The fast Fourier Poisson method for calculating Ewald sums

Darrin York, and Weitao Yang

Articles you may be interested in

A smooth particle mesh Ewald method
The Journal of Chemical Physics **103**, 8577 (1998); 10.1063/1.470117

Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems
The Journal of Chemical Physics **98**, 10089 (1998); 10.1063/1.464397

The effect of long-range electrostatic interactions in simulations of macromolecular crystals: A comparison of the Ewald and truncated list methods
The Journal of Chemical Physics **99**, 8345 (1998); 10.1063/1.465608

Multigrid methods for classical molecular dynamics simulations of biomolecules
The Journal of Chemical Physics **114**, 6578 (2001); 10.1063/1.1352646

Comparison of simple potential functions for simulating liquid water
The Journal of Chemical Physics **79**, 926 (1998); 10.1063/1.445869

Molecular dynamics with coupling to an external bath
The Journal of Chemical Physics **81**, 3684 (1998); 10.1063/1.448118

The fast Fourier Poisson method for calculating Ewald sums

Darrin York[a),b)] and Weitao Yang

Department of Chemistry, Duke University, Durham, North Carolina 27708-0354

The conventional Ewald expression for the electrostatic energy and forces is recast in a form that can be evaluated to high accuracy in order $N \log(N)$ steps using fast Fourier transforms. The fast Fourier Poisson method does not rely on interpolation approaches or Taylor/multipole expansions, and can be easily integrated with conventional molecular dynamics algorithms.

INTRODUCTION

A long standing obstacle to accurate molecular simulations concerns proper treatment of long-range electrostatic forces. Conventional molecular modeling force fields generally rely on point charge models for atoms and ions, and frequently employ periodic boundary conditions to approximate a macroscopic system. Evaluation of all pairwise Coulombic interactions using the "minimum image" convention requires order N^2 steps, and hence is a limitation for large systems. To reduce computational effort to an order N procedure, truncation schemes are typically employed that neglect interactions beyond a fixed cutoff. Cutoff methods (including the "minimum image" convention), however, ultimately lead to artificial behavior in molecular simulations where long-range electrostatic forces play a dominant role in the long-time behavior of the system.[1-3] An alternative method is to use Ewald sums[4,5] to evaluate the electrostatic potential due to the entire crystal lattice, and thus incorporate all the long-range effects.

Consider a neutral system of N point charges $q_1, q_2, \ldots, q_N$, at positions $r_1, r_2, \ldots, r_N$, in a unit cell with real space lattice vectors a_1, a_2, *and* a_3. The electrostatic energy of the system is defined by the interaction of each point charge with all the other point charges in the lattice,

$$E = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} q_i q_j \sum_{|n|=0}^{\infty} {}' |r_{ij} + n|^{-1} = \frac{1}{2} \sum_{i=1}^{N} q_i \varphi(r_i), \quad (1)$$

where the sum over n is a sum over lattice vectors $n = n_1 a_1 + n_2 a_2 + n_3 a_3$, and φ is the electrostatic potential. The prime symbol indicates the term for which $|r_{ij} + n| = 0$ is neglected. Hence, the potential φ in Eq. (1) does not include the infinite "self-energy" of the point charges, and is not a solution of Poisson's equation. Suppose we introduce a localized "screening" charge density around each point charge of equal magnitude and opposite sign.[5,6] Define the screening charge density $\rho_s(r)$ to be the sum of the localized densities. For concreteness, we choose the local densities to be spherical Gaussians

$$\rho_s(r) = -\sum_{j=1}^{N} q_j \left(\frac{\beta}{\sqrt{\pi}} \right)^3 e^{-\beta^2 |r - r_j|^2} = \sum_{j=1}^{N} \rho_s^j(r), \quad (2)$$

where the positive parameter β determines the width of the Gaussians. The potential φ in Eq. (1) can be rewritten

$$\varphi(r_i) = [\varphi(r_i) + \varphi_s(r_i)] - \varphi_s(r_i) = \varphi_{\text{real}}(r_i) + \varphi_{\text{recip}}(r_i). \quad (3)$$

The first term in Eq. (3), $\varphi_{\text{real}}(r_i)$, is a sum of short-range potentials associated with each point charge and its screening density, and can be evaluated in real space. The second term, $\varphi_{\text{recip}}(r_i)$, is the potential associated with the smooth periodic charge density $-\rho_s(r)$ and can be obtained by solving Poisson's equation in reciprocal space. The analytic expression for the electrostatic energy equation (1) with the choice of screening density equation (2) can be written[7,8]

$$
\begin{aligned}
E = {} & \frac{1}{2} \sum_{i=1}^{N} q_i \Bigg[\left(\sum_{j=1}^{N} {}' q_j \sum_{|n|=0}^{\infty} \frac{\text{erfc}(\beta|r_{ij}+n|)}{|r_{ij}+n|} - \frac{\beta}{\sqrt{\pi}} q_j \delta_{ij} \right) \\
& + \left(\sum_{j=1}^{N} q_j \frac{1}{\pi V} \sum_{|m| \neq 0}^{\infty} \frac{e^{-\pi^2 m^2/\beta^2}}{m^2} e^{i 2\pi m \cdot r_{ij}} \right) \Bigg] \\
& + J(D,P,\epsilon') \\
= {} & \frac{1}{2} \sum_{i=1}^{N} q_i [\varphi_{\text{real}}(r_i) + \varphi_{\text{recip}}(r_i)] + J(D,P,\epsilon') \\
= {} & E_{\text{real}} + E_{\text{recip}} + J(D,P,\epsilon'), \quad (4)
\end{aligned}
$$

where V is the volume of the unit cell, erfc (x) is the complementary error function, and E_{real} and E_{recip} are the real space and reciprocal space components of the Ewald energy, respectively. The last term $J(D,P,\epsilon')$ in Eq. (4) corrects for the nonuniform field associated with a macroscopic (but finite) crystal in a dielectric continuum, and depends on the dipole moment D of the unit cell, the shape P of the macroscopic crystal, and the external dielectric constant ϵ'.[7,8] This term vanishes for unit cells with zero dipole moment, or in the limit $\epsilon' \to \infty$. The sums in Eq. (4) are over real space (n) and reciprocal space (m) lattices, respectively, and are rapidly convergent. The parameter β adjusts the relative rates of convergence, but is otherwise arbitrary. If β is chosen such that the real space sum is negligible beyond some fixed cutoff, r_c, evaluation of Eq. (4) becomes an order N^2 procedure.[5] Adjusting β to optimize computational efficiency results in an order $N^{3/2}$ procedure,[9] which is still restrictive for large systems.

METHOD

Herein we describe a simple reformulation of Eq. (4) that can be evaluated efficiently to high accuracy in order $N \log(N)$ steps. The fast Fourier Poisson (FFP) method uses a modified expression for the real space term that can be evaluated using a conventional nonbond list in order N steps, and a reciprocal space term that can be evaluated using fast Fourier transforms (FFTs) in order $N \log(N)$ steps. The FFP method differs from other methods[10-17] in that it can achieve very high accuracy, and does not rely on interpolation techniques or Taylor/multipole expansions.

Consider the potential φ_{recip} associated with the charge density $-\rho_s$ in Eq. (2). These functions are related by Poisson's equation $\nabla^2 \varphi_{recip}(\mathbf{r}) = 4\pi\rho_s(\mathbf{r})$ [note $\nabla^2 \varphi_{recip}(\mathbf{r}) = -\nabla^2 \varphi_s(\mathbf{r})$]. From a computational standpoint, φ_{recip} can be evaluated on a grid using FFTs. The procedure is as follows: (1) evaluate $4\pi\rho_s$ at the Fourier transform (FT) grid points, (2) compute the forward real→complex FFT, (3) multiply the transformed components by $-(2\pi m)^{-2}$ (where $\mathbf{m}$ is the corresponding reciprocal space lattice vector), and (4) compute the reverse real←complex FFT to obtain the potential. Steps (1) and (3) are order N for a given grid density, whereas steps (2) and (4) are order $N \log(N)$. The gradient components can be obtained in a similar manner by replacing $-(2\pi m)^{-2}$ in step 3 by $-i\mathbf{m}/2\pi m^2$. The above procedure results in accurate determination of φ_{recip} and its gradient components at the FT grid points. Equation (4), however, requires knowing φ_{recip} at the point charge positions $\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N$. Fortunately, we can recast Eq. (4) in a form that does not have this restriction. In the same spirit as the separation of the potential in Eq. (3), separate the reciprocal space energy term,

$$E_{recip} = \frac{1}{2} \sum_{i=1}^{N} q_i \varphi_{recip}(\mathbf{r}_i) = \frac{1}{2} \int \rho(\mathbf{r}') \varphi_{recip}(\mathbf{r}') d^3 r'$$

$$= \frac{1}{2} \int [\rho(\mathbf{r}') + \rho_s(\mathbf{r}')] \varphi_{recip}(\mathbf{r}') d^3 r'$$

$$- \frac{1}{2} \int \rho_s(\mathbf{r}') \varphi_{recip}(\mathbf{r}') d^3 r', \quad (5)$$

where we have introduced the notation for the point charge density $\rho(\mathbf{r}) = \sum_{i=1}^{N} q_i \delta(\mathbf{r} - \mathbf{r}_i)$. The idea is that we replace the interaction of each point charge q_i with the potential φ_{recip} by the interaction of a Gaussian with the same net charge and at the same location, plus a correction (which is exact). In fact, one term in the correction exactly cancels the real space term in Eq. (4). The results for the energy expression and force are

$$E = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} q_i q_j \frac{erfc(\beta r_{ij}/\sqrt{2})}{r_{ij}} - \frac{\beta}{\sqrt{\pi}} \sum_{i=1}^{N} q_i^2$$
$$(r_{ij} < r_c)$$

$$- \frac{1}{2} \int \rho_s(\mathbf{r}') \varphi_{recip}(\mathbf{r}') d^3 r' + J(\mathbf{D}, P, \epsilon') \quad (6a)$$

and

$$\mathbf{f}_i = -q_i \sum_{\substack{j=1 \\ (r_{ij} < r_c)}}^{N} q_j \left(\beta \sqrt{\frac{2}{\pi}} e^{-\beta^2 r_{ij}^2/2} - \frac{erfc(\beta r_{ij}/\sqrt{2})}{r_{ij}} \right) \frac{\mathbf{r}_{ij}}{r_{ij}^2}$$

$$+ \int \rho_s^i(\mathbf{r}') \nabla \varphi_{recip}(\mathbf{r}') d^3 r' - \nabla_i J(\mathbf{D}, P, \epsilon'). \quad (6b)$$

The advantage of Eqs. (6a) and (6b) is that ρ_s, φ_{recip}, and $\nabla \varphi_{recip}$ are well represented at the FT grid points, hence the corresponding integrals can be computed accurately. The real space term in Eq. (6a) is modified from Eq. (4) by a factor of $1/\sqrt{2}$ in the erfc(x) argument. Note that if only the total energy E is desired (and not the potential in real space), the integral in Eq. (6a) can be evaluated in reciprocal space thereby avoiding a reverse FFT. Evaluation of the integral in Eq. (6b) for the force involves a local summation over the FT grid since the density ρ_s^i of Eq. (2) is highly localized. The FFP method is not restricted to orthogonal unit cells, and has the advantage that the energy and gradients are continuous functions of point charge position. As demonstrated below, Eqs. (6a) and (6b) can be evaluated efficiently to high accuracy using a conventional 9 Å cutoff for the real space sums, which makes implementation into conventional molecular dynamics programs straight forward.

RESULTS

The FFP method has been implemented as a series of simple self-contained FORTRAN subroutines. Here we examine the accuracy and computational scaling of the method. Two systems were chosen to test accuracy: (1) the conventional TIP3P water box of Jorgensen et al.[18] and (2) an ionic system consisting of 10 Na^+ ions and 10 Cl^- ions in a 40 Å cube with 1916 TIP3P waters. Accuracy and timing results for several FT grid densities are listed in Table I. Moderate accuracy is obtained (10^{-4} relative force accuracy) with a fairly course FT grid (~1.6 Å). Very high accuracy (10^{-10} relative force accuracy) can be obtained with a finer grid (~0.6 Å). In the current implementation, the main cost in obtaining high accuracy involves evaluation of the density ρ_s at the local FT grid points. In all cases a cutoff of 9 Å was used in the real space sums in Eqs. (6a) and (6b). For a given cutoff (r_c), the parameter β is determined so that the real space term erfc($\beta r_c/\sqrt{2}$)/r_c falls below a desired tolerance. The screening density around each atom is then evaluated at local grid points using a cutoff $r_c' = r_c/\sqrt{2}$. For any given grid density, this procedure scales linearly with the size of the system, and hence the overall scaling dependence is $N \log(N)$ resulting from the computation of the FFTs (Fig. 1).

CONCLUSION

The FFP method offers several advantages for calculating long-range electrostatic forces of large systems. Perhaps the most attractive is its simplicity. The method does not rely on interpolation approaches or on Taylor/multipole expansions, and can be integrated into conventional molecular dy-

TABLE I. Accuracy and timing of the FFP method.[a]

Grid (Å)	β	rmsE	rmsφ	rmsF	CPU
Water box[b]					
1.56	0.502	1.8×10^{-5}	2.2×10^{-3}	4.8×10^{-4}	2.1
1.17	0.553	6.8×10^{-7}	2.2×10^{-4}	3.8×10^{-5}	3.2
0.78	0.687	1.1×10^{-9}	3.4×10^{-7}	7.8×10^{-8}	7.9
0.59	0.799	2.5×10^{-11}	4.3×10^{-10}	2.3×10^{-10}	18.4
9 Å ...	...	2.5×10^{-2}	0.99	8.6×10^{-2}	0.7
Ion box[c]					
1.67	0.502	1.2×10^{-5}	7.3×10^{-3}	2.5×10^{-4}	18.2
1.25	0.553	7.3×10^{-7}	4.2×10^{-4}	1.6×10^{-5}	29.1
0.83	0.69	3.3×10^{-9}	1.1×10^{-6}	5.5×10^{-8}	83.8
0.63	0.763	3.5×10^{-11}	6.0×10^{-9}	2.6×10^{-10}	201.6
9 Å ...	...	1.6×10^{-3}	0.72	9.6×10^{-3}	6.1
18 Å ...	...	2.7×10^{-2}	0.71	5.1×10^{-3}	45.1

[a]Listed are the spacing between FT grid points (grid), the parameter β (see text), the relative energy error (rmsE): $\sqrt{(E-\hat{E})^2/E^2}$, the relative potential error (rmsφ): $\sqrt{\Sigma_i[\varphi(\mathbf{r}_i)-\hat{\varphi}(\mathbf{r}_i)]^2/\Sigma_i\varphi(\mathbf{r}_i)^2}$, the relative force error (rmsF): $\sqrt{\Sigma_i(\mathbf{f}_i-\hat{\mathbf{f}}_i)^2/\Sigma_i\mathbf{f}_i^2}$, and the CPU time required on an SGI Indigo R4000 Workstation (CPU). Here the energy E, potential φ, and forces $\mathbf{f}_i$ correspond to the exact quantities evaluated using the Ewald method excluding nearest image interactions of covalently bonded atoms, and $\hat{E}$, $\hat{\varphi}$, and $\hat{\mathbf{f}}_i$ are the corresponding approximate values evaluated with the FFP method. In all calculations the shape-dependent term $J(\mathbf{D},P,\epsilon)$ in Eq. (4) was chosen to be zero, corresponding to the commonly employed "tin-foil" ($\epsilon'\rightarrow\infty$) boundary conditions. Listed for comparison are results obtained using a conventional residue-based nonbond cutoff ($r_c=9$ Å, 18 Å).
[b]Box containing 216 Monte Carlo TIP3P waters.
[c]Box containing 10 Na^+ and 10 Cl^- ions and 1916 TIP3P waters, relaxed with 1000 steps conjugate gradient energy minimization and 10 ps molecular dynamics at 300 K.

namics codes that use a nonbond list. Moderate accuracy (10^{-4} relative force error) can be obtained with the FFP method very efficiently. High accuracy (10^{-10} relative force error) can be obtained by using a finer grid ($\sim$0.6 Å). Finally, since the formulation of the FFP approach relies on solving the potential of a smooth periodic charge density, in principle, the method can be generalized to arbitrary charge densities such as those encountered in real space electronic structure calculations.
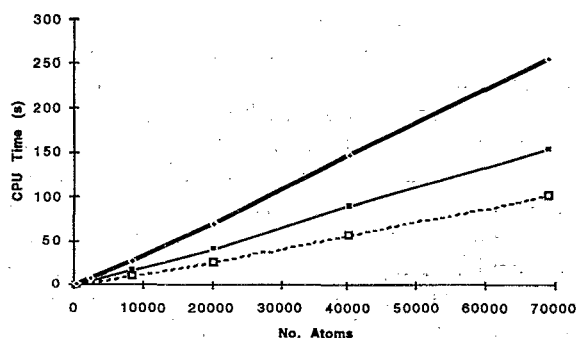

FIG. 1. Scaling of the FFP method. CPU times for evaluation of the energy and gradient are shown for the real space sum (thin solid line), reciprocal space sum (thin dotted line), and overall (thick solid line). Timings were performed on cubic water boxes (15, 30, 45, 60, 75, and 90 Å) using a 9 Å cutoff for the real space sum, and a 1.5 Å FT grid spacing (grid density 0.3 pts/Å^3). A Silicon Graphics Iris Indigo R4000 Workstation was used for all timings.

[1] H. Schreiber and O. Steinhauser, Biochemistry 31, 5856 (1992).
[2] H. Schreiber and O. Steinhauser, Chem. Phys. 168, 75 (1992).
[3] D. York, T. Darden, and L. Pedersen, J. Chem. Phys. 99, 8345 (1993).
[4] P. Ewald, Ann. Phys. (Leipzig) 64, 253 (1921).
[5] M. P. Allen and D. J. Tildesley, Computer Simulations of Liquids (Oxford University, New York, 1987).
[6] D. M. Heyes, J. Chem. Phys. 74, 1924 (1981).
[7] S. W. De Leeuw, J. W. Perram, and E. R. Smith, Proc. R. Soc. London, Ser. A 373, 27 (1980).
[8] E. R. Smith, Proc. R. Soc. London Ser. A 375, 475 (1981).
[9] J. W. Perram, H. G. Petersen, and S. W. De Leeuw, Mol. Phys. 65, 875 (1988).
[10] R. W. Hockney and J. W. Eastwood, Computer Simulation Using Particles (McGraw-Hill, New York, 1981).
[11] L. Greengard and V. Rokhlin, J. Comput. Phys. 73, 325 (1987).
[12] K. E. Schmidt and M. A. Lee, J. Stat. Phys. 63, 1223 (1991).
[13] J. A. Board, Jr., J. W. Causey, J. F. Leathrum, Jr., A. Windemuth, and K. Schulten, Chem. Phys. Lett. 198, 89 (1992).
[14] H.-Q. Ding, N. Karasawa, and W. A. Goddard III, J. Chem. Phys. 97, 4309 (1992).
[15] F. S. Lee and A. Warshel, J. Chem. Phys. 97, 3100 (1992).
[16] J. Shimada, H. Kaneko, and T. Takada, J. Comput. Chem. 15, 28 (1994).
[17] T. Darden, D. York, and L. Pedersen, J. Chem. Phys. 98, 10089 (1993).
[18] W. Jorgensen, J. Chandrasekhar, J. Madura, R. Impey, and M. Klein, J. Chem. Phys. 79, 926 (1983).