JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

On the Numerical Accuracy of Ewald, Smooth Particle Mesh Ewald, and Staggered Mesh Ewald Methods for Correlated Molecular Systems

Han Wang,*[,†] Pingwen Zhang,[‡] and Christof Schütte[†]

[†]Institute for Mathematics, Freie Universität Berlin, Germany

[‡]LMAM and School of Mathematical Sciences, Peking University, Beijing, P.R. China

ABSTRACT: In this work, we develop the accurate error estimates for three state-of-art algorithms of long-range electrostatic interaction in inhomogeneous and correlated molecular systems. They are the Ewald summation, the smooth particle mesh Ewald (SPME) and the staggered mesh Ewald methods. Two branches of force computation, namely the *ik*- and analytical differentiation, are considered. All the estimates are developed by proposing a more general framework: if the error force is of pairwise form, then the root-mean-square force error is composed of three additive parts, the homogeneity error, the inhomogeneity error and the correlation error. Computationally scalable estimates (estimating the errors at the cost $O(N \log N)$) are developed for all the considered algorithms. The effectiveness of the proposed estimates and the important role of the correlation error are carefully checked and demonstrated by example systems.

INTRODUCTION

The calculation of the long-range electrostatic interaction is a very important topic in the molecular simulations. Owing to the very slow decay of the interaction strength with respect to the distance between charges, the typical short-range fast algorithms (e.g., the cutoff + neighbor list algorithm) do not work, and the energy of the system is only conditionally convergent. One of the first algorithms handling this problem is the Ewald summation, dating back to the 1920s.[1] It has been shown that with a careful tuning of working parameters, the optimal computational cost of the Ewald summation is $O(N^{3/2})$,[2] which is not feasible for modern large scale molecular simulations. Therefore, several Ewald-based fast algorithms were proposed to reduce the computational cost to a scalable level: $O(N \log N)$. Some of them are the particle mesh Ewald (PME) method,[3] the smooth particle mesh Ewald (SPME) method,[4] and the particle−particle particle mesh (P3M) method.[5,6] It has been demonstrated that the SPME is actually a special case of the P3M method.[6,7] More recently, the staggered mesh or "interlacing" technique[8,9] was applied to SPME (called the staggered mesh Ewald)[10] and P3M (called the interlaced P3M),[11] and was shown to improve the accuracy of these algorithms greatly. In this paper, we focus on the Ewald summation, the SPME, and the staggered mesh Ewald methods. The Ewald summation is the starting point of all the mentioned fast algorithms. The SPME is the one of the most popular long-range algorithms among the mainstream molecular simulation packages, for example, AMBER,[12] GROMACS,[13,14] and NAMD.[15]

The aforementioned fast algorithms have achieved great success over the past three decades. However, the fact that the user should provide six working parameters for these methods may be problematic: it is difficult to find the most efficient parameter set in such a 6-dimensional parameter space. This problem can be solved by comparing the forces calculated by using different parameter sets on a representative snapshot of the system (see for example, ref 16). An other approach is the in-depth theoretical study of the error introduced by the algorithms, namely the error estimate,[17,11,17−22] by which the error is described by a function of the working parameters. Using the error estimates, a work flow has been designed to automatically determine the nearly optimal combination of parameters, in the sense that it is the fastest one satisfying a demanding accuracy.[22] However, so far, all error estimates assume the homogeneity of the charge distribution, and the independency of any pair of charges (i.e., the charges are uncorrelated), but neither of the assumptions is satisfactorily fulfilled in most molecular systems of practical interest. For example, the lipid bilayer membrane solving in water: the hydrophobic tail(s) of the lipids carry no charge, and the charges in the system are correlated by the covalent bonds, the van der Waals interaction, and the hydrogen bonding network, etc. Therefore, the current error estimates may be problematic, or even not applicable. Several papers have shown that due to the correlation of charges, the error estimates obviously deviate from the real error.[7,20,22]

The purpose of this manuscript is to develop a reliable error estimate in inhomogeneous and correlated molecular systems. We estimate the force error because the application is the molecular dynamics (MD) simulation. First, a general error estimate framework for the pairwise interaction is set up. We show that the error may be approximately decomposed into three additive parts: the homogeneity error, the inhomogeneity error and the correlation error.[23] The computational cost of the homogeneity and inhomogeneity error is $O(N \log N)$ when the error force kernel is of convolution form. The correlation error can also be estimated at a cost of $O(N \log N)$ by the nearest neighbor approximation technique, which assumes the error is dominated by short-range correlation. This is the case for most

Received: May 2, 2012
Published: July 25, 2012

   dx.doi.org/10.1021/ct300343y | *J. Chem. Theory Comput.* 2012, 8, 3243−3256

the liquid and gas systems far from critical point. Generally speaking, when the error force can be written as a pairwise form, all estimates are derived easily under this framework. Following this framework, we develop the error estimates for the Ewald summation, the SPME, and the staggered method Ewald methods. The estimates of the homogeneity and inhomogeneity errors are tested by two randomly distributed inhomogeneous charges systems, and is proved to be very sharp. The error estimates are also tested in a real inhomogeneous water system, in which the charge correlation plays a very important role on the force error. By including the nearest neighbor approximation of the correlation error, the quality of the estimates is impressively enhanced. This paper is ended by conclusions and remarks in the last section.

THEORY

Error Estimate for the Pairwise Interactions in a Charged System. In this paper, we consider a system with periodic boundary condition. Suppose it is composed of N charged particles located at $\mathbf{r}_1, \mathbf{r}_2,\cdots,\mathbf{r}_N$ with charges $q_1, q_2,\cdots,q_N$, respectively. We study the force on a testing particle located at $\mathbf{r}$, which feels interactions from all charges $q_1, q_2,\cdots,q_N$ in the system, and does not exert force on them. The *error force* is defined by the difference between the exact force and the calculated force on the testing particle,[23] and is denoted by $\Delta\mathbf{F}(\mathbf{r})$. The magnitude of the force computation error is defined by the root-mean-square (rms) error, which is the square root of the second moment of the error force: $\mathcal{E}(\mathbf{r}) = \langle|\Delta\mathbf{F}(\mathbf{r})|^2\rangle^{1/2}$. The notation $\langle\cdot\rangle$ means the ensemble average. We also estimate the ensemble average of the error force (also called mean error force), namely $\langle\Delta\mathbf{F}(\mathbf{r})\rangle$, which plays an important role in the error analysis of the inhomogeneous systems.[23]

One of the main results of this research is the following estimate for the error force that can be written as a summation of pairwise interactions:

Theorem 1 Let a periodic molecular system be composed of N charged particles located at $\mathbf{r}_1, \mathbf{r}_2,\cdots,\mathbf{r}_N$ with charges $q_1, q_2,\cdots,q_N$, respectively. If the error force of the testing particle with charge q has the form:

$$\Delta\mathbf{F}(\mathbf{r}) = q\sum_{j=1}^{N} q_j\mathbf{K}(\mathbf{r}, \mathbf{r}_j) \tag{1}$$

then the mean error force is

$$\langle\Delta\mathbf{F}(\mathbf{r})\rangle = q\int_{\mathbb{R}^3}\mathbf{K}(\mathbf{r}, \mathbf{r}')\,\rho_q(\mathbf{r}')\,d\mathbf{r}' \tag{2}$$

and the rms error is

$$\begin{aligned}\mathcal{E}^2(\mathbf{r}) = {} & q^2\int_{\mathbb{R}^3}|\mathbf{K}(\mathbf{r}, \mathbf{r}')|^2\,\rho_{q^2}(\mathbf{r}')\,d\mathbf{r}' \\ & + q^2\Big[\int_{\mathbb{R}^3}\mathbf{K}(\mathbf{r}, \mathbf{r}')\,\rho_q(\mathbf{r}')\,d\mathbf{r}'\Big]^2 \\ & + q^2\int_{\mathbb{R}^3\times\mathbb{R}^3}\mathbf{K}(\mathbf{r}, \mathbf{r}')\cdot\mathbf{K}(\mathbf{r}, \mathbf{r}'')\,C(\mathbf{r}', \mathbf{r}'')\,d\mathbf{r}'\,d\mathbf{r}''\end{aligned} \tag{3}$$

where

$$\rho_q(\mathbf{r}) = \Big\langle\sum_{j=1}^{N} q_j\delta(\mathbf{r} - \mathbf{r}_j)\Big\rangle \tag{4}$$

$$\rho_{q^2}(\mathbf{r}) = \Big\langle\sum_{j=1}^{N} q_j^2\delta(\mathbf{r} - \mathbf{r}_j)\Big\rangle \tag{5}$$

$$\rho^{(2)}(\mathbf{r}, \mathbf{r}') = \Big\langle\sum_{j\neq k} q_jq_k\delta(\mathbf{r} - \mathbf{r}_j)\,\delta(\mathbf{r}' - \mathbf{r}_k)\Big\rangle \tag{6}$$

$$C(\mathbf{r}, \mathbf{r}') = \rho^{(2)}(\mathbf{r}, \mathbf{r}') - \rho_q(\mathbf{r})\,\rho_q(\mathbf{r}') \tag{7}$$

Notice the definitions of $\rho_q(\mathbf{r})$, $\rho_{q^2}(\mathbf{r})$ and $\rho^{(2)}(\mathbf{r},\mathbf{r}')$ are periodically extended to $\mathbb{R}^3$, $\mathbb{R}^3$ and $\mathbb{R}^3\times\mathbb{R}^3$, respectively.

Proof By definition, the mean error force is

$$\begin{aligned}\langle\Delta\mathbf{F}(\mathbf{r})\rangle &= q\Big\langle\sum_{j=1}^{N} q_j\mathbf{K}(\mathbf{r}, \mathbf{r}_j)\Big\rangle \\ &= q\int_{\mathbb{R}^3}\mathbf{K}(\mathbf{r}, \mathbf{r}')\Big\langle\sum_{j=1}^{N} q_j\delta(\mathbf{r}' - \mathbf{r}_j)\Big\rangle\,d\mathbf{r}' \\ &= q\int_{\mathbb{R}^3}\mathbf{K}(\mathbf{r}, \mathbf{r}')\,\rho_q(\mathbf{r}')\,d\mathbf{r}'.\end{aligned}$$

The rms force error is calculated by

$$\begin{aligned}\langle|\Delta\mathbf{F}(\mathbf{r})|^2\rangle &= q^2\Big\langle\sum_{j,k} q_jq_k\mathbf{K}(\mathbf{r}, \mathbf{r}_j)\cdot\mathbf{K}(\mathbf{r}, \mathbf{r}_k)\Big\rangle \\ &= q^2\Big\langle\sum_{j=1}^{N} q_j^2\,|\mathbf{K}(\mathbf{r}, \mathbf{r}_j)|^2\Big\rangle + q^2\Big\langle\sum_{j\neq k} q_jq_k\mathbf{K}(\mathbf{r}, \mathbf{r}_j)\cdot\mathbf{K}(\mathbf{r}, \mathbf{r}_k)\Big\rangle \\ &= q^2\int_{\mathbb{R}^3}|\mathbf{K}(\mathbf{r}, \mathbf{r}')|^2\rho_{q^2}(\mathbf{r}')\,d\mathbf{r} + q^2\int_{\mathbb{R}^3\times\mathbb{R}^3}\mathbf{K}(\mathbf{r}, \mathbf{r}')\cdot\mathbf{K}(\mathbf{r}, \mathbf{r}'')\rho^{(2)}(\mathbf{r}', \mathbf{r}'')\,d\mathbf{r}'\,d\mathbf{r}'' \\ &= q^2\int_{\mathbb{R}^3}|\mathbf{K}(\mathbf{r}, \mathbf{r}')|^2\rho_{q^2}(\mathbf{r}')\,d\mathbf{r}' + q^2\Big[\int_{\mathbb{R}^3}\mathbf{K}(\mathbf{r}, \mathbf{r}')\rho_q(\mathbf{r}')\,d\mathbf{r}'\Big]^2 + q^2\int_{\mathbb{R}^3\times\mathbb{R}^3}\mathbf{K}(\mathbf{r}, \mathbf{r}')\cdot\mathbf{K}(\mathbf{r}, \mathbf{r}'')C(\mathbf{r}', \mathbf{r}'')\,d\mathbf{r}'\,d\mathbf{r}''\end{aligned} \tag{8}$$

Corollary 2 If the function $\mathbf{K}$ has the property

$$\mathbf{K}(\mathbf{r}, \mathbf{r}') = \mathbf{K}(\mathbf{r} - \mathbf{r}') \tag{9}$$

then the error estimates are convolutions, namely,

$$\langle\Delta\mathbf{F}(\mathbf{r})\rangle = q[\mathbf{K}*\rho_q](\mathbf{r}) \tag{10}$$

and

$$\begin{aligned}\mathcal{E}^2(\mathbf{r}) = {} & q^2[(\mathbf{K})^2*\rho_{q^2}](\mathbf{r}) + q^2[\mathbf{K}*\rho_q]^2(\mathbf{r}) \\ & + q^2\int_{\mathbb{R}^3\times\mathbb{R}^3}\mathbf{K}(\mathbf{r} - \mathbf{r}')\cdot\mathbf{K}(\mathbf{r} - \mathbf{r}'')C(\mathbf{r}, \mathbf{r}'')\,d\mathbf{r}'\,d\mathbf{r}''\end{aligned} \tag{11}$$

where the asterisk $(*)$ denotes the convolution.

Here are some remarks concerning Theorem 1 and Corollary 2:

•The densities $\rho_q(\mathbf{r})$ and $\rho_{q^2}(\mathbf{r})$ defined in eqs 4 and 5 are called *the first order charge distribution* and *the second order charge distribution*, respectively. $\rho^{(2)}(\mathbf{r},\mathbf{r}')$ in eq 6 is the *pairwise charge distribution*. When the positions of any two different charges are independent, $\rho^{(2)}(\mathbf{r},\mathbf{r}') = ((N-1)/N)\rho_q(\mathbf{r})\rho_q(\mathbf{r}')$. Since N is always a large number in real simulations, we take $\rho^{(2)}(\mathbf{r}, \mathbf{r}') = \rho_q(\mathbf{r})\rho_q(\mathbf{r}')$ for convenience. $C(\mathbf{r}, \mathbf{r}')$ defined in eq 7 is the *pairwise charge correlation function*, which describes the strength of the correlation between two different charges. When they are independent, $C(\mathbf{r}, \mathbf{r}')$ vanishes, otherwise $C(\mathbf{r}, \mathbf{r}') \neq 0$. Most previous error estimates assumed the independency of the particles in the system,[7,11,17−20,22,23] however, this could be problematic in the strongly correlated systems.[7,20,22] Also see the Example 3 presented later in this paper.

•The *pairwise form* of the interaction (see eq 1) is the key to derive the error estimates 2 and 3. The function $\mathbf{K}(\mathbf{r}, \mathbf{r}')$ is called *the error force kernel*. By setting $q_i = q_j = 1$, the Corollary 2 proves the error estimates for the short-range interaction in inhomogeneous systems, which was derived in ref 23.

•Analogous to the short-range error estimate,[23] the three terms on the right hand side (rhs) of eq 3 are called the homogeneity error, the inhomogeneity error, and the correlation error, respectively. This is denoted by

$$\mathcal{E}^2(\mathbf{r}) = \mathcal{E}_{\text{homo}}{}^2(\mathbf{r}) + \mathcal{E}_{\text{inhomo}}{}^2(\mathbf{r}) + \mathcal{E}_{\text{correlation}}(\mathbf{r}) \tag{12}$$

The first term on the rhs of eq 12 is called the "homogeneity error", but it does NOT mean this term is homogeneous or originated from a homogeneous density profile. It is named so because the error estimate only contains this term in the homogeneous and uncorrelated system. It is the term studied by most previous error estimates. Equation 1 calculates the homogeneity and inhomogeneity error at a cost of $O(N^2)$, and correlation error at a cost of $O(N^3)$. By writing the error estimates in the convolution form in the Corollary 2, the homogeneity error and the inhomogeneity error can be calculated at the cost of $O(N \log N)$ by the fast Fourier transform (FFT) (see ref 23 for details). However, the full calculation of the correlation error costs $O(N^2 \log N)$, because it involves a 2-fold convolution.

•A nonvanishing mean error force 2 was proved to be harmful to the calculation of the short-range interaction in inhomogeneous systems.[23] In the case of the charged system, this term is nonzero only when the system is NOT locally neutral, that is, a nonzero first order charge distribution. For most of the simulation cases, the local neutrality condition is satisfied, so the mean error force and the inhomogeneity error vanish.

Nearest Neighbor Approximation of the Correlation Error. The full computation of the correlation error costs at least $O(N^2 \log N)$, which is still not tractable. So we consider the nearest neighbor approximation to reduce the computational cost. Here the definition of the "nearest neighbors" is very flexible. Basically it means the nearby neighbors that are strongly correlated. It could be defined either by the charges falling in some certain neighboring range, or by charges connected by chemical bonds. Rewriting eq 3 yields

$$\begin{aligned}\langle |\Delta\mathbf{F}(\mathbf{r})|^2 \rangle &= q^2 \int_{\mathbb{R}^3} |\mathbf{K}(\mathbf{r}, \mathbf{r}')|^2 \rho_{q^2}(\mathbf{r}')\, d\mathbf{r}' \\ &\quad + q^2 \int_{\mathbb{R}^3\times\mathbb{R}^3} \mathbf{K}(\mathbf{r}, \mathbf{r}')\cdot\mathbf{K}(\mathbf{r}, \mathbf{r}'')\, \rho^{(2)}(\mathbf{r}', \mathbf{r}'')\, d\mathbf{r}'\, d\mathbf{r}''\end{aligned} \tag{13}$$

where the definition of $\rho^{(2)}(\mathbf{r}', \mathbf{r}'')$ is given by eq 6. We separately consider the contribution from the nearest neighbors and other contributions to the pairwise density $\rho^{(2)}(\mathbf{r}', \mathbf{r}'')$. Denoting $\Omega_j = \{k \mid \text{atom } k \text{ is one of the nearest neighbors of atom } j\}$, we have

$$\begin{aligned}\rho^{(2)}(\mathbf{r}', \mathbf{r}'') &= \Big\langle \sum_j \sum_{k\in\Omega_j} q_j q_k \delta(\mathbf{r}' - \mathbf{r}_j)\, \delta(\mathbf{r}'' - \mathbf{r}_k) \Big\rangle \\ &\quad + \Big\langle \sum_j \sum_{k\notin\Omega_j} q_j q_k \delta(\mathbf{r}' - \mathbf{r}_j)\, \delta(\mathbf{r}'' - \mathbf{r}_k) \Big\rangle\end{aligned} \tag{14}$$

We assume two charges are independent if they are not nearest neighbors. Then second term on the rhs of eq 14 is approximately $((N - \#(\Omega_j))/N)\rho_q(\mathbf{r}')\rho_q(\mathbf{r}'') \approx \rho_q(\mathbf{r}')\rho_q(\mathbf{r}'')$, where $\#(\Omega_j)$ is the number of nearest neighbors considered. The approximation is valid only when the number of neighbors in Ω_j is small compared with the number of charges in the system. Therefore, the first term on the rhs of 14 is actually the pairwise charge correlation function:

$$C(\mathbf{r}', \mathbf{r}'') \approx \Big\langle \sum_j \sum_{k\in\Omega_j} q_j q_k \delta(\mathbf{r}' - \mathbf{r}_j)\, \delta(\mathbf{r}'' - \mathbf{r}_k) \Big\rangle \tag{15}$$

By inserting eq 15 into the correlation error (the third line of eq 3), we have

$$\mathcal{E}_{\text{correlation}}(\mathbf{r}) \approx q^2 \sum_j \sum_{k\in\Omega_j} q_j q_k \langle \mathbf{K}(\mathbf{r}, \mathbf{r}_j)\cdot\mathbf{K}(\mathbf{r}, \mathbf{r}_k) \rangle_{\mathbf{r}_j,\mathbf{r}_k} \tag{16}$$

The quality of this approximation depends on how to define the "nearest neighbors". In the systems of liquid and gas far from critical point, the correlation is only short-ranged, so charges are nearest neighbors when their correlations are considered important. It is also possible to systematically improve the approximation by considering more and more neighbors as long as the number of neighbors is small compared with the size of the system.

As an example, we consider a typical three-point charge water model TIP3P:[24] each hydrogen atom carries a charge of $q_H = +0.417e$ and each oxygen atom carries a charge of $q_O = -0.834e$. There are two types of correlation in this system: (1) the bonded correlation that originates from the rigid H−O bond and H−O−H angle within one molecule, and (2) the nonbonded correlation or the intermolecular correlation that originates from a complex coupling of various intermolecular interactions, such as the van der Waals interaction, the hydrogen bonding, etc.

We first consider the bonded correlation, and define the nearest neighbors by the atoms within the same water molecule. Denoting the index set of the oxygen atoms by Ω_O, and the index set of the hydrogen atoms by Ω_H, we have

$$\mathcal{E}_{\text{correlation}}(\mathbf{r}) = q^2 \sum_{j\in\Omega_O}\sum_{k\in\Omega_j} q_j q_k \langle \mathbf{K}(\mathbf{r}, \mathbf{r}_j)\cdot\mathbf{K}(\mathbf{r}, \mathbf{r}_k)\rangle_{\mathbf{r}_j,\mathbf{r}_k} + q^2 \sum_{j\in\Omega_H}\sum_{k\in\Omega_j} q_j q_k \langle \mathbf{K}(\mathbf{r}, \mathbf{r}_j)\cdot\mathbf{K}(\mathbf{r}, \mathbf{r}_k)\rangle_{\mathbf{r}_j,\mathbf{r}_k}$$
$$= q^2 \int_{\mathbb{R}^3} \mathbf{K}(\mathbf{r}, \mathbf{r}')\cdot 2\langle q_H \mathbf{K}(\mathbf{r}, \mathbf{r}' + \mathbf{s}_O)\rangle \rho_O(\mathbf{r}')\mathrm{d}\mathbf{r}' + q^2 \int_{\mathbb{R}^3} \mathbf{K}(\mathbf{r}, \mathbf{r}')\cdot\langle q_O \mathbf{K}(\mathbf{r}, \mathbf{r}' - \mathbf{s}_O) + q_H \mathbf{K}(\mathbf{r}, \mathbf{r}' - \mathbf{s}_H)\rangle \rho_H(\mathbf{r}')\mathrm{d}\mathbf{r}' \quad (17)$$

If positions of the three atom within one molecule are $\mathbf{r}_{H_1}$, $\mathbf{r}_{H_2}$, and $\mathbf{r}_O$, then $\mathbf{s}_O = \mathbf{r}_{H_1} - \mathbf{r}_O$ and $\mathbf{s}_H = \mathbf{r}_{H_2} - \mathbf{r}_{H_1}$. The two ensemble averages in eq 17 are taken over all possible water directions at position $\mathbf{r}'$. As an approximation, we assume these averages are spatially uniform, namely independent with $\mathbf{r}'$. It may be problematic in the regions where the preference of water direction is different with others. If the error force kernel has the form $\mathbf{K}(\mathbf{r}, \mathbf{r}') = \mathbf{K}(\mathbf{r} - \mathbf{r}')$, the integrals can be written in the convolution form and accelerated by the FFT. One possible way to calculate the ensemble averages in eq 17 is by the Fourier transform of the error force kernel, for example

$$\langle q_H \mathbf{K}(\mathbf{r} - \mathbf{r}' - \mathbf{s}_O)\rangle = \frac{1}{V}\langle \sum_{\mathbf{m}} q_H \hat{\mathbf{K}}(\mathbf{m})\, e^{2\pi i \mathbf{m}\cdot(\mathbf{r}-\mathbf{r}'-\mathbf{s}_O)}\rangle = \frac{1}{V}\sum_{\mathbf{m}} \langle q_H\, e^{-2\pi i \mathbf{m}\cdot\mathbf{s}_O}\rangle \hat{\mathbf{K}}(\mathbf{m})\, e^{2\pi i \mathbf{m}\cdot(\mathbf{r}-\mathbf{r}')}$$

where "∧" denotes the forward Fourier transform. By denoting

$$\hat{T}_O(\mathbf{m}) = 2\langle q_H\, e^{-2\pi i \mathbf{m}\cdot\mathbf{s}_O}\rangle \quad (18)$$

$$\hat{T}_H(\mathbf{m}) = \langle q_O\, e^{2\pi i \mathbf{m}\cdot\mathbf{s}_O} + q_H\, e^{-2\pi i \mathbf{m}\cdot\mathbf{s}_H}\rangle \quad (19)$$

eq 17 becomes

$$\mathcal{E}_{\text{correlation}}(\mathbf{r}) = q^2 \int_{\mathbb{R}^3} \mathbf{K}(\mathbf{r} - \mathbf{r}')\cdot(\hat{T}_O\hat{\mathbf{K}})^\vee(\mathbf{r} - \mathbf{r}')\rho_O(\mathbf{r}')\,\mathrm{d}\mathbf{r}' + q^2 \int_{\mathbb{R}^3} \mathbf{K}(\mathbf{r} - \mathbf{r}')\cdot(\hat{T}_H\hat{\mathbf{K}})^\vee(\mathbf{r} - \mathbf{r}')\rho_H(\mathbf{r}')\,\mathrm{d}\mathbf{r}' \quad (20)$$

where "∨" denotes the backward Fourier transform. In the Fourier space, the error force kernel $\mathbf{K}(\mathbf{r})$ is influenced by the prefactors T_O and T_H, which are accounting for the bonded correlation within one water molecule. It should be noted that we define the force error by inserting a testing charge into the water system, and comparing the calculated and exact force on it. This definition implicitly assumes the independency between the testing particle and the system, as illustrated by Figure 1. It is consistent with the usual setting of the bonded pair atoms: the electrostatic interaction between them is considered to be *not* calculated. Throughout this paper, we remove this interaction from the standard long-range calculations.

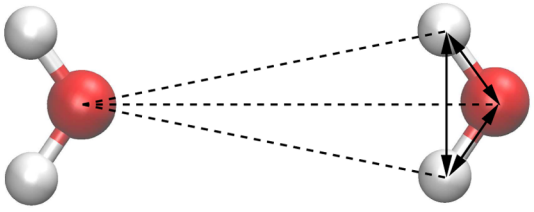

Figure 1. The schematic plot of two interacting water molecules. When we calculate the force error of the left oxygen atom (it is now a testing charge), the correlation with its bonded hydrogens are not considered. The correlations within any neighboring molecule are considered, as indicated by the solid lines with two arrows.

For the nonbonded correlation, we define the nearest neighbors by nonbonded atoms closer than a certain radius R_g (neighboring range). If the system is homogeneous and isotropic, the pair correlation function $\rho^{(2)}(\mathbf{r}', \mathbf{r}'')$ in eq 13 can be represented by the radial distribution functions (RDFs). Similar to the bonded correlation calculation, the nearest neighbor approximation of the nonbonded correlation error is estimated: the functions T_O and T_H are given by

$$\hat{T}_O(\mathbf{m}) = \int_0^{R_g} [q_H\rho_H\tilde{g}_{OH}(\mathbf{r}) + q_O\rho_O\tilde{g}_{OO}(\mathbf{r})]\frac{\sin(2\pi\mathbf{m}\mathbf{r})}{2\pi\mathbf{m}\mathbf{r}} 4\pi\mathbf{r}^2\,\mathrm{d}\mathbf{r} \quad (21)$$

$$\hat{T}_H(\mathbf{m}) = \int_0^{R_g} [q_H\rho_H\tilde{g}_{HH}(\mathbf{r}) + q_O\rho_O\tilde{g}_{OH}(\mathbf{r})]\frac{\sin(2\pi\mathbf{m}\mathbf{r})}{2\pi\mathbf{m}\mathbf{r}} 4\pi\mathbf{r}^2\,\mathrm{d}\mathbf{r} \quad (22)$$

where $\tilde{g}_{OO}(r) = g_{OO}(r) - 1$, $\tilde{g}_{OH}(r) = g_{OH}(r) - 1$, and $\tilde{g}_{HH}(r) = g_{HH}(r) - 1$. $g_{OO}(r)$, $g_{OH}(r)$, and $g_{HH}(r)$ are the RDFs of oxygen−oxygen, oxygen−hydrogen, and hydrogen−hydrogen, respectively. ρ_H and ρ_O are the homogeneous number densities of the hydrogen and oxygen atoms, respectively. For a short-range correlated system, the RDFs converge to 1 as r goes to infinity, so the integrals in eq 21 and 22 are convergent with respect to increasing R_g.

Both the bonded and nonbonded nearest neighbor approximations require some uniformity of the correlation in the system: the bonded estimate requires the uniformity of the water direction preference, while the nonbonded estimate requires the uniformity of the density distribution and the isotropicity of the system. These requirements should be carefully checked before using these error estimates. If the they are seriously violated, one can either use these estimates for some regions (or subsystems of the whole system) that still preserve the uniformity, or estimate the correlation error at the cost of $O(N^2 \log N)$, if the correlation in the system is highly inhomogeneous.

Ewald Summation and Its Error Estimate. The Ewald method divides the electrostatic interaction in to three parts: the direct part, the reciprocal part, and the correction part:

$$E = E_{\text{dir}} + E_{\text{rec}} + E_{\text{corr}} \quad (23)$$

where

$$E_{\text{dir}} = \frac{1}{2}\sum_{\mathbf{n}}^{*}\sum_{i,j=1}^{N} \frac{q_i q_j\,\mathrm{erfc}(\beta|\mathbf{r}_{ij} + \mathbf{n}|)}{|\mathbf{r}_{ij} + \mathbf{n}|} \quad (24)$$

$$E_{\text{rec}} = \frac{1}{2\pi V}\sum_{\mathbf{m}\neq 0} \frac{\exp(-\pi^2\mathbf{m}^2/\beta^2)}{\mathbf{m}^2} S(\mathbf{m})S(-\mathbf{m}) \quad (25)$$

$$E_{\text{corr}} = -\frac{\beta}{\sqrt{\pi}}\sum_{i=1}^{N} q_i^2 \quad (26)$$

The distance between the two particles is denoted by $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$. The lattice in the real space is denoted by $\boldsymbol{n} = n_1\boldsymbol{a}_1 + n_2\boldsymbol{a}_2 +$

$n_3\boldsymbol{a}_3$, where $\boldsymbol{a}_\alpha$, $\alpha = 1, 2, 3$ are box vectors. The structure factor $S(\mathbf{m})$ is defined by

$$S(\mathbf{m}) = \sum_{j=1}^{N} q_j \exp(2\pi i \mathbf{m}\cdot\mathbf{r}_j) \tag{27}$$

where $\mathbf{m} = m_1\mathbf{a}_1^* + m_2\mathbf{a}_2^* + m_3\mathbf{a}_3^*$ is the reciprocal space lattice vector. $\mathbf{a}_\alpha^*$, $\alpha = 1, 2, 3$ are conjugate reciprocal vectors of $\mathbf{a}_\alpha$, defined by $\mathbf{a}_\alpha \cdot \mathbf{a}_\beta^* = \delta_{\alpha\beta}$. and $V = \mathbf{a}_1 \times \mathbf{a}_2 \times \mathbf{a}_3$ is the volume of the box.

The complementary error function erfc(r) in eq 24 decays exponentially as r goes to infinity. Therefore, it is a short-range interaction in the real space, which can be calculated by the standard cutoff and neighbor list method[25] at a cost of $O(N)$. The reciprocal part also decays exponentially as the magnitude of the Fourier mode $|\mathbf{m}|$ increases. Therefore, in practice, the infinite summation in eq 25 is truncated, and only a finite summation is calculated:

$$E_{\mathrm{rec}}^{\mathrm{tr}} = \sum_{\substack{|m_\alpha|<K_\alpha/2 \\ \mathbf{m}\neq 0}} f(\mathbf{m})\, S(\mathbf{m})\, S(-\mathbf{m}) \tag{28}$$

where, for short,

$$f(\mathbf{m}) := \frac{\exp(-\pi^2\mathbf{m}^2/\beta^2)}{2\pi V \mathbf{m}^2} \tag{29}$$

and K_α is the number of Fourier modes used on direction α. The truncated reciprocal force acting on particle i is

$$\mathbf{F}_{\mathrm{rec}}^{\mathrm{tr}}(\mathbf{r}_i) = \sum_j q_i q_j \sum_{\substack{|m_\alpha|<K_\alpha/2 \\ \mathbf{m}\neq 0}} \mathbf{g}(\mathbf{m})\, \mathrm{e}^{2\pi i \mathbf{m}\cdot\mathbf{r}_{ij}} \tag{30}$$

where

$$\mathbf{g}(\mathbf{m}) = -4\pi \mathbf{m} i f(\mathbf{m}) \tag{31}$$

Both the error force kernel of the direct and reciprocal part of the Ewald summation can be written in the form of eq 9. The direct part is

$$\mathbf{K}_{\mathrm{dir}}^{\mathrm{cut}}(r) = \begin{cases} 0, & r \le r_c \\ \left[\dfrac{2\beta}{\sqrt{\pi}}\, \mathrm{e}^{-\beta^2 r^2} + \dfrac{\mathrm{erfc}(\beta r)}{r}\right]\dfrac{\mathbf{r}}{r^2}, & r > r_c \end{cases} \tag{32}$$

where $r = |\mathbf{r}|$, and r_c is the cutoff radius in the real space. The reciprocal error force kernel is given by

$$\mathbf{K}_{\mathrm{rec}}^{\mathrm{tr}}(\mathbf{r}) = \sum_{|m_\alpha|\ge K_\alpha/2} \mathbf{g}(\mathbf{m})\, \mathrm{e}^{2\pi i \mathbf{m}\cdot\mathbf{r}} \tag{33}$$

By the Corollary 2, the error estimate of the truncated Ewald summation is straightforward.

The SPME and Staggered Mesh Ewald Methods. As mentioned before, the optimal computational cost of the Ewald summation is $O(N^{3/2})$, which becomes not tractable when the system size is larger than several hundreds of charged particles. Both the PME and SPME methods are designed to reduce the computational cost in nearly the same way. Starting from the reciprocal part of the Ewald summation, they first calculate the term $\mathrm{e}^{2\pi i \mathbf{m}\cdot\mathbf{r}}$ on a uniform mesh in the real space. Therefore, all calculation of structure factors can be accelerated by the fast Fourier transform (FFT); that is why the computational cost of PME/SPME is reduced to $O(N \log N)$. Then, for any particle position $\mathbf{r}_i$, the value of $\mathrm{e}^{2\pi i \mathbf{m}\cdot\mathbf{r}_i}$ is interpolated by the known values on the neighboring mesh points. The PME method uses Lagrangian interpolation, while the SPME uses B-spline interpolation, which is more precise, and has higher order of smoothness. The working parameters of these methods are the splitting parameter β, the cutoff radius in real space r_c, the order of interpolation n, and the number of mesh points on each direction: K_α, $\alpha = 1, 2, 3$. Notice $K_\alpha/2$ is the truncation of the reciprocal summation.

The SPME method provides two possibilities of calculating the reciprocal force. The first one is to differentiate (with respect to the particle position) the reciprocal term of the truncated Ewald energy and then approximate the derived force eq 30 by the B-spline interpolation. This way is called the *ik-differentiation*. The alternative way notices the high-order smoothness of the B-spline interpolation, and derive the force by differentiating the B-spline approximated reciprocal Ewald energy (also truncated). The second way was proposed by the original SPME paper,[4] and is called the *analytical differentiation*. It has been shown that with the use of the same parameters, the *ik*-differentiation is more precise than the analytical differentiation, but it uses twice more FFTs, which is a bottleneck for the communication in parallel implementations.[22] For the *ik*-differentiation, the calculated force is not the negative gradient of the calculated energy. The analytical differentiation does not have this problem, but it (to the approximation precision) violates Newton's third law. Since the SPME method is not the point of the present paper, we refer the readers to, for example, refs 4,6 and 22 for details of this method.

The staggered mesh approach or originally call "interlacing" was recently applied to SPME (staggered mesh Ewald[10]). It calculates the reciprocal force on two meshes, one of which locates at the mesh element center of the other, and then the two reciprocal forces are averaged. It is shown that staggering the meshes cancels the error in the force computation, and greatly improves the accuracy. Implementing the staggered mesh method for SPME is relatively simple, and involves only a small modification to the existing codes. Obviously, the staggered mesh *ik*-/analytical differentiation costs twice as much on the reciprocal part as the original version.

Error Estimate of the SPME *ik*-Differentiation. From the precise calculation of the reciprocal summation to the SPME fast algorithm, the electrostatic interaction is approximated by two steps. The first step is the truncation of the reciprocal summation, that is, eq 28. The second step is the approximation of $\mathrm{e}^{2\pi i \mathbf{m}\cdot\mathbf{r}}$ by B-spline interpolation. It has been shown that the error due to the first step is negligible compared with the second step, at least for the parameters of practical interest.[22] Numerical examples supporting this argument in inhomogeneous systems are also given later in this paper. Therefore, to estimate the force error, it is enough to compare the SPME force with the *truncated* Ewald force 30.

As mentioned above, the SPME error mainly originates from approximating the term $\mathrm{e}^{2\pi i \mathbf{m}\cdot\mathbf{r}}$ by the B-spline interpolation. Denoting the error introduced by this approximation by $A(\mathbf{m},\mathbf{r})\, \mathrm{e}^{2\pi i \mathbf{m}\cdot\mathbf{r}}$, we are fortunate to have the analytical expression of $A(\mathbf{m},\mathbf{r})$:[26]

$$A(\mathbf{m}, \mathbf{r}) = \sum_{\alpha=1}^{3} \sum_{l\neq 0} Z_{\alpha,l}(\mathbf{m})(\mathrm{e}^{2\pi i l \mathbf{K}_\alpha \mathbf{a}_\alpha^*\cdot\mathbf{r}} - 1) \tag{34}$$

with

$$Z_{\alpha,l}(\mathbf{m}) = \frac{\left(\frac{2\pi m_\alpha}{K_\alpha} + 2\pi l\right)^{-n}}{\sum_{l=-\infty}^{+\infty}\left(\frac{2\pi m_\alpha}{K_\alpha} + 2\pi l\right)^{-n}} \tag{35}$$

Due to the symmetry of $\mathbf{g}(\mathbf{m})$ and $A(\mathbf{m},\mathbf{r})$, the "self-interacting" term of the *ik*-differentiation vanishes.[22] The error force is therefore

$$\Delta\mathbf{F}^{\mathrm{ik}}_{\mathrm{rec}}(\mathbf{r}) = q\sum_{\substack{|m_\alpha|<K_\alpha/2\\ \mathbf{m}\neq 0}} \mathbf{g}(\mathbf{m})\, A(\mathbf{m},\mathbf{r})\, e^{2\pi i\mathbf{m}\cdot\mathbf{r}} \sum_j q_j\, e^{-2\pi i\mathbf{m}\cdot\mathbf{r}_j} + q\sum_{\substack{|m_\alpha|<K_\alpha/2\\ \mathbf{m}\neq 0}} \mathbf{g}(\mathbf{m}) e^{2\pi i\mathbf{m}\cdot\mathbf{r}} \sum_j q_j A(\mathbf{m},-\mathbf{r}_j)\, e^{-2\pi i\mathbf{m}\cdot\mathbf{r}_j} \tag{36}$$

Unfortunately, the error force kernel cannot be written in the form of eq 9, so the error estimate derived by Corollary 2 cannot be used here. However, it is still possible to develop the error estimate with convolution form under some approximations. Here we only want to present the key idea and the final error estimate, and leave all details of the derivation to the appendix. We notice the term $e^{2\pi i l K_\alpha \mathbf{a}_\alpha^*\cdot\mathbf{r}}$ with $l \neq 0$ in eq 34 introduces some high-wavenumber terms in the expression of the error estimate. If K_α is large enough, then due to the thermal fluctuation of the system, the peaks and valleys of the high-wavenumber terms may cancel in the ensemble-averaged error estimate. Therefore, by neglecting these terms, it is proved that the mean error force of the *ik*-differentiation is

$$\langle\Delta\mathbf{F}^{\mathrm{ik}}_{\mathrm{rec}}(\mathbf{r})\rangle \approx -2q[(\sum_\alpha\sum_{l\neq 0}\mathbf{G}_{\alpha,l})*\rho_q](\mathbf{r}) \tag{37}$$

and the rms force error is

$$\langle|\Delta\mathbf{F}^{\mathrm{ik}}_{\mathrm{rec}}(\mathbf{r})|^2\rangle \approx 2q^2[(\sum_\alpha\sum_{l\neq 0}|\mathbf{G}_{\alpha,l}|^2)*\rho_{q^2}](\mathbf{r}) + 4q^2[(\sum_\alpha\sum_{l\neq 0}\mathbf{G}_{\alpha,l})^2*\rho_{q^2}](\mathbf{r}) + \langle\Delta\mathbf{F}^{\mathrm{ik}}_{\mathrm{rec}}(\mathbf{r})\rangle^2 \tag{38}$$

where

$$\mathbf{G}_{\alpha,l}(\mathbf{r}) = \sum_{\substack{|m_\alpha|<K_\alpha/2\\ \mathbf{m}\neq 0}} \mathbf{g}(\mathbf{m}) Z_{\alpha,l}(\mathbf{m})\, e^{2\pi i\mathbf{m}\cdot\mathbf{r}} \tag{39}$$

$\mathbf{G}_{\alpha,l}(\mathbf{r})$ does not depend on the coordinates of particles, so it can be calculated once and stored in a table for future use. The nearest neighbor approximation to the correlation error in a water system is

$$\mathcal{E}^{\mathrm{ik}}_{\mathrm{correlation}}(\mathbf{r}) = q^2[(\sum_\alpha\sum_{l\neq 0}\mathbf{G}_{\alpha,l}\cdot[\hat{T}_{\mathrm{O}}\hat{\mathbf{G}}_{\alpha,l}]^\vee)*\rho_{\mathrm{O}}](\mathbf{r}) + 4q^2[(\sum_\alpha\sum_{l\neq 0}\mathbf{G}_{\alpha,l})\cdot(\sum_\alpha\sum_{l\neq 0}[\hat{T}_{\mathrm{O}}\hat{\mathbf{G}}_{\alpha,l}]^\vee)*\rho_{\mathrm{O}}](\mathbf{r}) + q^2[(\sum_\alpha\sum_{l\neq 0}\hat{\mathbf{G}}_{\alpha,l}\cdot[\hat{T}_{\mathrm{H}}\hat{\mathbf{G}}_{\alpha,l}]^\vee)*\rho_{\mathrm{H}}](\mathbf{r}) + 4q^2[(\sum_\alpha\sum_{l\neq 0}\mathbf{G}_{\alpha,l})\cdot(\sum_\alpha\sum_{l\neq 0}[\hat{T}_{\mathrm{H}}\hat{\mathbf{G}}_{\alpha,l}]^\vee)*\rho_{\mathrm{H}}](\mathbf{r}) \tag{40}$$

Error Estimate of the SPME Analytical Differentiation. The error force of the analytical differentiation is

$$\Delta\mathbf{F}^{\mathrm{ana}}_{\mathrm{rec}}(\mathbf{r}) = \Delta\mathbf{F}^{\mathrm{ik}}_{\mathrm{rec}}(\mathbf{r}) + q\sum_{\substack{|m_\alpha|<K_\alpha/2\\ \mathbf{m}\neq 0}} -2f(\mathbf{m})\,\mathbf{B}(\mathbf{m},\mathbf{r})\, e^{2\pi i\mathbf{m}\cdot\mathbf{r}} \sum_j q_j\, e^{-2\pi i\mathbf{m}\cdot\mathbf{r}_j} + q^2\sum_{\substack{|m_\alpha|<K_\alpha/2\\ \mathbf{m}\neq 0}} -2f(\mathbf{m})\,\mathbf{B}(\mathbf{m},\mathbf{r}) \tag{41}$$

with $\boldsymbol{B}(\mathbf{m},\mathbf{r})$ defined by

$$\mathbf{B}(\mathbf{m},\mathbf{r}) = \nabla_{\mathbf{r}}A(\mathbf{m},\mathbf{r}) = \sum_{\alpha=1}^{3}\sum_{l\neq 0} Z_{\alpha,l}(\mathbf{m})\, 2\pi i l K_\alpha \mathbf{a}_\alpha^*\, e^{2\pi i l K_\alpha \mathbf{a}_\alpha^*\cdot\mathbf{r}} \tag{42}$$

The last term in eq 41 is due to the self-interaction,[10,11,27] which does not depend on the coordinates of particles. Therefore, it can be calculated explicitly and subtracted from the analytical force during the simulation. The computational cost of the self-interaction term is low compared with the SPME force calculation. It has been shown that the self-interaction error dominates when the charge density is low.[10] Therefore, we always remove the self-interaction term from the analytical differentiation.

Neglecting all high frequency terms, we have the mean error force of analytical differentiation:

$$\langle\Delta\mathbf{F}^{\mathrm{ana}}_{\mathrm{rec}}(\mathbf{r})\rangle \approx \langle\Delta\mathbf{F}^{\mathrm{ik}}_{\mathrm{rec}}(\mathbf{r})\rangle \tag{43}$$

The rms force error is

$$\langle|\Delta\mathbf{F}^{\mathrm{ana}}_{\mathrm{rec}}(\mathbf{r})|^2\rangle \approx q^2[(\sum_\alpha\sum_{l\neq 0}|\mathbf{G}_{\alpha,l}|^2)*\rho_{q^2}](\mathbf{r}) + q^2[(\sum_\alpha\sum_{l\neq 0}|\mathbf{G}_{\alpha,l}+\mathbf{F}_{\alpha,l}|^2)*\rho_{q^2}](\mathbf{r}) + 4q^2[(\sum_\alpha\sum_{l\neq 0}\mathbf{G}_{\alpha,l})^2*\rho_{q^2}](\mathbf{r}) + \langle\Delta\mathbf{F}^{\mathrm{ana}}_{\mathrm{rec}}(\mathbf{r})\rangle^2 \tag{44}$$

where

$$\mathbf{F}_{\alpha,l}(\mathbf{r}) = \sum_{\substack{|m_\alpha|<K_\alpha/2\\ \mathbf{m}\neq 0}} -4\pi i l K_\alpha \mathbf{a}_\alpha^* f(\mathbf{m})\, Z_{\alpha,l}(\mathbf{m})\, e^{2\pi i\mathbf{m}\cdot\mathbf{r}} \tag{45}$$

Similar to $\mathbf{G}_{\alpha,l}(\mathbf{r})$, $\mathbf{F}_{\alpha,l}(\mathbf{r})$ does not depends on the coordinates of particles, so it can be tabulated for the simulation. The nearest neighbor approximation to the correlation error is

$$\mathcal{E}^{\mathrm{ana}}_{\mathrm{correlation}}(\mathbf{r}) = q^2[\{\sum_\alpha\sum_{l\neq 0}(\mathbf{G}_{\alpha,l}+\mathbf{F}_{\alpha,l})\cdot[\hat{T}_{\mathrm{O}}(\hat{\mathbf{G}}_{\alpha,l}+\hat{\mathbf{F}}_{\alpha,l})]^\vee\} *\rho_{\mathrm{O}}](\mathbf{r}) + 4q^2[(\sum_\alpha\sum_{l\neq 0}\mathbf{G}_{\alpha,l})\cdot(\sum_\alpha\sum_{l\neq 0}[\hat{T}_{\mathrm{O}}\hat{\mathbf{G}}_{\alpha,l}]^\vee)*\rho_{\mathrm{O}}](\mathbf{r}) + q^2[\{\sum_\alpha\sum_{l\neq 0}(\mathbf{G}_{\alpha,l}+\mathbf{F}_{\alpha,l})\cdot[\hat{T}_{\mathrm{H}}(\hat{\mathbf{G}}_{\alpha,l}+\hat{\mathbf{F}}_{\alpha,l})]^\vee\}*\rho_{\mathrm{H}}](\mathbf{r}) + 4q^2[(\sum_\alpha\sum_{l\neq 0}\mathbf{G}_{\alpha,l})\cdot(\sum_\alpha\sum_{l\neq 0}[\hat{T}_{\mathrm{H}}\hat{\mathbf{G}}_{\alpha,l}]^\vee)*\rho_{\mathrm{H}}](\mathbf{r}) \tag{46}$$

Error Estimate of the Staggered Mesh Ewald Method. The shift of the mesh can be effectively treated as shifting all particles on the opposite direction. Since the system is periodic, the direction of shift is not important. For convenience, we shift all particles to the positive direction by half the mesh element size, namely by a vector $\mathbf{s} = \sum_\alpha(1/(2K_\alpha))\boldsymbol{a}_\alpha$. From eq 34, we have

$$\frac{1}{2}[A(\mathbf{m}, \mathbf{r}) + A(\mathbf{m}, \mathbf{r} + \mathbf{s})] = \sum_{\alpha=1}^{3} \sum_{l\neq 0} \frac{1}{2} Z_{\alpha,l}(\mathbf{m})[(1 + e^{\pi i l}) e^{2\pi i l K_\alpha \mathbf{a}_\alpha^* \cdot \mathbf{r}} - 2] \quad (47)$$

Therefore, when l is odd, $1 + e^{\pi i l}$ vanishes. So the $1/2[A(\mathbf{m}, \mathbf{r}) + A(\mathbf{m}, \mathbf{r} + \mathbf{s})]$ can be approximated by the leading order

$$\frac{1}{2}[A(\mathbf{m}, \mathbf{r}) + A(\mathbf{m}, \mathbf{r} + \mathbf{s})] \approx -\sum_{\alpha=1}^{3} \sum_{l\neq 0} Z_{\alpha,l}(\mathbf{m}) \quad (48)$$

The validity of this approximation and a similar approximation eq 54 will be numerically checked later. Then from eq 36, we have the error force of the staggered *ik*-differentiation:

$$\Delta \mathbf{F}_{\text{rec}}^{\text{ik,st}}(\mathbf{r}) \approx -2q \sum_j q_j \sum_\alpha \sum_{l\neq 0} \sum_{\substack{|m_\alpha|<K_\alpha/2 \\ \mathbf{m}\neq 0}} \mathbf{g}(\mathbf{m}) Z_{\alpha,l}(\mathbf{m}) e^{2\pi i \mathbf{m}\cdot(\mathbf{r}-\mathbf{r}_j)} \quad (49)$$

Therefore, the error force kernel is

$$\mathbf{K}_{\text{rec}}^{\text{ik,st}}(\mathbf{r}) = -2 \sum_\alpha \sum_{l\neq 0} \mathbf{G}_{\alpha,l}(\mathbf{r}) \quad (50)$$

then the Corollary 2 is applicable. We have the error estimates:

$$\langle \Delta \mathbf{F}_{\text{rec}}^{\text{ik,st}}(\mathbf{r}) \rangle \approx -2q[(\sum_\alpha \sum_{l\neq 0} \mathbf{G}_{\alpha,l}) * \rho_q](\mathbf{r}) \quad (51)$$

and the rms force error:

$$\langle |\Delta \mathbf{F}_{\text{rec}}^{\text{ik,st}}(\mathbf{r})|^2 \rangle \approx 4q^2[(\sum_\alpha \sum_{l\neq 0} \mathbf{G}_{\alpha,l})^2 * \rho_{q^2}](\mathbf{r}) + \langle \Delta \mathbf{F}_{\text{rec}}^{\text{ik,st}}(\mathbf{r}) \rangle^2 \quad (52)$$

The superscript "st" denotes the staggered mesh method. The mean error force of the staggered *ik*-differentiation is the same as the original *ik*-differentiation. The rms error is improved, because $\langle |\Delta \mathbf{F}_{\text{rec}}^{\text{ik}}(\mathbf{r})|^2 \rangle - \langle |\Delta \mathbf{F}_{\text{rec}}^{\text{ik,st}}(\mathbf{r})|^2 \rangle = 2q^2[(\sum_\alpha \sum_{l\neq 0} |\mathbf{G}_{\alpha,l}|^2) * \rho_{q^2}](\mathbf{r}) \geq 0$. The nearest neighbor approximation to the correlation error is

$$\mathcal{E}_{\text{correlation}}^{\text{ik,st}}(\mathbf{r}) = 4q^2[(\sum_\alpha \sum_{l\neq 0} \mathbf{G}_{\alpha,l}) \cdot (\sum_\alpha \sum_{l\neq 0} [\hat{T}_{\text{O}} \hat{\mathbf{G}}_{\alpha,l}]^\vee) * \rho_{\text{O}}](\mathbf{r}) + 4q^2[(\sum_\alpha \sum_{l\neq 0} \mathbf{G}_{\alpha,l}) \cdot (\sum_\alpha \sum_{l\neq 0} [\hat{T}_{\text{H}} \hat{\mathbf{G}}_{\alpha,l}]^\vee) * \rho_{\text{H}}](\mathbf{r}) \quad (53)$$

For the analytical differentiation, similarly we have

$$\frac{1}{2}[\mathbf{B}(\mathbf{m}, \mathbf{r}) + \mathbf{B}(\mathbf{m}, \mathbf{r} + \mathbf{s})] \approx 0 \quad (54)$$

Therefore, from eq 41, the error force of the staggered analytical differentiation is approximately the same as the staggered *ik*-differentiation and so are all the error estimates:

$$\Delta \mathbf{F}_{\text{rec}}^{\text{ana,st}}(\mathbf{r}) \approx \Delta \mathbf{F}_{\text{rec}}^{\text{ik,st}}(\mathbf{r}) \quad (55)$$

$$\langle \Delta \mathbf{F}_{\text{rec}}^{\text{ana,st}}(\mathbf{r}) \rangle \approx \langle \mathbf{F}_{\text{rec}}^{\text{ik,st}}(\mathbf{r}) \rangle \quad (56)$$

$$\langle |\Delta \mathbf{F}_{\text{rec}}^{\text{ana,st}}(\mathbf{r})|^2 \rangle \approx \langle |\Delta \mathbf{F}_{\text{rec}}^{\text{ik,st}}(\mathbf{r})|^2 \rangle \quad (57)$$

$$\mathcal{E}_{\text{correlation}}^{\text{ana,st}}(\mathbf{r}) \approx \mathcal{E}_{\text{correlation}}^{\text{ik,st}}(\mathbf{r}) \quad (58)$$

The correction of the self-interacting error is not necessary, because it is automatically canceled by the staggered mesh. The rms force error is improved because it is not difficult to see $\langle |\Delta \mathbf{F}_{\text{rec}}^{\text{ana}}(\mathbf{r})|^2 \rangle - \langle |\Delta \mathbf{F}_{\text{rec}}^{\text{ana,st}}(\mathbf{r})|^2 \rangle \geq 0$. Equation 55 implies, to the leading order of approximation, the staggered mesh *ik*- and analytical differentiation are equivalent. Considering the latter uses only one-half of FFTs as the former, it might be preferable in large-scale parallel simulations. From eqs 49 and 55, it is obvious that the Newton's third law is better preserved by the staggered mesh methods.

NUMERICAL TESTS

We consider three inhomogeneous charge systems to verify our error estimates. The first two are artificially designed to verify the homogeneity and inhomogeneity error estimates. The third is a liquid−vapor equilibrium water system, in order to test the quality of the nearest neighbor approximation for the correlated system.

Example 1: A Locally Neutral Inhomogeneous System. In this example, 41472 charges are put into a 14.90 nm × 7.45 nm × 7.45 nm periodic simulation box. One third of them are carrying a negative charge of −0.834*e*, and the other two-thirds are carrying a positive charge of +0.417*e*. The position of these charges are randomly generated subjecting to the number density shown in Figure 2. The number density of

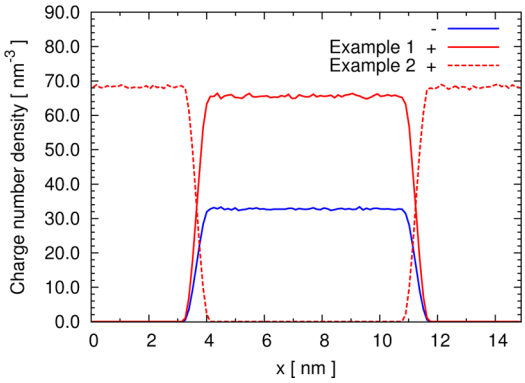

Figure 2. The charge number density of Example 1 and 2. The red solid line is the positive charge of Example 1. The red dashed line is the positive charge of Example 2. The blue line is the negative charge. Since the charges are uniformly distributed in the *y* and *z* directions, the densities are averaged on *y* and *z*, and plotted as a function of *x*.

the positive charge is twice as large as that of the negative charge, so the positive and negative charges cancel, and the system is *locally* neutral. Notice the amount of charges and the density distributions are intentionally chosen the same as in Example 3, for an easy comparison. The only difference is that the positions of charges are independent in this example, while they are correlated in Example 3.

Figure 3 presents the real rms error (by points) and the corresponding error estimates (by lines) of the truncated Ewald direct part, the truncated Ewald reciprocal part, the original *ik*-/analytical differentiation and the staggered mesh *ik*-/analytical differentiation. All the error estimates are consistently accurate compared with the real errors. The parameters are chosen to be the same for an easy comparison among the reciprocal space methods. The errors of FFT based fast Ewald methods are 8 orders of magnitudes larger than that of the truncated Ewald

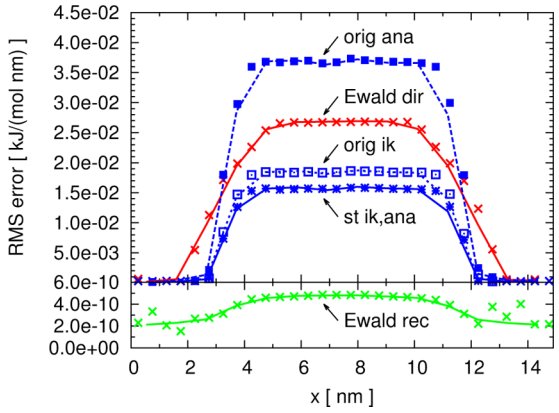

Figure 3. Example 1: the real rms errors and the corresponding estimates. Different colors denote different parts of the Ewald summation: (Red) direct part of Ewald summation; (green) reciprocal part of Ewald summation, calculated by truncation, (×) real error, (—) estimated error; (blue) reciprocal part of Ewald summation, calculated by fast algorithms, (■) real error of the original analytical differentiation, (⊡) real error of the original *ik*-differentiation, (×) real error of the staggered mesh analytical differentiation, (+) real error of the staggered mesh *ik*-differentiation. (---) estimated original analytical differentiation; (···) estimated original *ik*-differentiation. (—) estimated staggered mesh *ik*-/analytical differentiation. The rms errors are averaged over *y* and *z* directions, and are plotted against *x* axis. The cutoff in the real space is 1.31 nm, the number of freedom in the reciprocal space is 120 × 60 × 60, the parameter β is 2.5 nm^{-1}, and the order of B-spline interpolation is 6.

summation. This supports the argument addressed earlier: the error introduced by truncating the reciprocal Ewald summation is much smaller than that introduced by approximating $e^{2\pi i m \cdot r}$. The original *ik*-differentiation is more accurate than the original analytical differentiation. With the staggered mesh, the error of the analytical differentiation is reduced by more than 50%, while the accuracy of the *ik*-differentiation is only marginally improved. The errors of the staggered mesh *ik*- and analytical differentiation are identical, which supports the theoretical prediction. The original *ik*-differentiation uses 4 FFT transforms, and the original analytical differentiation uses only 2. Therefore, the staggered mesh analytical differentiation uses the same amount of FFTs as the original *ik*-differentiation, and saves one-half of the FFTs compared to the staggered mesh *ik*-differentiation. As mentioned before, the FFT is a bottleneck of communication on massive parallel supercomputers, so the methods using less FFTs may have advantages on these machines.

Example 2: Separated Positive and Negative Charge. In this example, the setting is nearly the same as Example 1. The only difference is that all positive charges are moved to the region where the density of negative charge is low, see Figure 2. In this system, the negative charges are separated from their counterions, but the whole system is kept neutral. This case rarely happens in real simulation, but it serves as a good test of the error estimates in extreme situations.

Figure 4 shows the real mean error forces (by points) and the corresponding error estimates (by lines) of the truncated Ewald direct part, truncated Ewald reciprocal part, the *ik*-/analytical differentiation and the staggered mesh *ik*-/analytical differentiation. All notations in the figure are the same as Example 1. Although the charge distribution is not usual in this case, the error estimates are still sharp. The two peaks of the direct mean error force present at the positive−negative

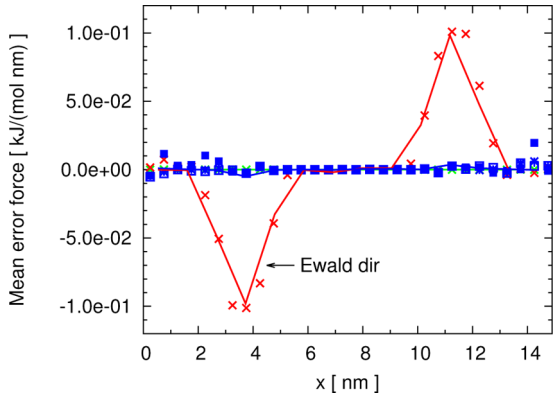

Figure 4. Example 2: the real mean error forces and the corresponding estimates. All symbols and working parameters are the same as in Figure 3. The reciprocal mean error forces vanish, so the points and lines overlap each other. The errors are averaged over *y* and *z* directions and are plotted against the *x* axis.

interface, due to the separation of the positive and negative charges, that is, a nonvanishing and fast changing first order charge distribution ρ_q. Similar phenomena were reported by the authors in ref 23. Surprisingly, the reciprocal mean error forces do not present any singularity, and vanish all over the simulation region.

Figure 5 presents the real and estimated rms error of Example 2. The the original *ik*-differentiation is not shown,

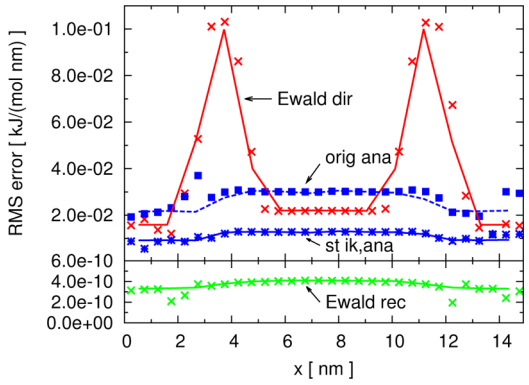

Figure 5. Example 2: the real rms error and the corresponding estimates. The symbols and working parameters used in this figure are the same as Figure 3. The real and estimated error of the original *ik*-differentiation are very closed to the staggered mesh Ewald results, so they are not shown in this figure for clarity, The errors are averaged over *y* and *z* directions, and are plotted against *x* axis.

because it is nearly indistinguishable from the staggered mesh *ik*-/analytical differentiation in this plot. All the error estimates are consistent with the real errors. Stemming from the inhomogeneity error, two peaks of the direct error form at the interfacial regions, and are much larger than the error in bulk regions. The reciprocal rms errors only contain homogeneity contributions, due to the vanished mean error force. Similar with the method of Example 1, the truncated Ewald method is much more precise than the fast algorithms, and the staggered mesh improves the accuracy of analytical differentiation by more than 50%.

Example 3: A Water System in Gas−Liquid Phase Equilibrium. This example studies the gas−liquid phase equilibrium of a water system. The size of the simulation box

dx.doi.org/10.1021/ct300343y | J. Chem. Theory Comput. 2012, 8, 3243−3256

and number of charges are the same as in Example 1: 13824 TIP3P water molecules[24] are put into a 14.90 nm × 7.45 nm × 7.45 nm periodic simulation box. The molecules dynamics simulations of this system was performed by GROMACS 4.5.[14] The system is coupled to a velocity rescale thermostat[28] at temperature 300 K, and simulated long enough to reach the equilibrium. After the equilibrium, the water molecules separate into the liquid and vapor phases. The number densities of oxygen and hydrogen atoms are the same as in Example 1, see Figure 2. There were 50 consequential snapshots of the system taken along the MD trajectory with a time interval of 20 ps. These snapshots are used to calculate the real error and the charge densities for the error estimates.

The real and estimated errors of the original SPME (both *ik*- and analytical differentiation) and the staggered mesh Ewald (both *ik*- and analytical differentiation) are in Figures 6 and 7,

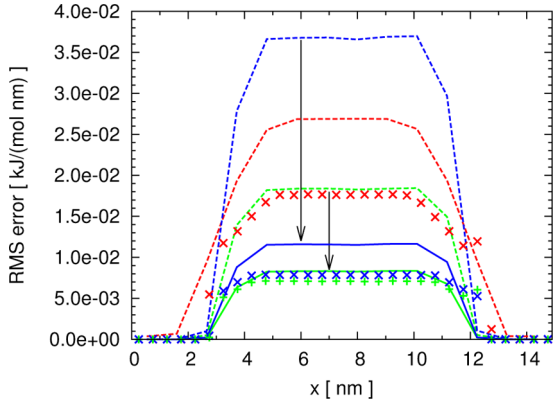

Figure 6. Example 3: the real rms errors and the corresponding estimates of original *ik*- and analytical differentiation. Red: the direct part. Green: the original *ik*-differentiation. Blue: the original analytical differentiation. " × " or "+": the real error. Dashed line: the error estimate, without the correlation error estimate. Solid line: the error estimate, with the bonded nearest neighbor approximation of the correlation error. The working parameters are the same as Example 1. The black arrows indicate the improvement of error estimate by including bonded nearest neighbor approximation of correlation error.

respectively. Unlike previous examples, the real error of *ik*- and analytical differentiation are nearly the same. The staggered mesh *ik*- and analytical differentiation are also equivalent in this case, and are ∼70% more accurate than the original *ik*-/analytical differentiation. In this case, all estimates without considering the correlation error deviate from the real errors. The quality of the direct part error estimate is still good (though not perfect): the error is overestimated by 50%. The rms errors of the *ik*-differentiation, analytical differentiation, and the staggered mesh *ik*-/analytical differentiation are overestimated by 2.6, 4.7, and 6.8 times, respectively. Notice the only difference between this example and Example 1 (in which the estimates were very sharp) is the correlation of the charges in the system, which plays a very important role in the rms error. More surprisingly, the correlation actually reduces the error in the force computation. Thanks to the reviewer of this paper, one likely explanation is that the electrostatic forces exerted by bonded pairs are anticorrelated due to opposing charge signs. The extreme case is the overlapping of one oxygen atom with two hydrogen atoms: the charges cancels, so the error vanishes. In the case of the TIP3P water model, the H−O bond is 0.1 nm, which is much shorter than the typical center-

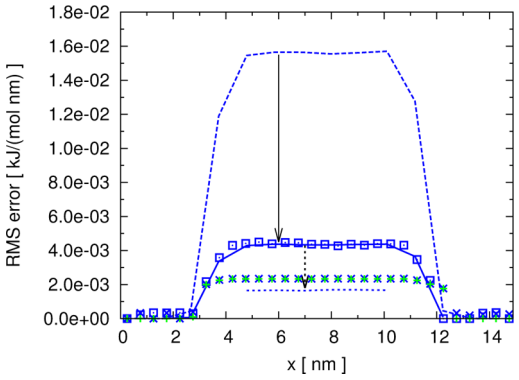

Figure 7. Example 3: the real rms errors and the corresponding estimates. (Blue ×) the real error of the staggered mesh analytical differentiation; (blue ⊡) the real error of the randomized COM test, calculated by the staggered mesh analytical differentiation; (green +) the real error of the staggered mesh *ik*-differentiation; it is overlapping with the analytical differentiation; (blue ---) the error estimate, without the correlation error; (blue —) the error estimate, with the bonded nearest neighbor approximation of the correlation error; (blue ···) the error estimate for the liquid phase including both the bonded and nonbonded nearest neighbor approximations. The rms errors are averaged over *y* and *z* directions, and are plotted against *x* axis. The working parameters are the same as Example 1. The black arrow indicates the improvement of error estimate by including bonded nearest neighbor approximation. The black dotted arrow indicates the improvement of further including nonbonded nearest neighbor approximation.

of-mass (COM) distance between two neighboring water molecules (0.31 nm), therefore, the bonded correlation tends to neutralize the molecule and reduce the force error.

By including the bonded nearest neighbor approximation of the correlation error, the quality of all error estimates is improved greatly, as indicated by the solid arrows in Figures 6 and 7. Now the rms errors of *ik*-differentiation, analytical differentiation and staggered mesh *ik*-/analytical differentiation are overestimated by a factor of 1.16, 1.44, and 1.86, respectively. Reference 22 showed that the error estimate (*ik*-differentiation in a homogeneous water system) that is even three times larger than the real error still works well in the parameter tuning process, so the error estimate for all methods are now good enough for the parameter tuning application. Since we worry about the validity of the leading order approximation in eqs 48 and 54, we consider the staggered mesh Ewald method in the following testing case (we call it the randomized COM test): randomly redistribute the water molecules according to the same COM distribution, while keeping the bonds and direction of the molecules unchanged. A perfect error estimate including the bonded nearest neighbor approximation should be very close to the real error of the randomized COM test, because there is NO nonbonded correlation in the test system. Since the only approximation made in the error estimate is the leading order approximation, the difference between the estimate and the real error of the randomized COM test comes from the higher order contributions discarded by eqs 48 and 54. In Figure 7, the overlapping of the estimate with the real error justifies the effectiveness of the leading order approximation. We will investigate this issue in a wider parameter range later. The prerequisite of the bonded nearest neighbor approximation is basically fulfilled: the water direction preference in the bulk liquid and vapor regions is uniform (actually it is isotropic).

dx.doi.org/10.1021/ct300343y | J. Chem. Theory Comput. 2012, 8, 3243−3256

Although we observe (not shown) a preference of the dipole moment along the x direction at the liquid−vapor interface, it is relatively weak and has no obvious side-effect on the error estimate.

The nonbonded nearest neighbor approximation (eqs 21 and 22) cannot be directly used in this example, because the system is inhomogeneous, and the RDFs are not well-defined. However, the system contains a comparatively homogeneous liquid region, where we can compute the RDFs and the nonbonded nearest neighbor approximation. The resulting error estimate is presented by the dotted blue line in Figure 7. (We do not consider the estimate in the vapor region because the error is much smaller than the liquid region.) The neighboring range R_g is chosen to be 1.2 nm. The error estimate converges when R_g is larger than 0.6 nm (see Figure 8). As indicated by Figure 7, the rms error estimate is further

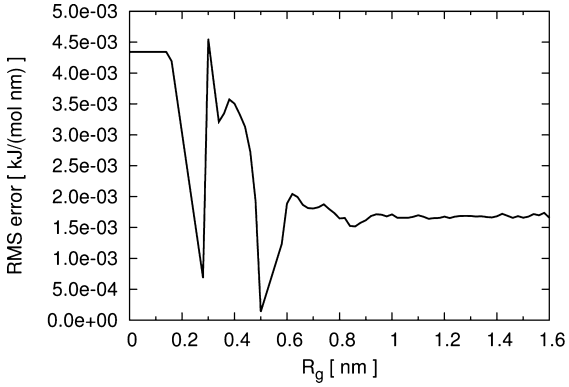

Figure 8. The convergence of the nonbonded nearest neighbor approximation with respect to the neighboring range R_g.

improved by the nonbonded nearest neighbor approximation. However, there is still a discrepancy between the real and estimated error, which we cannot explain. It may originate from assuming that the testing particle is independent with the system, but we cannot check this point, because investigating the correlation between the testing particle and the system is beyond the theoretical framework of the present paper. In Figure 7, the improvement marked by the solid arrow is much larger than that marked by the dotted arrow, so the bonded correlation dominates the correlation error, compared to the nonbonded correlation. If not stated otherwise, we only consider the bonded nearest neighbor approximation.

In Figures 9−12, we present the actual and estimated rms error of the liquid phase as a function of the Ewald splitting parameter β, using various cutoff radii, interpolation orders, and reciprocal grid spacings. Figure 9 and Figure 10 show the results of the original *ik*-/analytical differentiation, while Figure 11 and Figure 12 present the staggered mesh *ik*-/analytical differentiation. In Figures 9 and 11, the interpolation order is fixed to $n = 6$, and we plot the errors using the direct space cutoff radii r_c = 0.9, 1.1, 1.3, and 1.7 nm, and the reciprocal space grid spacings h = 0.248, 0.124, and 0.062 nm. In Figures 10 and 12, the grid spacing is fixed to h = 0.248 nm, and we plot interpolation orders of n = 4, 6, and 8. Like the homogeneous water system studied in the literature (see for example ref 7,22), when β is small the direct error dominates, when β is large the reciprocal error dominates. Therefore, given a set of parameters (r_c, n, and h), the β at the crossover of the direct and reciprocal errors achieves the nearly optimal accuracy

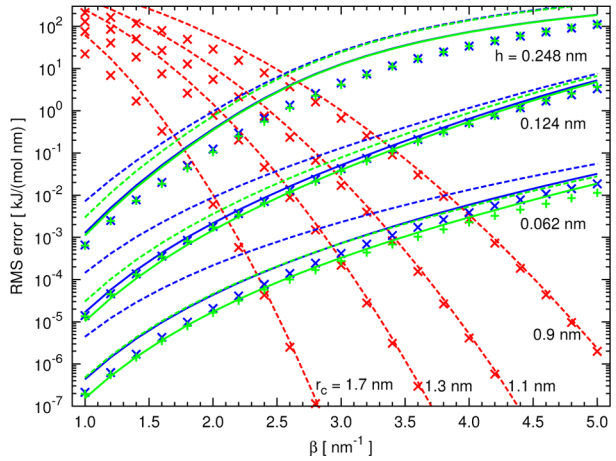

Figure 9. Example 3: the actual and estimated rms error of the liquid phase as a function of the Ewald splitting parameter β, with various cutoff radii and grid spacings. (Red) the direct part; (green) the original *ik*-differentiation; (blue) the original analytical differentiation. " × " and " + " denote the actual error. The solid and dashed lines denote the error estimates with and without bonded nearest neighbor approximation, respectively. The plotted direct space cutoff radii are r_c = 0.9, 1.1, 1.3, and 1.7 nm. The reciprocal space grid spacings h are approximately 0.248, 0.124, and 0.062 nm, corresponding to grids 60 × 30 × 30, 120 × 60 × 60, and 240 × 120 × 120, respectively. The order of the B-spline interpolation is 6.

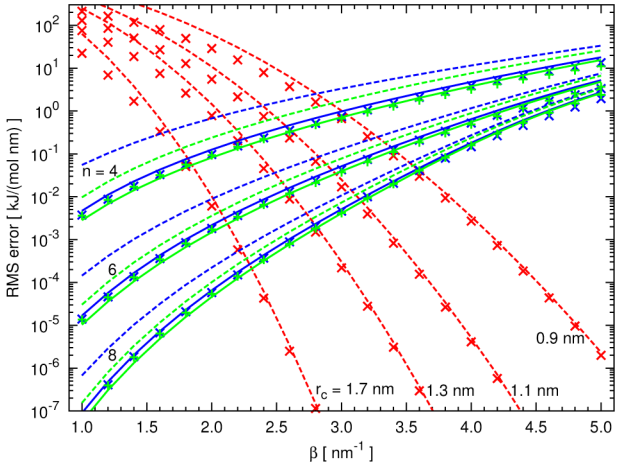

Figure 10. Example 3: the actual and estimated rms error of the liquid phase as a function of the Ewald splitting parameter β, with various cutoff radii and interpolation orders. The symbols are the same as in Figure 9. The plotted orders of the B-spline interpolation are n = 4, 6, and 8. The reciprocal space grid spacing h is approximately 0.124 nm, corresponding to a 120 × 60 × 60 grid.

and should be used in simulations. For all cutoff radii, the quality of the direct error estimate is satisfactory when the direct error is smaller than 10^{-1} kJ/(mol nm). It is also observed that the quality of the error estimate is better for larger β than for smaller β. As expected, Figures 9−12 show that the smaller grid spacings and larger interpolation orders enable more accurate force computation. Unlike the uncorrelated testing examples 1 and 2, the actual precisions of the original *ik*- and analytical differentiation are nearly the same (see Figures 9 and 10). This implies that the latter may be more preferable than the former for the massive parallel water simulations, because it needs only one-half FFTs of the former. With the bonded nearest neighbor approximation, the

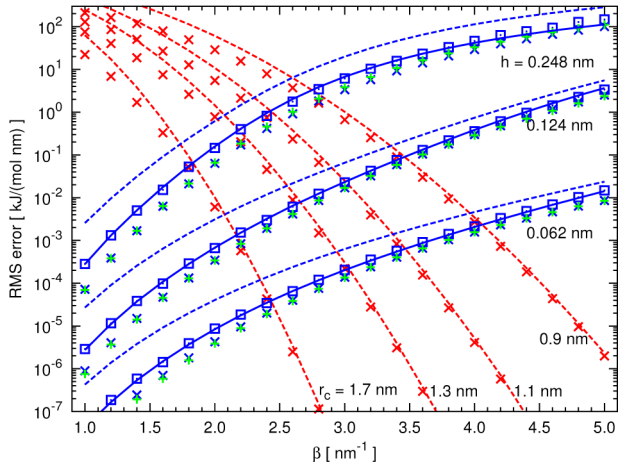

Figure 11. Example 3: the actual and estimated rms error of the liquid phase as a function of the Ewald splitting parameter β, with various cutoff radii and grid spacings. (Red) the direct part; (green) the staggered mesh *ik*-differentiation; (blue) the staggered mesh analytical differentiation. " × " and " + " denote the actual error. The solid and dashed lines denote the error estimates with and without bonded nearest neighbor approximation, respectively. The error estimates for the staggered mesh *ik*- and analytical differentiations are identical, so only the blue lines are presented. The blue "⊡" indicates the real error of the randomized COM test, calculated by the staggered mesh analytical differentiation. The same results of the *ik*-differentiation are not shown. The plotted direct space cutoff radii are r_c = 0.9, 1.1, 1.3, and 1.7 nm. The reciprocal space grid spacings h are approximately 0.248, 0.124, and 0.062 nm, corresponding to grids 60 × 30 × 30, 120 × 60 × 60, and 240 × 120 × 120, respectively. The order of the B-spline interpolation is 6.

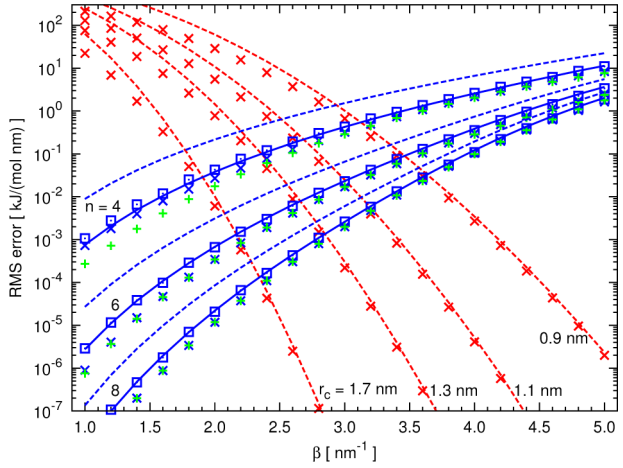

Figure 12. Example 3: the actual and estimated rms error of the liquid phase as a function of the Ewald splitting parameter β, with various cutoff radii and interpolation orders. The symbols are the same as in Figure 11. The plotted orders of the B-spline interpolation are n = 4, 6, and 8. The reciprocal space grid spacing h is approximately 0.124 *nm*, corresponding to a 120 × 60 × 60 grid.

reciprocal error estimates are impressively improved compared to those without considering the charge correlation, and are able to correctly catch the trends of real errors with various working parameters (Figures 9−12). The improvement of the original analytical differentiation and the staggered mesh Ewald method is larger than that of the original *ik*-differentiation. As discussed before, we compare the error estimate (including bonded nearest neighbor approximation) with the real error of the randomized COM test. In most cases in Figure 11 and Figure 12, a perfect overlapping is observed, which justifies the leading order approximation proposed in eqs 48 and 54. However, we do observe exceptions at large β when the grid spacing is h = 0.248 nm (see Figure 11), and at small β when the interpolation order is n = 4 (see Figure 12). The deviations indicate that the higher order terms discarded by eqs 48 and 54 may play a role in these situations. This may also explain the unexpected deviation of the staggered mesh *ik*-differentiation from the analytical differentiation at small β, n = 4.

■ CONCLUSIONS AND REMARKS

We proposed the error estimates for three state-of-art algorithms calculating the long-range electrostatic interaction in a molecular system. They were the Ewald summation, the smooth particle mesh Ewald (SPME, both *ik*- and analytical differentiation), and the staggered mesh Ewald (both *ik*- and analytical differentiation) methods. Unlike the previous error estimates, the new error estimates did NOT assume the homogeneity and uncorrelation of the system, which was a too strong assumption for most molecular simulations of practical interest. A general error estimate framework was proposed to study all the aforementioned estimates. When the error force has a pairwise form, the rms error was proved to be composed of three additive parts: the homogeneity error, the inhomogeneity error, and the correlation error. Moreover, the framework quantitatively provided the estimate of all these error components, and suggested a computationally scalable way (at the cost of $O(N \log N)$) to calculate them.

Some minor contributions of the present work are as follows: (1) In a locally neutral charge system, the inhomogeneity error was proved to vanish, therefore, no adaptive method nor force correction[23] is necessary. Fortunately, most molecular systems of practical interests are locally neutral. (2) We explicitly gave the expression of the self-interaction term in the analytical differentiation, which was proved to dominate the force error in low density systems.[10] The expression helped to remove the self-interaction term at a comparatively low computational cost. (3) The error estimate showed that the staggered mesh Ewald (both *ik*- and analytical differentiation) are always more precise than the original SPME. (4) Unlike the original analytical differentiation that violates the Newton's third law, the staggered mesh analytical differentiation preserves it to the leading order. (5) We also proved the equivalence of the staggered mesh *ik*- and analytical differentiation, to the leading order of the force error. The staggered mesh analytical differentiation requires only one-half the FFTs as the *ik*-differentiation, so the former may be preferable for massive parallel simulation, due to the bottleneck of all-to-all communication required by the parallel FFT.

The effectiveness of the proposed error estimates was verified by three numerical tests: two ideal inhomogeneous cases, in which the charges were uncorrelated, and one real water system, in which the charges were correlated. In the ideal cases, the all error estimates were sharp, even in an extreme case: positive and negative charges were globally separated. In the water system, all estimates overestimated the real error, because of the correlation of charges. The quality of the real space error estimates was acceptable. The error estimates for the reciprocal space fast algorithms were impressively improved by including the bonded first neighbor approximation to the correlation error. This indicated among all possible correlations, the rigid bond and angle correlation within one molecule is dominant.

dx.doi.org/10.1021/ct300343y | J. Chem. Theory Comput. 2012, 8, 3243−3256

This paper only focuses on the accuracy of the mentioned long-range algorithms. We do not touch the topic of comparing the efficiency of these algorithms. To be fair, we propose optimizing the working parameters for each algorithm with a predetermined accuracy before comparing the real execution speed. Such a comparison includes the topic of the hardware architecture, the communication bandwidth, the software implementation, the scalability of the problem, the inhomogeneity of the system, etc. They are far beyond the scope of the current work, and will be discussed in the following research.

ERROR ESTIMATE OF THE SMOOTH PARTICLE MESH EWALD METHOD

The *ik*-Differentiation

Starting from eq 36, we have the error force kernel of the *ik*-differentiation:

$$\mathbf{K}_{\mathrm{rec}}^{\mathrm{ik}}(\mathbf{r},\mathbf{r}') = \sum_{\substack{|m_\alpha|<K_\alpha/2\\ \mathbf{m}\neq 0}} [A(\mathbf{m},\mathbf{r}) + A(\mathbf{m},-\mathbf{r}')]\mathbf{g}(\mathbf{m}) \times \mathrm{e}^{-2\pi i\mathbf{m}\cdot(\mathbf{r}-\mathbf{r}')} \tag{59}$$

By inserting eq 34, the error force kernel is

$$\mathbf{K}_{\mathrm{rec}}^{\mathrm{ik}}(\mathbf{r},\mathbf{r}') = \sum_{\substack{|m_\alpha|<K_\alpha/2\\ \mathbf{m}\neq 0}} \sum_{\alpha=1}^{3}\sum_{l\neq 0} (\mathrm{e}^{2\pi i l K_\alpha \mathbf{a}_\alpha^*\cdot\mathbf{r}} + \mathrm{e}^{-2\pi i l K_\alpha \mathbf{a}_\alpha^*\cdot\mathbf{r}'} - 2) \times Z_{\alpha,l}(\mathbf{m})\mathbf{g}(\mathbf{m})\,\mathrm{e}^{-2\pi i\mathbf{m}\cdot(\mathbf{r}-\mathbf{r}')} \tag{60}$$

To calculate the mean error force estimate, we should calculate the integral of $\int \mathbf{K}_{\mathrm{rec}}^{ik}(\mathbf{r},\mathbf{r}')\rho_q(\mathbf{r}')\,\mathrm{d}\mathbf{r}$. By eq 60, we should calculate $\int \mathrm{e}^{-2\pi i l\mathbf{K}_\alpha\mathbf{a}_\alpha^*\cdot\mathbf{r}'}\rho_q(\mathbf{r}')\mathrm{e}^{2\pi i\mathbf{m}\cdot\mathbf{r}''}\,\mathrm{d}\mathbf{r}$, which is the inverse Fourier transform of $\mathrm{e}^{-2\pi i l K_\alpha \mathbf{a}_\alpha^*\cdot\mathbf{r}'}\rho_q(\mathbf{r}')$. Notice, for example, when $\alpha = 1$ we have

$$[\rho_q(\mathbf{r}')\mathrm{e}^{-2\pi i l K_1\mathbf{a}_1^*\cdot\mathbf{r}'}]^{\vee}(m_1,m_2,m_3) = \check{\rho}(m_1 - lK_1, m_2, m_3) \tag{61}$$

Since $|m_1| < K_1/2$ and $l \neq 0$, $\check{\rho}(m_1 - lK_1, m_2, m_3)$ is the high wavenumber (higher than $K_\alpha/2$) part of the density profile. If the density is smooth enough, this term can be reasonably neglected. Therefore,

$$\langle \Delta\mathbf{F}_{\mathrm{rec}}^{\mathrm{ik}}(\mathbf{r})\rangle = q\int_{\mathbb{R}^3} \mathbf{K}_{\mathrm{rec}}^{\mathrm{ik}}(\mathbf{r},\mathbf{r}')\rho_q(\mathbf{r}')\,\mathrm{d}\mathbf{r}' \approx q\{[\sum_\alpha\sum_{l\neq 0}(\mathrm{e}^{2\pi i l\mathbf{K}_\alpha\mathbf{a}_\alpha^*\cdot\mathbf{r}} - 2)\mathbf{G}_{\alpha,l}]*\rho_q\}(\mathbf{r}) \tag{62}$$

Further neglecting the high-wavenumber term $\mathrm{e}^{2\pi i l K_\alpha\mathbf{a}_\alpha^*\cdot\mathbf{r}}$, we reach

$$\langle \Delta\mathbf{F}_{\mathrm{rec}}^{\mathrm{ik}}(\mathbf{r})\rangle \approx -2q[(\sum_\alpha\sum_{l\neq 0}\mathbf{G}_{\alpha,l})*\rho_q](\mathbf{r}) \tag{63}$$

To calculate the rms error, we square the error force kernel:

$$|\mathbf{K}_{\mathrm{rec}}^{\mathrm{ik}}(\mathbf{r},\mathbf{r}')|^2 = \sum_{\substack{|m_\alpha|<K_\alpha/2\\ \mathbf{m}\neq 0}}\sum_{\substack{|n_\alpha|<K_\alpha/2\\ \mathbf{n}\neq 0}}\sum_{\alpha,\beta}\sum_{l_1,l_2}\mathbf{g}(\mathbf{m})\cdot \mathbf{g}(\mathbf{n})Z_{\alpha,l_1}(\mathbf{m})Z_{\beta,l_2}(\mathbf{n}) \times (\mathrm{e}^{2\pi i l_1 K_\alpha\mathbf{a}_\alpha^*\cdot\mathbf{r}} + \mathrm{e}^{-2\pi i l_1 K_\alpha\mathbf{a}_\alpha^*\cdot\mathbf{r}'} - 2) \times (\mathrm{e}^{-2\pi i l_2 K_\beta\mathbf{a}_\beta^*\cdot\mathbf{r}} + \mathrm{e}^{2\pi i l_2 K_\beta\mathbf{a}_\beta^*\cdot\mathbf{r}'} - 2) \times \mathrm{e}^{-2\pi i(\mathbf{m}-\mathbf{n})\cdot(\mathbf{r}-\mathbf{r}')} \tag{64}$$

By carefully calculating $(\mathrm{e}^{2\pi i l_1\mathbf{K}_\alpha\mathbf{a}_\alpha^*\cdot\mathbf{r}} + \mathrm{e}^{-2\pi i l_1\mathbf{K}_\alpha\mathbf{a}_\alpha^*\cdot\mathbf{r}'} - 2)(\mathrm{e}^{-2\pi i l_2\mathbf{K}_\beta\mathbf{a}_\beta^*\cdot\mathbf{r}} + \mathrm{e}^{2\pi i_2 2K_\beta\mathbf{a}_\beta^*\cdot\mathbf{r}'} - 2)$:

$$\begin{aligned}(\mathrm{e}^{2\pi i l_1\mathbf{K}_\alpha\mathbf{a}_\alpha^*\cdot\mathbf{r}} + \mathrm{e}^{-2\pi i l_1\mathbf{K}_\alpha\mathbf{a}_\alpha^*\cdot\mathbf{r}'} - 2)(\mathrm{e}^{-2\pi i l_2\mathbf{K}_\beta\mathbf{a}_\beta^*\cdot\mathbf{r}} + \mathrm{e}^{2\pi i l_2\mathbf{K}_\beta\mathbf{a}_\beta^*\cdot\mathbf{r}'} - 2) &= \mathrm{e}^{2\pi i(l_1\mathbf{K}_\alpha\mathbf{a}_\alpha^* - l_2\mathbf{K}_\beta\mathbf{a}_\beta^*)\cdot\mathbf{r}} + \mathrm{e}^{-2\pi i(l_1\mathbf{K}_\alpha\mathbf{a}_\alpha^* - l_2\mathbf{K}_\beta\mathbf{a}_\beta^*)\cdot\mathbf{r}'}\\ &- 2(\mathrm{e}^{2\pi i l_1\mathbf{K}_\alpha\mathbf{a}_\alpha^*\cdot\mathbf{r}} + \mathrm{e}^{-2\pi i l_1\mathbf{K}_\alpha\mathbf{a}_\alpha^*\cdot\mathbf{r}'})\\ &- 2(\mathrm{e}^{2\pi i l_1\mathbf{K}_\alpha\mathbf{a}_\alpha^*\cdot\mathbf{r}} + \mathrm{e}^{-2\pi i l_1\mathbf{K}_\alpha\mathbf{a}_\alpha^*\cdot\mathbf{r}'}) + 4\end{aligned} \tag{65}$$

The first two terms on the rhs of eq 65 are not high-wavenumber terms only when $l_1 = l_2$ and $\alpha = \beta$. The third and fourth terms are always high-wavenumber terms, because $l_1 \neq 0$ and $l_2 \neq 0$. Therefore, we have the error estimate for the rms homogeneity error:

$$\begin{aligned}[\mathcal{E}_{\mathrm{homo}}^{\mathrm{ik}}(\mathbf{r})]^2 &= q^2\int |\mathbf{K}_{\mathrm{rec}}^{\mathrm{ik}}(\mathbf{r},\mathbf{r}')|^2\rho_{q^2}(\mathbf{r}')\mathrm{d}\mathbf{r}'\\ &\approx 2q^2[(\sum_\alpha\sum_{l\neq 0}|\mathbf{G}_{\alpha,l}|^2)*\rho_{q^2}](\mathbf{r})\\ &+ 4q^2[|\sum_\alpha\sum_{l\neq 0}\mathbf{G}_{\alpha,l}|^2*\rho_{q^2}](\mathbf{r})\end{aligned} \tag{66}$$

It is also possible to calculate the first neighbor approximation for the *ik*-differentiation. First, we have:

$$\begin{aligned}\langle \mathbf{K}_{\mathrm{rec}}^{\mathrm{ik}}(\mathbf{r},\mathbf{r}'+\mathbf{s})\rangle &= \sum_{\substack{|m_\alpha|<K_\alpha/2\\ \mathbf{m}\neq 0}}\sum_{\alpha=1}^{3}\sum_{l\neq 0} Z_{\alpha,l}(\mathbf{m})\mathbf{g}(\mathbf{m})\\ &\times \langle(\mathrm{e}^{2\pi i l\mathbf{K}_\alpha\mathbf{a}_\alpha^*\cdot\mathbf{r}} + \mathrm{e}^{-2\pi i l\mathbf{K}_\alpha\mathbf{a}_\alpha^*\cdot(\mathbf{r}'+s)} - 2)\\ &\mathrm{e}^{-2\pi i\mathbf{m}\cdot(\mathbf{r}-\mathbf{r}'-\mathbf{s})}\rangle\\ &\approx \sum_{\substack{|m_\alpha|<K_\alpha/2\\ \mathbf{m}\neq 0}}\sum_{\alpha=1}^{3}\sum_{l\neq 0}(\mathrm{e}^{2\pi i l\mathbf{K}_\alpha\mathbf{a}_\alpha^*\cdot\mathbf{r}} - 2)\\ &Z_{\alpha,l}(\mathbf{m})\mathbf{g}(\mathbf{m})\hat{T}_s(\mathbf{m})\mathrm{e}^{-2\pi i\mathbf{m}\cdot(\mathbf{r}-\mathbf{r}')}\end{aligned} \tag{67}$$

So

$$\begin{aligned}&\mathbf{K}_{\mathrm{rec}}^{\mathrm{ik}}(\mathbf{r},\mathbf{r}')\cdot\langle\mathbf{K}_{\mathrm{rec}}^{\mathrm{ik}}(\mathbf{r},\mathbf{r}'+\mathbf{s})\rangle\\ &= \sum_{\substack{|m_\alpha|<K_\alpha/2\\ \mathbf{m}\neq 0}}\sum_{\substack{|n_\alpha|<K_\alpha/2\\ \mathbf{n}\neq 0}}\sum_{\alpha,\beta}\sum_{l_1,l_2}\mathbf{g}(\mathbf{m})\cdot\mathbf{g}(\mathbf{n})Z_{\alpha,l_1}(\mathbf{m})Z_{\beta,l_2}(\mathbf{n})\\ &\hat{T}_s(\mathbf{n}) \times (\mathrm{e}^{2\pi i l_1\mathbf{K}_\alpha\mathbf{a}_\alpha^*\cdot\mathbf{r}} + \mathrm{e}^{-2\pi i l_1\mathbf{K}_\alpha\mathbf{a}_\alpha^*\cdot\mathbf{r}'} - 2)\\ &(\mathrm{e}^{-2\pi i l_2\mathbf{K}_\beta\mathbf{a}_\beta^*\cdot\mathbf{r}} - 2) \times \mathrm{e}^{-2\pi i(\mathbf{m}-\mathbf{n})\cdot(\mathbf{r}-\mathbf{r}')}\end{aligned} \tag{68}$$

Inserting this into eq 17 while neglecting the high frequency terms yields

$$\mathcal{E}^{ik}_{correlation}(\mathbf{r}) = q^2[(\sum_{\alpha}\sum_{l\neq 0}\mathbf{G}_{\alpha,l}\cdot[\hat{T}_O\hat{\mathbf{G}}_{\alpha,l}]^{\vee})*\rho_O](\mathbf{r}) + 4q^2[(\sum_{\alpha}\sum_{l\neq 0}\mathbf{G}_{\alpha,l})\cdot(\sum_{\alpha}\sum_{l\neq 0}[\hat{T}_O\hat{\mathbf{G}}_{\alpha,l}]^{\vee})*\rho_O](\mathbf{r}) + q^2[(\sum_{\alpha}\sum_{l\neq 0}\mathbf{G}_{\alpha,l}\cdot[\hat{T}_H\hat{\mathbf{G}}_{\alpha,l}]^{\vee})*\rho_H](\mathbf{r}) + 4q^2[(\sum_{\alpha}\sum_{l\neq 0}\mathbf{G}_{\alpha,l})\cdot(\sum_{\alpha}\sum_{l\neq 0}[\hat{T}_H\hat{\mathbf{G}}_{\alpha,l}]^{\vee})*\rho_H](\mathbf{r}) \quad (69)$$

The Analytical Differentiation

Starting from eq 41, we have the error force kernel of the analytical differentiation (notice the self-interacting term is removed):

$$\mathbf{K}^{ana}_{rec}(\mathbf{r},\mathbf{r}') = \sum_{\substack{|\mathbf{m}_\alpha|<K_\alpha/2\\ \mathbf{m}\neq 0}}\{[A(\mathbf{m},\mathbf{r}) + A(\mathbf{m},-\mathbf{r}')]\mathbf{g}(\mathbf{m}) - 2\mathbf{B}(\mathbf{m},\mathbf{r})f(\mathbf{m})\}e^{-2\pi i\mathbf{m}\cdot(\mathbf{r}-\mathbf{r}')} \quad (70)$$

By inserting eqs 34 and 42, we have

$$\mathbf{K}^{ana}_{rec}(\mathbf{r},\mathbf{r}') = \sum_{\substack{|\mathbf{m}_\alpha|<\mathbf{K}_\alpha/2\\ \mathbf{m}\neq 0}}\sum_{\alpha=1}^{3}\sum_{l\neq 0}[(e^{2\pi il\mathbf{K}_\alpha\mathbf{a}^*_\alpha\cdot\mathbf{r}} + e^{-2\pi il\mathbf{K}_\alpha\mathbf{a}^*_\alpha\cdot\mathbf{r}'} - 2)\mathbf{g}(\mathbf{m}) - 4\pi il\mathbf{K}_\alpha\mathbf{a}^*_\alpha e^{2\pi il\mathbf{K}_\alpha\mathbf{a}^*_\alpha\cdot\mathbf{r}}f(\mathbf{m})] \times Z_{\alpha,l}(\mathbf{m})e^{-2\pi i\mathbf{m}\cdot(\mathbf{r}-\mathbf{r}')} \quad (71)$$

Following the same idea as the error estimate of *ik*-differentiation,

$$\langle\Delta\mathbf{F}^{ana}_{rec}(\mathbf{r})\rangle \approx -2q[(\sum_{\alpha}\sum_{l\neq 0}\mathbf{G}_{\alpha,l})*\rho_q](\mathbf{r}) \quad (72)$$

To calculate the rms error, we start from

$$|\mathbf{K}^{ana}_{rec}(\mathbf{r},\mathbf{r}')|^2 = \sum_{\substack{|\mathbf{m}_\alpha|<K_\alpha/2\\ \mathbf{m}\neq 0}}\sum_{\substack{|n_\alpha|<K_\alpha/2\\ \mathbf{n}\neq 0}}\sum_{\alpha,\beta}\sum_{l_1,l_2}[(e^{2\pi il_1\mathbf{K}_\alpha\mathbf{a}^*_\alpha\cdot\mathbf{r}} + e^{-2\pi il_1\mathbf{K}_\alpha\mathbf{a}^*_\alpha\cdot\mathbf{r}'} - 2)\mathbf{g}(\mathbf{m}) - 4\pi il_1\mathbf{K}_\alpha\mathbf{a}^*_\alpha e^{2\pi il_1\mathbf{K}_\alpha\mathbf{a}^*_\alpha\cdot\mathbf{r}}f(\mathbf{m})] \times [(e^{2\pi il_2\mathbf{K}_\beta\mathbf{a}^*_\beta\cdot\mathbf{r}} + e^{-2\pi il_2\mathbf{K}_\beta\mathbf{a}^*_\beta\cdot\mathbf{r}'} - 2)\mathbf{g}(\mathbf{n}) - 4\pi il_2\mathbf{K}_\beta\mathbf{a}^*_\beta e^{2\pi il_2\mathbf{K}_\beta\mathbf{a}^*_\beta\cdot\mathbf{r}}f(\mathbf{n})] \times Z_{\alpha,l_1}(\mathbf{m})Z_{\beta,l_2}(\mathbf{n})e^{-2\pi i(\mathbf{m}-\mathbf{n})\cdot(\mathbf{r}-\mathbf{r}')} \quad (73)$$

Similar to the calculation of eq 65, one can eventually prove the error estimate for the homogeneity rms error:

$$[\mathcal{E}^{ana}_{homo}(\mathbf{r})]^2 = q^2\int|\mathbf{K}^{ana}_{rec}(\mathbf{r},\mathbf{r}')|^2\rho_{q^2}(\mathbf{r}')d\mathbf{r}' \approx q^2[(\sum_{\alpha}\sum_{l\neq 0}|\mathbf{G}_{\alpha,l}|^2)*\rho_{q^2}](\mathbf{r}) + q^2[(\sum_{\alpha}\sum_{l\neq 0}|\mathbf{G}_{\alpha,l}+\mathbf{F}_{\alpha,l}|^2)*\rho_{q^2}](\mathbf{r}) + 4q^2[|\sum_{\alpha}\sum_{l\neq 0}\mathbf{G}_{\alpha,l}|^2*\rho_{q^2}](\mathbf{r}) \quad (74)$$

To calculate the first neighbor approximation to the correlation error,

$$\langle\mathbf{K}^{ana}_{rec}(\mathbf{r},\mathbf{r}'+\mathbf{s})\rangle = \sum_{\substack{|\mathbf{m}_\alpha|<K_\alpha/2\\ \mathbf{m}\neq 0}}\sum_{\alpha=1}^{3}\sum_{l\neq 0}\langle[(e^{2\pi il\mathbf{K}_\alpha\mathbf{a}^*_\alpha\cdot\mathbf{r}} + e^{-2\pi il\mathbf{K}_\alpha\mathbf{a}^*_\alpha\cdot(\mathbf{r}'+\mathbf{s})} - 2)\mathbf{g}(\mathbf{m}) - 4\pi il\mathbf{K}_\alpha\mathbf{a}^*_\alpha e^{2\pi il\mathbf{K}_\alpha\mathbf{a}^*_\alpha\cdot\mathbf{r}}f(\mathbf{m})]Z_{\alpha,l}(\mathbf{m}) e^{-2\pi i\mathbf{m}\cdot(\mathbf{r}-\mathbf{r}'-\mathbf{s})}\rangle \approx \sum_{\substack{|\mathbf{m}_\alpha|<K_\alpha/2\\ \mathbf{m}\neq 0}}\sum_{\alpha=1}^{3}\sum_{l\neq 0}[(e^{2\pi il\mathbf{K}_\alpha\mathbf{a}^*_\alpha\cdot\mathbf{r}} - 2)\mathbf{g}(\mathbf{m}) - 4\pi il\mathbf{K}_\alpha\mathbf{a}^*_\alpha e^{2\pi il\mathbf{K}_\alpha\mathbf{a}^*_\alpha\cdot\mathbf{r}}f(\mathbf{m})]T_s(\mathbf{m})Z_{\alpha,l}(\mathbf{m})e^{-2\pi i\mathbf{m}\cdot(\mathbf{r}-\mathbf{r}')} \quad (75)$$

Inserting $\langle K^{ana}_{rec}(\mathbf{r},\mathbf{r}'+s)\rangle$ into eq 17, we have the first neighbor approximation to the correlation error:

$$\mathcal{E}^{ana}_{correlation}(\mathbf{r}) = q^2[\{\sum_{\alpha}\sum_{l\neq 0}(\mathbf{G}_{\alpha,l}+\mathbf{F}_{\alpha,l})\cdot[\hat{T}_O(\hat{\mathbf{G}}_{\alpha,l}+\hat{\mathbf{F}}_{\alpha,l})]^{\vee}\}*\rho_O](\mathbf{r}) + 4q^2[(\sum_{\alpha}\sum_{l\neq 0}\mathbf{G}_{\alpha,l})\cdot(\sum_{\alpha}\sum_{l\neq 0}[\hat{T}_O\hat{\mathbf{G}}_{\alpha,l}]^{\vee})*\rho_O](\mathbf{r}) + q^2[\{\sum_{\alpha}\sum_{l\neq 0}(\mathbf{G}_{\alpha,l}+\mathbf{F}_{\alpha,l})\cdot[\hat{T}_H(\hat{\mathbf{G}}_{\alpha,l}+\hat{\mathbf{F}}_{\alpha,l})]^{\vee}\}*\rho_H](\mathbf{r}) + 4q^2[(\sum_{\alpha}\sum_{l\neq 0}\mathbf{G}_{\alpha,l})\cdot(\sum_{\alpha}\sum_{l\neq 0}[\hat{T}_H\hat{\mathbf{G}}_{\alpha,l}]^{\vee})*\rho_H](\mathbf{r}) \quad (76)$$

AUTHOR INFORMATION

Corresponding Author

*E-mail: han.wang@fu-berlin.de.

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Ewald, P. P. *Ann. Phys.* **1921**, *369*, 253–287.
(2) Perram, J.; Petersen, H.; De Leeuw, S. *Mol. Phys.* **1988**, *65*, 875–893.
(3) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
(4) Essmann, U.; Perera, L.; Berkowitz, M.; Darden, T.; Lee, H.; Pedersen, L. *J. Chem. Phys.* **1995**, *103*, 8577.
(5) Hockney, R. W.; Eastwood, J. W. *Computer Simulation Using Particles*; IOP: London, 1988.
(6) Deserno, M.; Holm, C. *J. Chem. Phys.* **1998**, *109*, 7678.
(7) Ballenegger, V.; Cerdà, J.; Holm, C. *J. Chem. Theory Comput.* **2012**, *8*, 936–947.

(8) Chen, L.; Bruce Langdon, A.; Birdsall, C. K. *J. Comput. Phys.* **1974**, *14*, 200−222.
(9) Eastwood, J. W. In *Computational Methods in Classical and Quantum Physics*; Hooper, M. B., Ed.; Advance Publications: London, 1976.
(10) Cerutti, D.; Duke, R.; Darden, T.; Lybrand, T. *J. Chem. Theory Comput.* **2009**, *5*, 2322−2338.
(11) Neelov, A.; Holm, C. *J. Chem. Phys.* **2010**, *132*, 234103.
(12) Case, D.; Cheatham, T., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K., Jr; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. *J. Comput. Chem.* **2005**, *26*, 1668−1688.
(13) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.; Berendsen, H. *J. Comput. Chem.* **2005**, *26*, 1701−1718.
(14) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(15) Phillips, J.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(16) Abraham, M. J.; Gready, J. E. *J. Comput. Chem.* **2011**, *32*, 2031−2040.
(17) Kolafa, J.; Perram, J. *Mol. Simul.* **1992**, *9*, 351−368.
(18) Hummer, G. *Chem. Phys. Lett.* **1995**, *235*, 297−302.
(19) Petersen, H. *The J. Chem. Phys.* **1995**, *103*, 3668.
(20) Deserno, M.; Holm, C. *J. Chem. Phys.* **1998**, *109*, 7694.
(21) Stern, H.; Calkins, K. *J. Chem. Phys.* **2008**, *128*, 214106.
(22) Wang, H.; Dommert, F.; Holm, C. *J. Chem. Phys.* **2010**, *133*, 034117.
(23) Wang, H.; Schüette, C.; Zhang, P. *Phys. Rev. E* **2012**, *86*, 026704.
(24) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. *J. Chem. Phys.* **1983**, *79*, 926.
(25) Frenkel, D.; Smit, B. *Understanding Molecular Simulation*, 2nd ed.; Academic Press: San Diego, CA, 2002.
(26) Schoenberg, I. J. *Cardinal Spline Interpolation*; Society for Industrial and Applied Mathematics: Philadelphia, PA, 1973.
(27) Ballenegger, V.; Cerdà, J.; Holm, C. *Comput. Phys. Commun.* **2011**, *182*, 1919−1923.
(28) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.