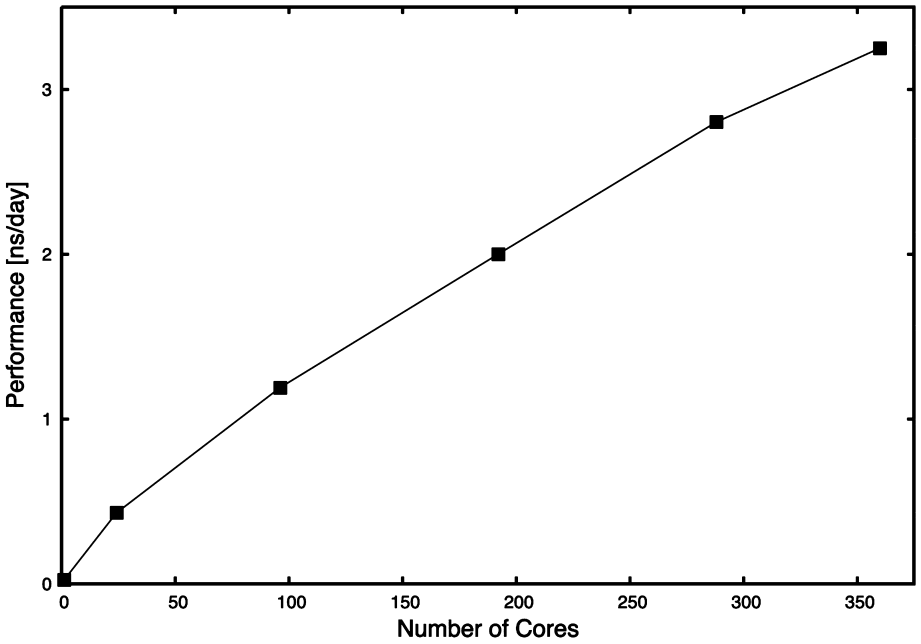


Figure 6. Performance (in nanoseconds per day) as a function of the number of cores for a cubic box of water containing 4000 molecules with an edge of 49.32 Å.

electrostatic potential. As usual in this parallel paradigm, a small amount of communication between processes responsible for adjacent parts of the 3D grid is required to compute the reciprocal potential. Once the iterations for the charges and the Lagrange multipliers have converged, the calculation of the forces is exactly the same as for calculating the forces due to permanent charges in Tinker-HP, and this requires only a small amount of communication for the reciprocal part. Figure 6 shows the parallel performance on the Occigen machine at GENCI (CINES, Montpellier, France) with Intel Xeon E5−2690v3 processors (24 Haswell cores at 2.6 GHz per node). We have in the present case used a self-consistent procedure in each time step for the polarizable molecular dynamics, since this allows the use of advanced multi-time-step strategies[63] that significantly reduces the computational cost for a fixed amount of sampling. The charge and Lagrange variables could alternatively be propagated by an extended Lagrange formalism using fictive masses, in analogy with the virtual particles in the Drude model, but this would require the use of significantly smaller time steps.[64]

CONCLUSIONS

We argue that the Bond Capacity model is an attractive component of designing new force field methods as it allows representation of molecular polarization using only charges for representing the electrostatic interaction. It ensures charge conservation without the need for additional constraint parameters, and it can be parametrized directly using results from electronic structure calculations. It furthermore provides a smooth connection between the covalent bonded and dissociation limits. The Bond Capacity model differs from other polarization models by not determining the electrostatic variables (charges) by a minimization of the electrostatic energy. This formally implies solution of a very large number of response equations in order to obtain molecular gradients needed for molecular dynamics simulations, and this would make such nonvariational models unfeasible. We show that by employing a Lagrange technique it is possible to calculate energy gradients of parametrized energy models using nonvariational polarization multipoles with only a small computational overhead compared to variational models. The additional requirement consists of solving a single set of linear equations similar to those for determining the electrostatic variables themselves. The Bond Capacity model is implemented with periodic boundary conditions in the Tinker-HP program package, and proof-of-concept molecular dynamics simulations are presented. We note that the Bond Capacity model has been coupled with continuum solvent models, which may be useful for calculations not employing periodic boundary conditions.[65] The presented framework is completely general and provides additional flexibility in designs of new polarizable and future reactive force fields.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.9b00721.

Alternative derivation of the Bond Capacity equations and a short review of the Ewald summation method and the Smooth Particle Mesh Ewald approximation (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: frj@chem.au.dk.

ORCID

Pier Paolo Poier: 0000-0003-0907-7242
Louis Lagardère: 0000-0002-7251-0910
Jean-Philip Piquemal: 0000-0001-6615-9426
Frank Jensen: 0000-0002-4576-5838

Notes

The authors declare no competing financial interest.

DOI: 10.1021/acs.jctc.9b00721
J. Chem. Theory Comput. 2019, 15, 6213−6224

■ ACKNOWLEDGMENTS

P.P.P. and F.J. acknowledge support from the Danish Council for Independent Research, Grant No. 4181-00030B. Computations have been performed thanks to GENCI on the Occigen machine (CINES, Montpellier, France) on grant A0050707671. This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 810367), project EMC2 (J.P.P.). J.P.P. acknowledges the funding of the Institut Universitaire de France.

■ REFERENCES

(1) Frenkel, D.; Smit, B. *Understanding Molecular Simulation*, 2nd ed.; Academic Press, Inc.: Orlando, FL, 2001.
(2) Rapaport, D. C. *The Art of Molecular Dynamics Simulation*, 2nd ed.; Cambridge University Press: New York, 2004.
(3) Rahman, A. Correlations in the Motion of Atoms in Liquid Argon. *Phys. Rev.* **1964**, *136*, A405−A411.
(4) Verlet, L. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **1967**, *159*, 98−103.
(5) Karplus, M.; Weaver, D. L. Protein-folding dynamics. *Nature* **1976**, *260*, 404−406.
(6) Case, D. A.; Cheatham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668−1688.
(7) Brooks, B. R.; et al. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(8) Jorgensen, W. L.; Tirado-Rives, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666 PMID: 27557051. .
(9) Plazinski, W.; Lonardi, A.; Hünenberger, P. H. Revision of the GROMOS 56A6CARBO force field: Improving the description of ring-conformational equilibria in hexopyranose-based carbohydrates chains. *J. Comput. Chem.* **2016**, *37*, 354−365.
(10) Shaw, D. E.; et al. Anton, a Special-purpose Machine for Molecular Dynamics Simulation. *Commun. ACM* **2008**, *51*, 91−97.
(11) Lane, T. J.; Shukla, D.; Beauchamp, K. A.; Pande, V. S. To milliseconds and beyond: challenges in the simulation of protein folding. *Curr. Opin. Struct. Biol.* **2013**, *23*, 58−65 Folding and binding/Protein-nucleic acid interactions. .
(12) Piana, S.; Klepeis, J. L.; Shaw, D. E. Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **2014**, *24*, 98−105 Folding and binding/Nucleic acids and their protein complexes. .
(13) Robinson, M. K.; Monroe, J. I.; Shell, M. S. Are AMBER Force Fields and Implicit Solvation Models Additive? A Folding Study with a Balanced Peptide Test Set. *J. Chem. Theory Comput.* **2016**, *12*, 5631−5642 PMID: 27731628. .
(14) Henriksen, N. M.; Gilson, M. K. Evaluating Force Field Performance in Thermodynamic Calculations of Cyclodextrin Host−Guest Binding: Water Models, Partial Charges, and Host Force Field Parameters. *J. Chem. Theory Comput.* **2017**, *13*, 4253−4269 PMID: 28696692. .
(15) Robustelli, P.; Piana, S.; Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E4758−E4766.
(16) Rappe, A. K.; Goddard, W. A. Charge equilibration for molecular dynamics simulations. *J. Phys. Chem.* **1991**, *95*, 3358−3363.
(17) Applequist, J.; Carl, J. R.; Fung, K.-K. Atom dipole interaction model for molecular polarizability. Application to polyatomic molecules and determination of atom polarizabilities. *J. Am. Chem. Soc.* **1972**, *94*, 2952−2960.
(18) Drude, P. *The Theory of Optics*; Longmans, Green, and Co.: New York, 1902.
(19) Huang, J.; Simmonett, A. C.; Pickard, F. C.; MacKerell, A. D.; Brooks, B. R. Mapping the Drude polarizable force field onto a multipole and induced dipole model. *J. Chem. Phys.* **2017**, *147*, 161702.
(20) Wang, Z.-X.; Zhang, W.; Wu, C.; Lei, H.; Cieplak, P.; Duan, Y. Strike a balance: Optimization of backbone torsion parameters of AMBER polarizable force field for simulations of proteins and peptides. *J. Comput. Chem.* **2006**, *27*, 781−790.
(21) Lamoureux, G.; Roux, B. Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm. *J. Chem. Phys.* **2003**, *119*, 3025−3039.
(22) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B* **2010**, *114*, 2549−2564 PMID: 20136072. .
(23) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P. Anisotropic, Polarizable Molecular Mechanics Studies of Inter- and Intramolecular Interactions and Ligand-Macromolecule Complexes. A Bottom-Up Strategy. *J. Chem. Theory Comput.* **2007**, *3*, 1960−1986 PMID: 18978934. .
(24) Jakobsen, S.; Jensen, F. Searching the Force Field Electrostatic Multipole Parameter Space. *J. Chem. Theory Comput.* **2016**, *12*, 1824−1832 PMID: 26925529. .
(25) Mei, Y.; Simmonett, A. C.; Pickard, F. C.; DiStasio, R. A.; Brooks, B. R.; Shao, Y. Numerical Study on the Partitioning of the Molecular Polarizability into Fluctuating Charge and Induced Atomic Dipole Contributions. *J. Phys. Chem. A* **2015**, *119*, 5865−5882 PMID: 25945749. .
(26) Nistor, R. A.; Polihronov, J. G.; Müser, M. H.; Mosey, N. J. A generalization of the charge equilibration method for nonmetallic materials. *J. Chem. Phys.* **2006**, *125*, 094108.
(27) Smirnov, K. S. Assessment of split-charge equilibration model for development of polarizable force fields. *Modell. Simul. Mater. Sci. Eng.* **2015**, *23*, 074006.
(28) Chelli, R.; Procacci, P.; Righini, R.; Califano, S. Electrical response in chemical potential equalization schemes. *J. Chem. Phys.* **1999**, *111*, 8569−8575.
(29) Chen, J.; Martínez, T. J. QTPIE: Charge transfer with polarization current equalization. A fluctuating charge model with correct asymptotics. *Chem. Phys. Lett.* **2007**, *438*, 315−320.
(30) Lee Warren, G.; Davis, J. E.; Patel, S. Origin and control of superlinear polarizability scaling in chemical potential equalization methods. *J. Chem. Phys.* **2008**, *128*, 144110.
(31) Stone, A. *The Theory of Intermolecular Forces*, 2nd ed.; Oxford University Press: Oxford, UK, 2001.
(32) Misquitta, A. J.; Stone, A. J. Distributed polarizabilities obtained using a constrained density-fitting algorithm. *J. Chem. Phys.* **2006**, *124*, 024111.
(33) Loboda, O.; Ingrosso, F.; Ruiz-López, M. F.; Szalewicz, K.; Millot, C. Geometry-dependent distributed polarizability models for the water molecule. *J. Chem. Phys.* **2016**, *144*, 034304.
(34) Morita, A.; Kato, S. Ab Initio Molecular Orbital Theory on Intramolecular Charge Polarization: Effect of Hydrogen Abstraction on the Charge Sensitivity of Aromatic and Nonaromatic Species. *J. Am. Chem. Soc.* **1997**, *119*, 4021−4032.
(35) Cho, M. Correlation between electronic and molecular structure distortions and vibrational properties. I. Adiabatic approximations. *J. Chem. Phys.* **2003**, *118*, 3480−3490.
(36) Ishida, T. Optimal Charge and Charge Response Determination through Conformational Space: Global Fitting Scheme for Representative Charge and Charge Response Kernel. *J. Phys. Chem. A* **2008**, *112*, 7035−7046 PMID: 18610947. .
(37) Nakano, H.; Yamamoto, T.; Kato, S. A wave-function based approach for polarizable charge model: Systematic comparison of polarization effects on protic, aprotic, and ionic liquids. *J. Chem. Phys.* **2010**, *132*, 044106.

DOI: 10.1021/acs.jctc.9b00721
J. Chem. Theory Comput. 2019, 15, 6213−6224

(38) Isegawa, M.; Kato, S. Polarizable Force Field for Protein with Charge Response Kernel. *J. Chem. Theory Comput.* **2009**, *5*, 2809−2821 PMID: 26631793. .
(39) Poier, P. P.; Jensen, F. Describing Molecular Polarizability by a Bond Capacity Model. *J. Chem. Theory Comput.* **2019**, *15*, 3093−3107 PMID: 30920212. .
(40) Feynman, R. P. Forces in Molecules. *Phys. Rev.* **1939**, *56*, 340−343.
(41) Lagardère, L.; Jolly, L.-H.; Lipparini, F.; Aviat, F.; Stamm, B.; Jing, Z. F.; Harger, M.; Torabifard, H.; Cisneros, G. A.; Schnieders, M. J.; Gresh, N.; Maday, Y.; Ren, P. Y.; Ponder, J. W.; Piquemal, J.-P. Tinker-HP: a massively parallel molecular dynamics package for multiscale simulations of large complex systems with advanced point dipole polarizable force fields. *Chem. Sci.* **2018**, *9*, 956−972.
(42) Nalewajski, R. F.; Korchowiec, J.; Zhou, Z. Molecular hardness and softness parameters and their use in chemistry. *Int. J. Quantum Chem.* **1988**, *34*, 349−366.
(43) Thole, B. Molecular polarizabilities calculated with a modified dipole interaction. *Chem. Phys.* **1981**, *59*, 341−350.
(44) Handy, N. C.; Schaefer, H. F. On the evaluation of analytic energy derivatives for correlated wave functions. *J. Chem. Phys.* **1984**, *81*, 5031−5033.
(45) Helgaker, T.; Jørgensen, P. Configuration-interaction energy derivatives in a fully variational formulation. *Theoretica chimica acta* **1989**, *75*, 111−127.
(46) Mayer, A.; Åstrand, P.-O. A ChargeDipole Model for the Static Polarizability of Nanostructures Including Aliphatic, Olephinic, and Aromatic Systems. *J. Phys. Chem. A* **2008**, *112*, 1277−1285 PMID: 18198848. .
(47) Giovannini, T.; Puglisi, A.; Ambrosetti, M.; Cappelli, C. Polarizable QM/MM Approach with Fluctuating Charges and Fluctuating Dipoles: The QM/FQF Model. *J. Chem. Theory Comput.* **2019**, *15*, 2233−2245 PMID: 30875213. .
(48) Voora, V. K.; Jordan, K. D. Nonvalence Correlation-Bound Anion States of Spherical Fullerenes. *Nano Lett.* **2014**, *14*, 4602−4606 PMID: 24978808. .
(49) Choi, T. H.; Jordan, K. D. Model potential study of non-valence correlation-bound anions of (C60)n clusters: the role of electric field-induced charge transfer. *Faraday Discuss.* **2019**, *217*, 547−560.
(50) Liang, T.; Shin, Y. K.; Cheng, Y.-T.; Yilmaz, D. E.; Vishnu, K. G.; Verners, O.; Zou, C.; Phillpot, S. R.; Sinnott, S. B.; van Duin, A. C. Reactive Potentials for Advanced Atomistic Simulations. *Annu. Rev. Mater. Res.* **2013**, *43*, 109−129.
(51) Maxwell, J. C. *A Treatise on Electricity and Magnetism; Cambridge Library Collection - Physical Sciences*; Cambridge University Press, 2010; Vol. *1*; pp i−iv.
(52) Cheatham, T. E. I.; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A. Molecular Dynamics Simulations on Solvated Biomolecular Systems: The Particle Mesh Ewald Method Leads to Stable Trajectories of DNA, RNA, and Proteins. *J. Am. Chem. Soc.* **1995**, *117*, 4193−4194.
(53) Norberg, J.; Nilsson, L. On the Truncation of Long-Range Electrostatic Interactions in DNA. *Biophys. J.* **2000**, *79*, 1537−1553.
(54) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(55) Pulay, P. Convergence acceleration of iterative sequences. the case of scf iteration. *Chem. Phys. Lett.* **1980**, *73*, 393−398.
(56) Kolafa, J. Time-reversible always stable predictor−corrector method for molecular dynamics of polarizable molecules. *J. Comput. Chem.* **2004**, *25*, 335−342.
(57) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(58) Lagardère, L.; Lipparini, F.; Polack, E.; Stamm, B.; Cancès, E.; Schnieders, M.; Ren, P.; Maday, Y.; Piquemal, J.-P. Scalable Evaluation of Polarization Energy and Associated Forces in Polarizable Molecular Dynamics: II. Toward Massively Parallel Computations Using Smooth Particle Mesh Ewald. *J. Chem. Theory Comput.* **2015**, *11*, 2589−2599 PMID: 26575557. .
(59) Rohwedder, T.; Schneider, R. An analysis for the DIIS acceleration method used in quantum chemistry calculations. *J. Math. Chem.* **2011**, *49*, 1889.
(60) Lipparini, F.; Lagardère, L.; Stamm, B.; Cancès, E.; Schnieders, M.; Ren, P.; Maday, Y.; Piquemal, J.-P. Scalable Evaluation of Polarization Energy and Associated Forces in Polarizable Molecular Dynamics: I. Toward Massively Parallel Direct Space Computations. *J. Chem. Theory Comput.* **2014**, *10*, 1638−1651 PMID: 26512230. .
(61) Soper, A. K. The Radial Distribution Functions of Water as Derived from Radiation Total Scattering Experiments: Is There Anything We Can Say for Sure? *ISRN Phys. Chem.* **2013**, *2013*, 1−68.
(62) Bowers, K. J.; Dror, R. O.; Shaw, D. E. The midpoint method for parallelization of particle simulations. *J. Chem. Phys.* **2006**, *124*, 184109.
(63) Lagardère, L.; Aviat, F.; Piquemal, J.-P. Pushing the Limits of Multiple-Time-Step Strategies for Polarizable Point Dipole Molecular Dynamics. *J. Phys. Chem. Lett.* **2019**, *10*, 2593−2599.
(64) Ngo, V. A.; Fanning, J. K.; Noskov, S. Y. Comparative Analysis of Protein Hydration from MD simulations with Additive and Polarizable Force Fields. *Advanced Theory and Simulations* **2019**, *2*, 1800106.
(65) Poier, P. P.; Jensen, F. Including Implicit Solvation in the Bond Capacity Polarization Model. *J. Chem. Phys.* **2019**, *151*, 114118.

DOI: 10.1021/acs.jctc.9b00721
J. Chem. Theory Comput. 2019, 15, 6213−6224