Harmonic surface mapping algorithm for fast electrostatic sums

Qiyuan Zhao, Jiuyang Liang, and Zhenli Xu

Harmonic surface mapping algorithm for fast electrostatic sums

Qiyuan Zhao,[1] Jiuyang Liang,[2] and Zhenli Xu[3,a)]
[1]*Zhiyuan College, Shanghai Jiao Tong University, Shanghai 200240, China*
[2]*School of Mathematical Sciences, Shanghai Jiao Tong University, Shanghai 200240, China*
[3]*School of Mathematical Sciences, Institute of Natural Sciences and MoE Key Lab of Scientific and Engineering Computing, Shanghai Jiao Tong University, Shanghai 200240, China*

We propose a harmonic surface mapping algorithm (HSMA) for electrostatic pairwise sums of an infinite number of image charges. The images are induced by point sources within a box due to a specific boundary condition which can be non-periodic. The HSMA first introduces an auxiliary surface such that the contribution of images outside the surface can be approximated by the least-squares method using spherical harmonics as basis functions. The so-called harmonic surface mapping is the procedure to transform the approximate solution into a surface charge and a surface dipole over the auxiliary surface, which becomes point images by using numerical integration. The mapping procedure is independent of the number of the sources and is considered to have a low complexity. The electrostatic interactions are then among those charges within the surface and at the integration points, which are all the forms of Coulomb potential and can be accelerated straightforwardly by the fast multipole method to achieve linear scaling. Numerical calculations of the Madelung constant of a crystalline lattice, electrostatic energy of ions in a metallic cavity, and the time performance for large-scale systems show that the HSMA is accurate and fast, and thus is attractive for many applications. *Published by AIP Publishing.* https://doi.org/10.1063/1.5044438

I. INTRODUCTION

Calculations of electrostatic interactions are among the most important components in molecular simulations of systems at the micro/nano scale such as biomolecules, membranes, electrochemical energy devices, and soft materials.[1–3] Due to the long-range nature, fast algorithms have to be used to speed up the pairwise interactions within a finite simulation volume with a specified boundary condition. To mimic the system environment, a periodic boundary condition (PBC) is often used, and mostly, Ewald-based lattice summations such particle mesh Ewald[4–6] or particle-particle particle-mesh method[7–9] have been employed for the electrostatic interaction. When the simulation systems are nonperiodic such as those with interfaces, the 3D Ewald-based techniques can be time-consuming since a larger simulation volume has to be used to reduce the artifact due to the PBC assumption, and the development of new techniques such as Ewald-type methods for quasi-2D systems[10–14] and non-Ewald methods[15–17] remains the crucial theme for simulating electrostatic phenomena.

In many problems, electrostatic interactions in a box with specific boundary conditions can be represented as the sum of an infinite number of charged particles using the method of images and recursive reflections between boundary faces.[18] The objective of this paper is to develop a harmonic surface mapping algorithm (HSMA) to transform the infinite sum into a finite one such that the accelerating techniques such as the fast multipole method (FMM)[19–22] can be simply applied to achieve an $O(N)$ complexity. The HSMA removes the difficulty of solving boundary integral equations by introducing an auxiliary surface away from the central box. The idea of the auxiliary surface has been used in scattering problems,[23] multiphase flows,[24] and electrostatics,[25,26] which is successful because it allows that the nearest-neighbor interactions within the surface is summed directly and the distant interactions can be approximated by a small number of basis functions using least-squares fittings. The harmonic surface mapping developed in this work maps the contribution of the distant interactions into a surface integral such that it can be approximated by discrete images on surfaces with a high order of convergence due to the use of the Fibonacci integration. Essentially, the HSMA can belong to the method of fundamental solutions,[27,28] and dramatically all approximate fundamental functions (the images) of the HSMA are located on the auxiliary surfaces and their strengths are obtained by the harmonic surface mapping and the numerical integration, and thus the convergence can be then ensured. Numerical examples are performed to show the efficiency of the HSMA which is accelerated by both the FMM and graphics processing units (GPUs).

The HSMA is potentially useful for many electrostatic problems in the presence of non-periodic boundaries. For instance, for quasi-2D systems with ions confined by parallel dielectric/metallic media, the images within the auxiliary surface are produced by a combination of the mirror reflections and the periodic extension, for which the direct summation

---

a)Electronic mail: xuzl@sjtu.edu.cn

of the infinite reflective images can be costly. Moreover, the HSMA can be useful for systems with an irregular boundary once the images within the auxiliary surface can be well approximated. We show an example to demonstrate that the algorithm can accurately calculate the electrostatic interaction of ions within metallic boundaries, which is considered difficult due to the divergence of the traditional image-reflection method.

II. METHOD

Consider a charged system of N point sources located at $\{\mathbf{r}_j, j = 1, \ldots, N\}$ in a cubic domain Ω of side length L. Let $\Phi(\mathbf{r})$ be the electric potential distribution due to these point sources and a specific boundary condition. Within Ω, the potential satisfies the Poisson's equation,

$$-\nabla^2 \Phi(\mathbf{r}) = 4\pi \sum_{j=1}^{N} q_j \delta(\mathbf{r} - \mathbf{r}_j),\tag{1}$$

and in many situations, the solution can be expressed as an infinite sum,

$$\Phi(\mathbf{r}) = \sum_{j=1}^{N} \frac{q_j}{|\mathbf{r}_j - \mathbf{r}|} + \sum_{i=N+1}^{\infty} \frac{q_i}{|\mathbf{r}_i - \mathbf{r}|},\tag{2}$$

where the first term is the direct potential of the point sources and the second term describes the contribution of the infinite number of image charges which are introduced to satisfy the boundary condition. For instance, if the boundary condition is periodic, one has,

$$\Phi(\mathbf{r}) = \sum_{\mathbf{k}} \sum_{j=1}^{N} \frac{q_j}{|\mathbf{r}_j + \mathbf{k}L - \mathbf{r}|},\tag{3}$$

where $\mathbf{k}$ runs over all three-dimensional integer vectors and the images are periodic copies of the source charges; if the boundary is a dielectric interface, i.e., the boundary conditions become the continuities of the potential and electric displacement, then one gets,[29]

$$\Phi(\mathbf{r}) = \sum_{\mathbf{k}} \sum_{j=1}^{N} \frac{(\gamma)^{|\mathbf{k}|} q_j}{|\mathbf{r}_{j,\mathbf{k}} + \mathbf{k}L - \mathbf{r}|},\tag{4}$$

where $\gamma = (1 - \varepsilon)/(1 + \varepsilon)$ with ε being the dielectric ratio between the exterior and the interior of the simulation box, $|\mathbf{k}|$ represents the sum of absolute values of all components, and $\mathbf{r}_{j,\mathbf{k}} = (-1)^{\mathbf{k}} \circ \mathbf{r}_j$ uses the element-by-element power and product, and the images are constructed by iterative reflections of the sources due to the six boundary faces. Similar sums can be found when the boundary condition is the combination of different types of boundary conditions such as periodic, Dirichlet, Neumann, Robin, and dielectric-jump conditions. These infinite sums are often slowly convergent (dielectric boundaries), conditionally convergent (periodic boundaries), or even divergent (conducting limit $\gamma \to -1$ of dielectric boundaries), and the direct pairwise summation for the energy and force calculation is often difficult. In practical simulations, Ewald-based lattice-summation methods[4,5] are often employed for an $O(N \log N)$ calculations through the use of FFT

acceleration. These methods are efficient for periodic systems, but less efficient for nonperiodic systems considering a large buffer zone has to be introduced to remove the artifact of periodicity. Alternatively, non-Ewald methods such as reaction-field methods[30–33] have been studied in the literature for calculating the long-range interaction. For the Poisson's equation or the modified Helmholtz equation in 2D, the method of images has been used to solve the boundary-value problem such that the adaptive FMM can be used to calculate the volume integral.[34,35]

In this work, we introduce the HSMA for the summation of infinite charges. It is composed of two steps: one approximates the infinite sum outside a given surface by a harmonic series, followed by a harmonic mapping from this series onto a surface integral which can be discretized into contributions of point charges and thus the FMM or GPU acceleration can be directly applied. We describe the details of the HSMA in the following content.

A. Spherical harmonics approximation

Consider the pairwise sum of the infinite particle system given in Eq. (2). As schematically shown in Fig. 1, we introduce an artificial boundary $\partial\Omega_s$ of the spherical shape to separate the infinite system such that domain Ω_s includes Ω and they share the same center at the origin. The spherical radius is R_s. Let $\{q_j, j = 1, \ldots, N_s\}$ be the set of those charges within the sphere, where the first N charges are the sources and the remaining $(N_s - N)$ charges are the images near the central box. The electric potential can be then rewritten into the sum of two contributions, $\Phi(\mathbf{r}) = \Phi_{\text{in}} + \Phi_{\text{out}}$, where

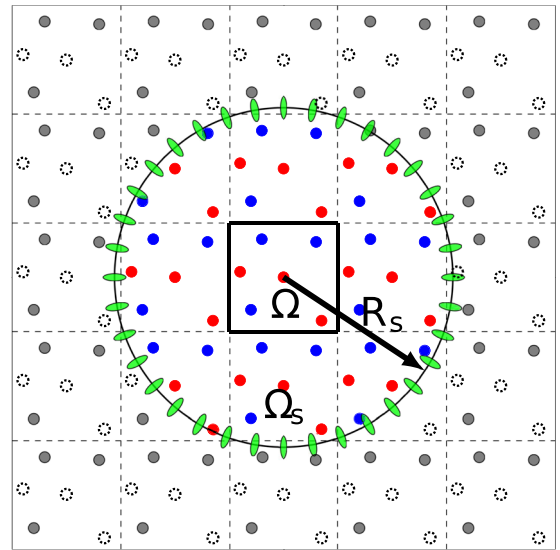$$\Phi_{\text{in}} = \sum_{j=1}^{N_s} \frac{q_j}{|\mathbf{r}_j - \mathbf{r}|},\tag{5}$$

FIG. 1. Schematic illustration of the HSMA for an infinite particle system, where Ω is the central box and particles within it are the sources. $\partial\Omega_s$ is an auxiliary boundary. The contribution of infinite images outside $\partial\Omega_s$ is approximated by point and dipolar images (elliptical symbols) on the surface.

and Φ_{out} is the electric potential produced by all images outside $\partial\Omega_s$.

Since Φ_{out} is a harmonic function in the domain Ω, it can be approximated by a truncated spherical harmonic series[25] using the spherical coordinates,

$$\Phi_{\text{out}} = \sum_{n=0}^{P} \sum_{m=-n}^{n} A_{nm} B_n^m(r, \vartheta, \varphi), \qquad (6)$$

where $B_n^m = r^n Y_n^m(\vartheta, \varphi)$ are orthogonal bases and Y_n^m is the spherical harmonic function of degree n and order m. The number of the basis functions is $N_b = (P+1)^2$.

To determine the coefficients $\{A_{nm}\}$, one shall select N_m monitoring points, $\{\mathbf{x}_i, i = 1, \ldots, N_m\}$, which are nearly uniformly distributed on the boundary, and find the coefficients such that the boundary condition (nonperiodic) is satisfied at these points. Let $\mathcal{L}(\Phi, \partial_{\mathbf{n}}\Phi) = 0$ for $\mathbf{r} \in \partial\Omega$ be the boundary condition for the Poisson's equation, where $\mathcal{L}$ is a linear operator and Φ is the sum of Φ_{in} and Φ_{out}. The N_m-dimensional residual vector $\mathbf{v} = (v_1, \ldots, v_{N_m})^T$ is then defined by these monitoring points such that $v_i = \mathcal{L}(\Phi(\mathbf{x}_i), \partial_{\mathbf{n}}\Phi(\mathbf{x}_i))$. The coefficients of the truncated spherical harmonic series are calculated by minimizing the L_2 norm of the residual vector,

$$\{A_{nm}\} = \text{argmin}\|\mathbf{v}\|_2. \qquad (7)$$

Because the number of monitoring points is usually bigger than the number of basis functions, $N_m > N_b$, the minimization is generally done by using the discrete least-squares method, resulting in an $N_b \times N_b$ linear system. The contribution of Φ_{in} appears in the right-hand side of the system and the complexity to calculate it is $O(N_s N_m)$ without the acceleration or $O(N_s + N_m)$ with the FMM acceleration.

The PBC is different from other boundary conditions, as the distribution of monitoring points on $\partial\Omega$ would lead to an ill-conditioned fitting matrix. To avoid this problem, one should choose the monitoring points $\{\mathbf{x}_i\}$ on the circumsphere of the central box Ω. Let $\widetilde{\mathbf{x}}_i = \mathbf{x}_i - \mathbf{k}L$ with $\mathbf{k}$ being the index vector such that $\widetilde{\mathbf{x}}_i$ is located within the central box. Then the ith component of the residual vector is then given by $v_i = \Phi(\mathbf{x}_i) - \Phi(\widetilde{\mathbf{x}}_i)$. The minimization for the L_2 norm of the residual vector is then better conditioned. The condition number can be further reduced when the monitoring points are uniformly distributed over the circumsphere. In this work, we use the Fibonacci grid,[36] which arranges the monitoring points along a spiral lattice such that they are close to a uniform distribution on the spherical surface, namely, the polar and azimuth angles of these points are $\vartheta_i = \arcsin((2i-1)/N_m - 1)$ and $\varphi_i = 2i\pi\omega$ for $i = 1, \ldots, N_m$, with $\omega = (\sqrt{5}-1)/2$ being the golden ratio.

Let α be the ratio between R_s and the radius of the circumsphere of the central box. An error bound of the spherical harmonic expansion for the electric potential truncated at $n = P$ is given by[25]

$$\epsilon_{\text{potl}} \leq \frac{4\pi d_0 R_s^2 \max |q_i|}{(\alpha-1)(P-2)} \left(\frac{1}{\alpha}\right)^P, \qquad (8)$$

where d_0 is a constant which is close to the number density of the particles. Since the force on each particle is used in molecular dynamics simulations, it is important to estimate the error of the gradient of the potential. Let $\epsilon_{\text{grad}}^{\ell}$ be the ℓth component of the three-dimensional error vector, then the error bound of the potential gradient is,

$$\epsilon_{\text{grad}}^{\ell} \leq \frac{4\pi d_0 R_s \max |q_i|}{(\alpha-1)(P-2)} \left(P + \frac{\alpha}{\alpha-1}\right) \left(\frac{1}{\alpha}\right)^{P-1}, \qquad (9)$$

for $\ell = 1, 2, 3$. See the Appendix for the proof. Due to the spectral convergence with the degree of the harmonics, it is expected that a small value of P can provide an accurate approximation of the solution.

B. Harmonic surface mapping

We use the Green's second identity to transform the spherical harmonic expansion of Φ_{out} into a surface integral over $\partial\Omega_s$. This will further speed up the calculation because the approximation to the surface integral leads to the sum of point images, and state-of-the-art accelerating techniques can be simply used. Essentially, the so-called harmonic surface mapping is the relation between the spherical harmonic bases and the fundamental solution bases. It avoids the solution of ill-conditioned linear systems by directly using the images on surfaces for the least-squares fitting since it benefits from the orthogonality of spherical harmonics.

Let $G(\mathbf{r}, \mathbf{r}') = 1/|\mathbf{r} - \mathbf{r}'|$ be the free-space Green's function. Since Φ_{out} satisfies the Laplace equation $\nabla^2 \Phi_{\text{out}} = 0$ for $\mathbf{r} \in \Omega$, using the Green's second identity leads us to

$$\Phi_{\text{out}} = \frac{1}{4\pi} \int_{\partial\Omega_s} \left[G \frac{\partial\Phi_{\text{out}}(\mathbf{r}')}{\partial\mathbf{n}'} - \Phi_{\text{out}}(\mathbf{r}') \frac{\partial G}{\partial\mathbf{n}'} \right] dS', \qquad (10)$$

where $\mathbf{n}'$ represents the unit outer normal direction at $\mathbf{r}'$. Equation (10) describes that the potential Φ_{out} is a sum of a surface charge and a surface dipole, which are analytically given by the spherical harmonic series. Defining $\sigma(\mathbf{r}') = \partial_{\mathbf{n}'}\Phi_{\text{out}}(\mathbf{r}')$ and using Eq. (6), one has

$$\sigma(\mathbf{r}') = \frac{1}{r'} \sum_{n=0}^{P} \sum_{m=-n}^{n} n A_n^m B_n^m(r', \vartheta', \varphi'). \qquad (11)$$

In order to represent the surface dipole in Eq. (10) in the form of the direct Coulomb potential and thus the HSMA is more efficient, the surface dipole is approximated by two surface charges. The normal derivative of the Green's function in Eq. (10) is approximated by the central difference,

$$\frac{\partial G}{\partial\mathbf{n}'} = \frac{1}{\Delta r} [G(\mathbf{r}^+, \mathbf{r}) - G(\mathbf{r}^-, \mathbf{r})], \qquad (12)$$

with $\mathbf{r}^{\pm} = (r' \pm \Delta r/2, \vartheta', \varphi')$. Let $\Omega_s^{\pm}$ represent the spheres of radii $R_s \pm \Delta r/2$, which are concentric with Ω_s, and let us define

$$\mu^{\pm}(\mathbf{r}') = \frac{R_s^2}{\Delta r (R_s \pm \Delta r/2)^2} \Phi_{\text{out}}(\mathbf{r}'). \qquad (13)$$

We can then write Φ_{out} as the sum of the following three integrals:

$$\Phi_{\text{out}}(\mathbf{r}) = \frac{1}{4\pi} \left(\int_{\partial\Omega_s} \frac{\sigma(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dS' + \int_{\partial\Omega_s^-} \frac{\mu^-(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dS' \right.$$
$$\left. - \int_{\partial\Omega_s^+} \frac{\mu^+(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dS' \right), \qquad (14)$$

where the charge densities on the three surfaces are given analytically. The last two terms are from the approximation of the surface dipole, which is an $O(\Delta r^2)$ approximation and thus is accurate when Δr is small.

The surface integrals are not singular and can be approximated by traditional numerical quadratures. One difficulty for integral over a spherical surface is the appropriate distribution of sampling points as the crystal grids will include defects leading to the loss of accuracy if uniform weights are applied. Fibonacci numerical integration[37] is introduced to optimize the approximation of the integrals, which achieves the order of accuracy $\sim N_o^{-6}$ for N_o grid points. Let F_1 and F_2 be two successive Fibonacci numbers with $F_1 < F_2$, and $f(\mathbf{r}')$ be the integrand. The Fibonacci numerical integral is written as

$$\int_{\partial\Omega_s} f(\mathbf{r}')dS' \approx \frac{2\pi R_s^2}{F_2} \sum_{j=0}^{F_2} [1 + \cos(\pi z_j)][f(\mathbf{r}_{2j+1}) + f(\mathbf{r}_{2j+2})], \tag{15}$$

where $z_j = (-1 + 2j/F_2)$, $\mathbf{r}_{2j+1} = (R_s, \vartheta_j, \varphi_j)$, $\mathbf{r}_{2j+2} = (R_s, \vartheta_j, \pi + \varphi_j)$, $\vartheta_j = \arccos(z_j + \sin(\pi z_j)/\pi)$, and $\varphi_j = \pi j F_1/F_2$. Using this quadrature to the three integrals in Eq. (14), we transform the contribution of infinite images outside $\partial\Omega_s$ into finite images on Ω_s and $\Omega_s^{\pm}$,

$$\Phi_{\text{out}}(\mathbf{r}) = \sum_{j=1}^{N_o} \frac{Q_j}{|\mathbf{R}_j - \mathbf{r}|}, \tag{16}$$

where $\mathbf{Q}_j$ and $\mathbf{R}_j$ are the charge and location of the jth one of a total of $N_o = 6(F_2 + 1)$ images.

The numerical integrations over the surfaces are accurate as long as the integration points are not small due to the sixth order of convergence, and thus the dominant error for the mapping is due to the finite-difference approximation of the surface dipole. The truncation error for the approximation given in Eq. (12) can be expressed as $\Delta r^2 \partial_r^3 G(\mathbf{r}, \mathbf{r}')/24$ for $r' \in [R_s - \Delta r/2, R_s + \Delta r/2]$, which has a small prefactor of Δr^2 in the case that α is not close to one.

C. Algorithm steps and complexity

We describe the details of the HSMA which is composed of the preparation and update steps.

The preparation step generates monitoring points $\{\mathbf{x}_i, i = 1, \ldots, N_m\}$ and integration points $\{\mathbf{R}_j, j = 1, \ldots, N_o\}$ and constructs the fitting and mapping matrices. The monitoring points are distributed either on the circumsphere of the central box Ω for a periodic boundary condition or on $\partial\Omega$ for other boundary conditions. The fitting matrix does not depend on the source charges, and the QR factorization for the least squares is performed in this step. The mapping matrix calculates the spherical harmonic expansion at the integration points and its (ℓ, k) entry represents the value of the kth harmonic basis function at point $\mathbf{R}_\ell$. In the calculation, the complex basis functions $B_n^{\pm|m|}(\mathbf{r})$ can be redefined into two real functions with the real and imaginary parts and a recursive process is introduced to calculate these functions.

In the update step, images within Ω_s are first generated. The least-square problem is solved to determine the

TABLE I. Complexity of the HSMA using the FMM.

Preparation step	Complexity
Construct the fitting matrix	$O(N_m N_b)$
Factorize the fitting matrix	$O(N_b^3)$
Construct the mapping matrix	$O(N_o N_b)$

Update step	Complexity
Generate images in Ω_s	$O(N_s)$
Solve the least-squares problems	$O(N_s + N_m + N_b N_m)$
Assign images on surfaces	$O(N_b N_o)$
Calculate the forces and energies	$O(N_s + N_o)$

coefficients $\{A_{nm}, n = 0, \ldots, P \text{ and } m = -n, \ldots, n\}$ of the spherical harmonic expansion, followed by the calculation of the charges of the images $\{Q_j, j = 0, \ldots, N_o\}$ at the integration points. The electric energy and forces of each source can be determined by summing up the contribution from both the charges $\{q_j, j = 1, \ldots, N_s\}$ and $\{Q_j, j = 1, \ldots, N_o\}$.

Since the number of sources N is generally much bigger than the number of monitoring points N_m, the number of bases N_b and the number of integration points N_o, and the most time-consuming calculations are the determination of the potential due to the N_s charges within sphere Ω_s. Both FMM and GPU accelerations are applied to speed up the calculations. The complexity of each step of the algorithm with the FMM acceleration is summarized in Table I. By using the FMM, the complexity of calculating the right-hand side of the linear least-squares system is reduced to $O(N_s + N_m)$, and the solution of the system is $O(N_b N_m)$ as the QR factorization has been done in the preparation step. Moreover, the use of FMM reduces the complexity of the force and energy calculation to $O(N_s + N_o)$. The GPU acceleration, which has been widely used in different electrostatics algorithms,[38–41] is also implemented for our computer program. It does not reduce the asymptotic scaling, but it is usually much faster than the single-core FMM when the number of sources is no more than one million.

III. RESULTS

The performance of the HSMA is tested by three examples. The computer program of the algorithm can be found on GitHub.[42] In the calculations, the central box is cubic. The radius of $\partial\Omega_s$ takes $R_s = \sqrt{3}L\alpha/2$ where L is the edge length of the central box and $\alpha > 1$ is a parameter to determine the size of the auxiliary surface. The number of monitoring points is fixed to be $N_m = 2P^2$ which approximates $2N_b$. The grid size for the central difference takes $\Delta r = 10^{-5}R_s$ if without additional description.

In the first example, the Madelung constant of a NaCl cubic-like crystalline lattice is calculated. The Madelung constant is used to determine the electrostatic energy of an ion in a crystal. In the setup, 64 unit source charges are evenly arranged on a lattice grid of edge length $L = 2$, with neighboring charges having opposite charge amount. The PBC is specified. The numerical error is computed by comparing the exact value of the Madelung constant $-1.747\,564\,594\,633\,182\,19$.[43]
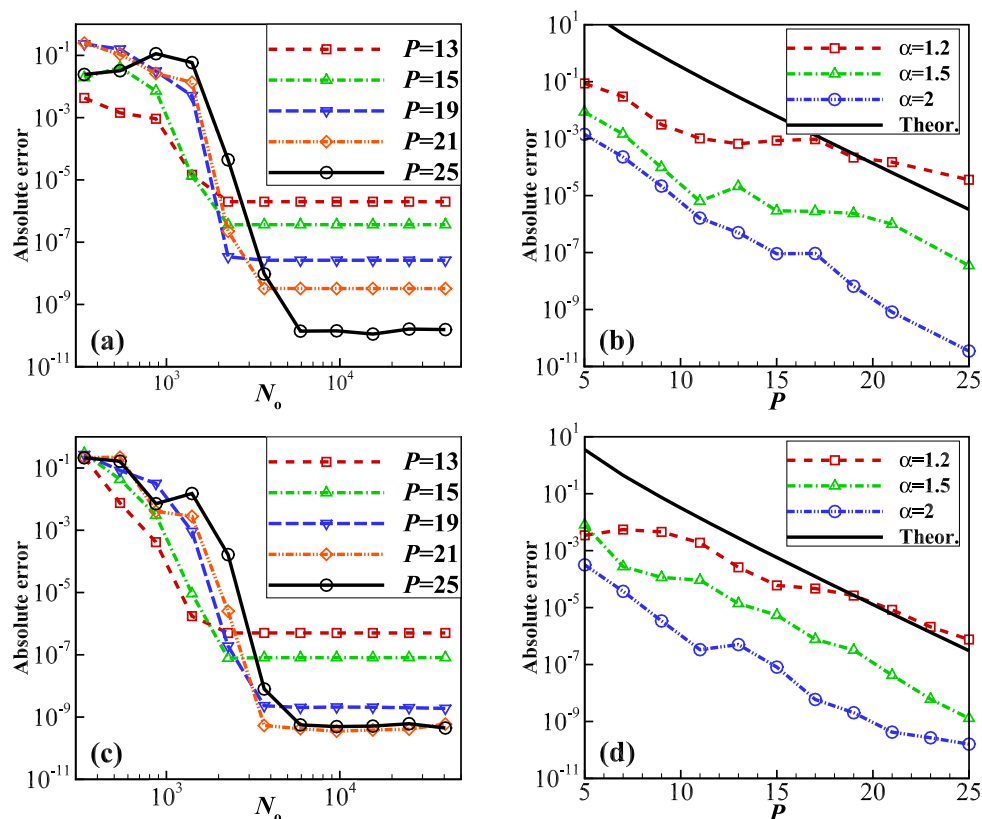
FIG. 2. Errors in the potential of the HSMA as functions of the number of the integration points for different P and the number of the spherical harmonic bases for different α. The theoretical lines correspond to parameter $\alpha = 2$. [(a) and (b)] Periodic boundary condition; [(c) and (d)] Dirichlet boundary condition.

Figure 2(a) shows the error convergence with the number of integration points N_o, where $\alpha = 2$ is taken and the results of five different P are displayed. All the curves reach platforms when $N_o > 3600$, i.e., the Fibonacci number in the numerical integration $F_2 \geq 610$. The platform in the panel means that the dominate error source is from other parts of approximation, demonstrating the rapid convergence of the numerical integration. Figure 2(b) shows the error convergence with the increase of the truncated degree of the spherical harmonic bases P, where $N_o = 5928$, i.e., the Fibonacci number $F_2 = 987$, is fixed and the results of three different α are illustrated, together with the theoretical estimate with $\alpha = 2$ (d_0 takes the average number density in the central box). It is observed that the rate of convergence is in agreement with the theoretical estimate and the actual error is much smaller than that of the estimate because the charge neutrality leads to the error cancellation.

In the second example, we take $L = 2$ as before but specify a Dirichlet boundary condition (DBC) on the boundary of the central box which includes three source points: a +2 charge at $(0.1, 0, 0)$ and two -1 charges at $(0.8, 0.8, 0.8)$ and $(-0.9, -0.9, -0.9)$ in Cartesian coordinates. The same accuracy test as the first example is made and the results are shown in Figs. 2(c) and 2(d). In the results, the error in the total electrostatic energy is calculated, where the reference "exact" solution is obtained by the HSMA method but setting $\alpha = 10$ and $P = 30$. The image series given by Eq. (4) at the conducting limit is divergent and we take $\Delta r = 10^{-4} R_s$ to avoid the influence of round-off error to

the accuracy. Similar performance as the PBC can be observed from the two panels.

In molecular dynamics simulations, the gradient of the potential is calculated to obtain the force of each charged particle. It is important to validate the accuracy of the force calculation of the HSMA. We use the same systems with the PBC and the DBC as the previous two examples and calculate the forces on each charge for varying parameters. We calculate the maximum absolute value of three force components of all the source charges, and compare it with the "exact" reference solution for the error. The reference solution is calculated using the force balance condition for the PBC and using the HSMA with high-accurate parameters ($\alpha = 10$ and $P = 30$) for the DBC. The results are present in Fig. 3. Similar performance on the error convergence as the potential calculation can be observed except that the error values are about an order of magnitude higher than those of the potential. Figures 3(b) and 3(d) also shows the agreement on the convergence rate between the numerical error and the theoretical bound in the case of $\alpha = 2$.

Next consider systems with the increase of the number of source charges N in the central box. These charges are initially distributed at the sites of a uniformly spaced lattice, each with unit charge of a random sign, and the charge neutrality is remained during the initialization. The PBC is used and the energy by the Ewald summation method[44] is calculated as the "exact" reference solution. The GPU acceleration for the pairwise summation is used for solving the linear-squares
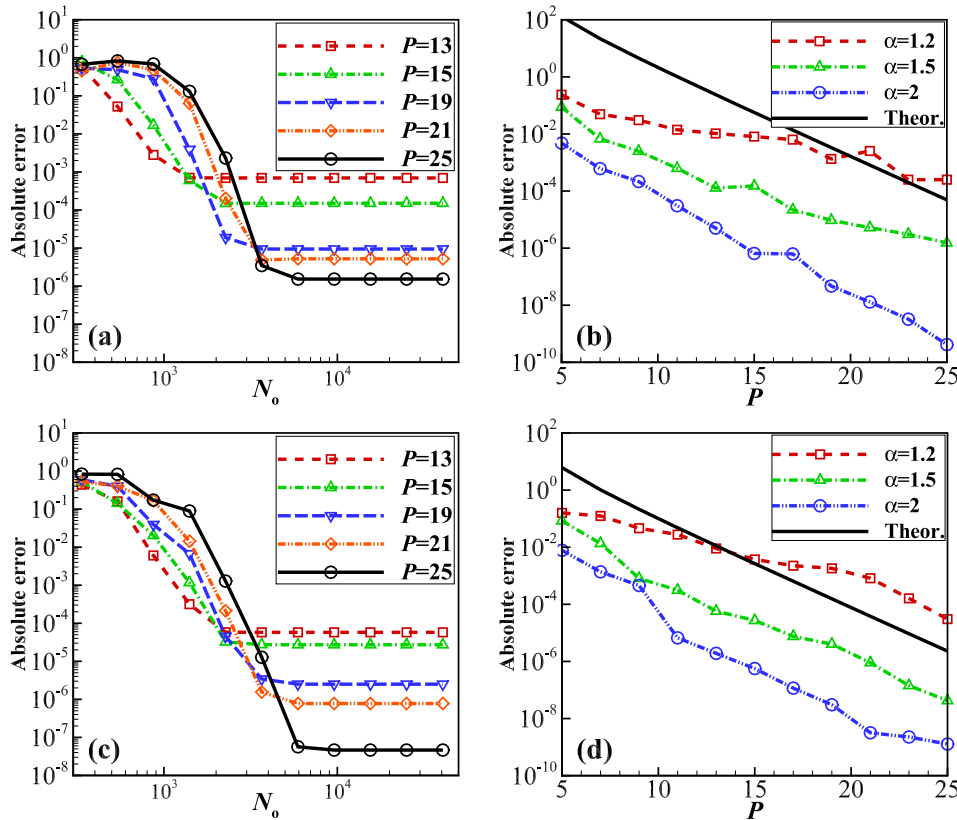
FIG. 3. Errors in the force of the HSMA as functions of the number of the integration points for different P and the number of the spherical harmonic bases for different α. The theoretical lines correspond to parameter $\alpha = 2$. [(a) and (b)] Periodic boundary condition; [(c) and (d)] Dirichlet boundary condition.

problems and the energy calculation. Figure 4(a) shows the error convergence as function of the integration points for systems from $1k$ to $512k$ particles using parameters $\alpha = 2$ and $P = 25$. The error bars represent the standard deviation from mean errors for 5 runs of different initial distributions. These curves have almost the same tendency, in accord with those present in Fig. 2(a), demonstrating that the convergence is almost independent of the number of source points. Figure 4(a) displays the results as function of P with varying α for given $N = 512k$ and $N_o = 5928$. Again, the error decreases rapidly with the increase of P and using a larger α will significantly improve the results, which is in agreement with the results in Fig. 2(b) and the theoretical prediction.

We now move to the timing of the HSMA for these systems. Both the FMM and the GPU accelerations are used together with the direct sum. The simulations of the direct sum and the FMM acceleration are run on an Intel Xeon E5-2680

v4 (14 Cores, 2.40 GHz, 35 MB Cache, 9.6 GT) machine, and all cores are used for each calculation. The publicly available software package FMM3DLIB[20,22] is adopted for the FMM acceleration, where the FMM precision is set as 10^{-6}. The GPU calculations are run using two NVIDIA Tesla P100 GPUs or two NVIDIA Tesla K80 GPUs. In the GPU acceleration, we utilize N_s threads divided into several blocks to calculate pairwise interactions. If the number of particles is large, in order to reduce the shared-memory-bank conflicts in the evaluation, one optimizes the storage structure of particle descriptions or uses the GPU whose architecture supports concurrent reads from multiple threads to a single shared memory address.[45] We set the HSMA parameters $\alpha = 1.3$, $P = 25$, and $N_o = 5928$. Figure 5 illustrates the timing results with the increase of the source charge number. Both the brute force summations with multicore and GPU accelerations show a quadratic scaling, while the FMM acceleration shows a linear scaling. By
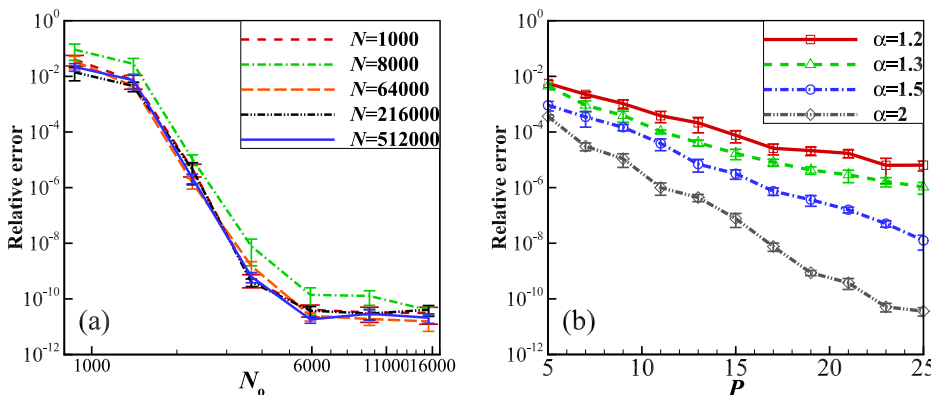

FIG. 4. Errors of the HSMA for the potential as function of (a) the number of integration points for different N, and (b) the number of the spherical harmonic bases for different α and given $N = 512k$.
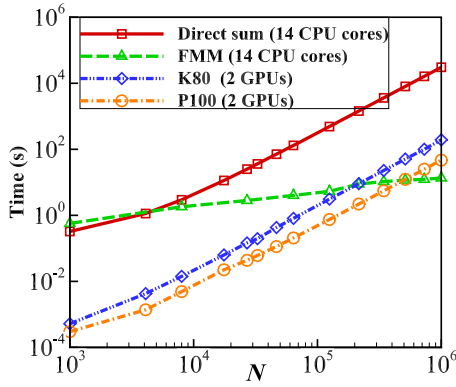
FIG. 5. Comparison of the computation times for the HSMA accelerated with multiple cores, FMM and GPUs (both K80 and P100). The parameters are $\alpha = 1.3$, $P = 25$, and $N_o = 5928$.

comparing the FMM with the direct summation, the breakeven point is ~4k for the HSMA use. When the source number N is less than $100k$, the brute calculation with the GPU acceleration is promising and shows an improvement of 3-4 orders of magnitude in the time cost by comparing it with the direct summation. It is remarked that the performance of different methods depends on how the algorithm is implemented, and the breakeven point can be largely varied if the FMM or the direct summation is optimized. In practice, the most important is the simulation time used for each step. We can observe from Fig. 5 that both the GPU times are less than 1 s for $N = 100k$ and the FMM time is slightly over 1 s, showing that the HSMA will be promising for practical molecular dynamics simulations.

IV. CONCLUDING REMARKS

In summary, the HSMA is proposed for evaluating the electrostatic sum in a cubic box with a general boundary condition. The algorithm transforms the electrostatic problem into the pairwise summation of finite charges, which can be easily accelerated by the FMM to achieve a linear complexity or by the GPUs to reach a 3-4 orders of improvement in time cost. One important feature of the HSMA is its high accuracy. For the mapped integrals, errors from the numerical integration and the approximation of the surface dipole using the central difference is small by increasing the integration points with minor influence to the time cost. The error mainly comes from the truncation of the spherical harmonic series since a small number of terms P is expected for the practical use. Overall, the HSMA can be considered as a method of spectral convergence. For a large-scale system, P can be slightly increased to reduce the radius of the auxiliary surface, and thus the time cost can be significantly reduced.

The HSMA can be very useful in many simulations of molecular systems, in particular, when the system is non-periodic and the Ewald-based lattice summation is not very efficient, e.g., when a solid boundary is present near an electrolyte. Moreover, the HSMA can be promising to serve as a fast solver of the Poisson's equation with a general source term, where the source can be discretized as the sum of many point sources and thus the continuation within the auxiliary surface can be performed.

In many applications, partial periodic systems with dielectric jumps are often studied, for example, electrolytes between two charged dielectric interfaces.[46] It is worthy to note that the minimization step for the harmonic expansion may be not straightforward when the boundary is a dielectric interface as the interface conditions require the solution information of the exterior domain of the simulation box in the form of an infinite pairwise Coulomb sum.[29,47] In this case, the representation of the harmonic series expansion with to-be-determined coefficients is also required, leading to a bigger linear system from the minimization step. The HSMA approach for this problem is our ongoing project. The comparison of the HSMA with the Ewald-based algorithms, the implementation for molecular dynamics simulations, and more applications of the HSMA are also the future work.

ACKNOWLEDGMENTS

APPENDIX: ERROR BOUND OF THE POTENTIAL GRADIENT

We estimate the error of the potential gradient using the truncated spherical harmonic series (6) for the approximation of Φ_{out} which is due to all images outside the auxiliary sphere Ω_s.

Let R_0 be the radius of the circumsphere of the central box and $\alpha = R_s/R_0 > 1$. For any charge q at $\mathbf{r}_i$ outside the auxiliary sphere, the potential can be written as

$$\phi(\mathbf{r}) = \frac{q}{|\mathbf{r} - \mathbf{r}_i|} = \sum_{n=0}^{\infty} \frac{q r^n}{r_i^{n+1}} P_n(\cos\varphi), \qquad (A1)$$

where φ is the angle of $\mathbf{r}$ and $\mathbf{r}_i$. If one truncates the series at $n = P$ and defines the truncation error of its gradient by $\mathbf{E} = (E_1, E_2, E_3)^T$, then

$$|\mathbf{E}| = \left| \nabla \sum_{n=P+1}^{\infty} \frac{q r^n}{r_i^{n+1}} P_n(\cos\varphi) \right|. \qquad (A2)$$

By a simple calculation, one can find that each component of the error vector satisfies

$$E_\ell \leq \sum_{n=P+1}^{\infty} \frac{n r^{n-1}}{r_i^{n+1}} = \left(\frac{r}{r_i}\right)^P \frac{(r_i - r)(P+1) + r}{r_i(r_j - r)^2}. \qquad (A3)$$

Now for all images outside Ω_s, the number density of the image particles can be written as $n(\mathbf{r}) = \sum_{\mathbf{r}_i \in \Omega_s^c} \delta(\mathbf{r} - \mathbf{r}_i)$. Let $\epsilon_{\text{grad}}^\ell$ be the ℓth component of the three-dimensional error vector of the potential gradient using the truncated spherical harmonic expansion (6). Let $q_m = \max |q_i|$. Then we have

$$\epsilon_{\text{grad}}^{\ell} \leq \sum_{\mathbf{r}_i \in \Omega_s^c} |q_i| \left(\frac{r}{r_i}\right)^P \frac{(r_i - r)(P + 1) + r}{r_i(r_i - r)^2}$$

$$\leq q_m \int_{\Omega_s^c} n(\mathbf{x}) \left(\frac{R_0}{x}\right)^P \frac{C}{x} d\mathbf{x}$$

$$\leq 4\pi d_0 q_m C \int_{R_s}^{\infty} \left(\frac{R_0}{x}\right)^P x \, dx, \tag{A4}$$

where $C = [(R_s - R_0)(P + 1) + R_0]/(R_s - R_0)^2$, and we have used the properties $r < R_0$ and $r_i > R_s$ in the second inequality. d_0 is a constant which approximates the average number density $n_0 = N/V$ with V being the volume of the central box, $d_0 \approx n_0$. An accurate value of d_0 depends on the uniformity of the charges, and the upper bound is an approximation if we replace it by n_0. Calculating the integral and using the relation $\alpha = R_s/R_0$, we finally obtain an error bound,

$$\epsilon_{\text{grad}}^{\ell} \leq \frac{4\pi d_0 R_s q_m}{(\alpha - 1)(P - 2)} \left(P + \frac{\alpha}{\alpha - 1}\right) \left(\frac{1}{\alpha}\right)^{P-1}, \tag{A5}$$

which shows the spectral convergence with the order of spherical harmonic expansion. The error bound linearly depends on the number density of the particle, and the ratio α plays the most important role for the accuracy of the approximation.

[1] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications* (Academic Press, New York, 2002).
[2] R. H. French, V. A. Parsegian, R. Podgornik, R. F. Rajter, A. Jagota, J. Luo, D. Asthagiri, M. K. Chaudhury, Y.-M. Chiang, S. Granick *et al.*, Rev. Mod. Phys. **82**, 1887 (2010).
[3] D. A. Walker, B. Kowalczyk, M. O. de la Cruz, and B. A. Grzybowski, Nanoscale **3**, 1316 (2011).
[4] T. A. Darden, D. M. York, and L. G. Pedersen, J. Chem. Phys. **98**, 10089 (1993).
[5] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. Pedersen, J. Chem. Phys. **103**, 8577 (1995).
[6] H. G. Petersen, J. Chem. Phys. **103**, 3668 (1995).
[7] R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles* (Taylor & Francis, 1988).
[8] J. Shimada, H. Kaneko, and T. Takada, J. Comput. Chem. **14**, 867 (1993).
[9] B. A. Luty, M. E. Davis, I. G. Tironi, and W. F. Van Gunsteren, Mol. Simul. **14**, 11 (1994).
[10] M. Mazars, Phys. Rep. **500**, 43 (2010).
[11] A. Arnold and C. Holm, Comput. Phys. Commun. **148**, 327 (2002).
[12] A. Arnold, J. de Joannis, and C. Holm, J. Chem. Phys. **117**, 2496 (2002).
[13] Z. Hu, J. Chem. Theory Comput. **10**, 5254 (2014).
[14] A. P. dos Santos, M. Girotto, and Y. Levin, J. Chem. Phys. **147**, 184105 (2017).
[15] I. Fukuda and H. Nakamura, Biophys. Rev. **4**, 161 (2012).
[16] W. Cai, S. Deng, and D. Jacobs, J. Comput. Phys. **223**, 846 (2007).
[17] Y. Liang, Z. Xu, and X. Xing, New J. Phys. **17**, 083082 (2015).
[18] J. D. Jackson, *Classical Electrodynamics*, 3rd ed. (John Wiley & Sons, New York, 2001).
[19] L. Greengard and V. Rokhlin, J. Comput. Phys. **73**, 325 (1987).
[20] L. Greengard and V. Rokhlin, Acta Numer. **6**, 229 (1997).
[21] L. Ying, G. Biros, and D. Zorin, J. Comput. Phys. **196**, 591 (2004).
[22] H. Cheng, L. Greengard, and V. Rokhlin, J. Comput. Phys. **155**, 468 (1999).
[23] A. Gillman and A. Barnett, J. Comput. Phys. **248**, 309 (2013).
[24] G. Marple, A. Barnett, A. Gillman, and S. Veerapaneni, SIAM J. Sci. Comput. **38**, B740 (2016).
[25] N. A. Gumerov and R. Duraiswami, J. Comput. Phys. **272**, 307 (2014).
[26] Q. Zhao, P. Liu, and Z. Xu, Commun. Comput. Phys. **24**, 1241 (2018).
[27] T. Shigeta, D. L. Young, and C.-S. Liu, J. Comput. Phys. **231**, 7118 (2012).
[28] M. H. Cho and A. H. Barnett, Opt. Express **23**, 1775 (2015).
[29] P. K. Yang, S. H. Liaw, and C. Lim, J. Phys. Chem. B **106**, 2973 (2002).
[30] G. King and A. Warshel, J. Chem. Phys. **91**, 3647 (1989).
[31] H. Alper and R. M. Levy, J. Chem. Phys. **99**, 9847 (1993).
[32] Y. Lin, A. Baumketner, S. Deng, Z. Xu, D. Jacobs, and W. Cai, J. Chem. Phys. **131**, 154103 (2009).
[33] D. Beglov and B. Roux, J. Chem. Phys. **100**, 9050 (1994).
[34] F. Ethridge and L. Greengard, SIAM J. Sci. Comput. **23**, 741 (2001).
[35] H. Cheng, J. Huang, and T. J. Leiterman, J. Comput. Phys. **211**, 616 (2006).
[36] R. Swinbank and R. J. Purser, Q. J. R. Meteorol. Soc. **132**, 1769 (2006).
[37] J. H. Hannay and J. F. Nye, J. Phys. A **37**, 11591 (2004).
[38] W. M. Brown, A. Kohlmeyer, S. J. Plimpton, and A. N. Tharrington, Comput. Phys. Commun. **183**, 449 (2012).
[39] S. Le Grand, A. W. Goetz, and R. C. Walker, Comput. Phys. Commun. **184**, 374 (2013).
[40] W. Geng and F. Jacob, Comput. Phys. Commun. **184**, 1490 (2013).
[41] R. Adelman, N. A. Gumerov, and R. Duraiswami, IEEE Trans. Magn. **53**, 7002311 (2017).
[42] See https://github.com/liangjiuyang/ for the source files of the computer program.
[43] E. Madelung, Phys. Z. **19**, 524 (1918).
[44] P. P. Ewald, Ann. Phys. **369**, 253 (1921).
[45] L. Nyland, M. Harris, and J. Prins, *GPU Gems 3* (Addison-Wesley Professional, 2007), pp. 677–694.
[46] A. Arnold and C. Holm, Adv. Polym. Sci. **185**, 59 (2005).
[47] Z. Xu, W. Cai, and X. Cheng, Commun. Comput. Phys. **9**, 1056 (2011).