Accuracy and efficiency of the particle mesh Ewald method

Henrik G. Petersen

Citation: J. Chem. Phys. **103**, 3668 (1995); doi: 10.1063/1.470043
View online: http://dx.doi.org/10.1063/1.470043
View Table of Contents: http://jcp.aip.org/resource/1/JCPSA6/v103/i9
Published by the American Institute of Physics.

Accuracy and efficiency of the particle mesh Ewald method

Henrik G. Petersen
Department of Mathematics and Computer Science, Odense University, DK-5230 Odense M, Denmark

(Received 22 March 1995; accepted 23 May 1995)

In this article, a recently proposed method called the particle mesh Ewald (PME) method for computing the long ranged Coulomb interactions in for example molecular dynamics simulations is studied. The PME method has a complexity $\mathcal{O}(N \log N)$, where N is the total number of charges. This complexity should in particular be compared with the complexity $\mathcal{O}(N^{3/2})$ for the well known Ewald method and $\mathcal{O}(N)$ for the rather new (but already famous) fast multipole method (FMM). However, these complexities say nothing about which method is fastest at some finite N. The purpose of this article is thus to study the PME method and compare its efficiency with the Ewald method and the fast multipole method. To enable this, a theoretical estimate for the accuracy of the PME method as function of its truncation parameters is derived. It is shown that this estimate is very precise by comparing it with results obtained from molecular dynamics simulations of a molten NaCl. Based on this estimate and very careful time experiments, the overall necessary time overhead for the PME method as function of N and a required accuracy is predicted. By a direct comparison with a similar prediction for the Ewald method and by studying existing Ewald-FMM comparisons, it is found that the PME method is significantly faster than both the Ewald method and the fast multipole method in the important decades $N \simeq 10^4$–10^5. © *1995 American Institute of Physics.*

I. INTRODUCTION

With the large increase of computer power over the past and present years, it is becoming possible to simulate systems subject to long-ranged interactions of recently impossible scale. Problems in which such systems are considered are molecular dynamics and Monte Carlo simulations of condensed matter containing charged and/or dipolar interactions, energy minimization of biological macromolecules, astrophysical simulations, and simulations of colloidal suspensions. There is now widespread agreement that the most accurate and efficient way of suppressing finite sample size surface effects is to embed the system in truly periodic boundary conditions. Different methods are available for calculating the long-ranged interactions of such systems. These methods have different complexities in both the accuracy and the number of particles. Below, we shall review the most important methods and discuss how these methods have been compared with respect to performance. However, we shall first turn the attention on the very important issue of having a significant measure for the accuracy of the calculation of the interactions. This accuracy of course depends on the level of truncation for any method considered for the calculation. When performing simulations it is important to control the accuracy properly and when comparing different methods with respect to speed, the comparison should be made at the same accuracy. As accuracies on the energy may be obscured by unimportant constant contributions, is very sensitive to fluctuations and can provide no local information, we shall consider accuracies on the forces only. This accuracy may be obtained by different measures. In information distributed equilibrium simulations the average accuracy is most important. This accuracy should be estimated by

$$\Delta f = \left[\frac{1}{N} \sum_{j=1}^{N} \Delta f_j^2 \right]^{1/2}, \tag{1}$$

where Δf_j denotes the error on the total force on the jth particle and N is the total number of particles. This estimate suppresses fluctuations, is widely used, very significant and can be easily obtained from computational experiments. However, in rare events simulations such as diffusion over a potential barrier (and maybe even in all simulations) it is also important to control local errors. These local errors can be controlled by controlling the maximum error on the total force on a single particle. We shall measure this error by the worst to average accuracy ratio

$$\gamma = \frac{\max_j |\Delta f_j|}{\Delta f}.$$

When the $|\Delta f_j|$'s are Gaussianly distributed values of γ around 3 for $N = 10^3$–10^5 should be expected. Throughout this paper, we consider comparisons with respect to the average accuracy Δf although we for each method give a brief discussion of typical γ-values.

The most well-known (and used) method to simulate systems in truly periodic boundary conditions is the Ewald method (see Sec. II and Refs. 4 and 5). Although the Ewald method almost always has been applied to charged or dipolar systems, techniques similar to those used for deriving the Ewald potential for charged and dipolar systems can be used to obtain Ewald-type rapidly computational forms for other potential as well.[6] We shall however for simplicity only consider electrically neutral systems of charges in this article. One of the advantages of the Ewald method is that there exists very reliable estimates of the average accuracy given by Eq. (1) as a function of the parameters of the method.[1] These estimates allow a detailed analysis of the performance of the method and also facilitates a determination of its parameters that is optimal in the sense that the parameters minimize, the overall computation time for any required accuracy. We shall return to this estimate and the optimal choice of parameters, but at present it should be mentioned

 0021-9606/95/103(9)/3668/12/$6.00

that for an optimal choice of the parameters the overall computation time grows as $N^{3/2}$ the coefficient being accuracy dependent.[6] For the Ewald method, the values of γ are claimed to be in the Gaussian range.[1] As the Ewald method is widely used and extremely well understood, we shall use it as a reference method for other methods with respect to performance throughout this paper.

The Ewald method is very suitable for calculating the electrostatic interactions of a modest number of particles ($N=10^2$–10^3). However, with the increasing computer power it has become possible to simulate much larger systems ($N=10^4$–10^5), and alternative methods have therefore become important. A recently proposed method[3] usually called the fast multipole method (FMM) or Greengard–Rokhlin method is based on expanding the potential in spherical harmonics and use of a tesselation of the simulation box into a hierarchy of subcells to organize transfers of the expansions. It has been shown that the computation time grows linearly with N. The coefficient is as in the Ewald method accuracy dependent, but a formula for the average accuracy as function of the truncation level of the expansion has not been found, and thus it is not possible to make an average accuracy based theoretical performance analysis and optimization. However, recently an accurate estimate of the worst case error of a single pair interaction has been found[10,11] based on which a performance optimization may be obtained.[12] Based on the estimate in Ref. 10, an average accuracy estimate has actually been found for the two dimensional case.[13]

Many experimental comparisons have been made between the Ewald method and the fast multipole method and breakevens (that is the number of charges below which the Ewald method is faster and above which the fast multipole method is faster) between $N=300$,[17] a few thousand[15] and 100 000 (Ref. 14) have been claimed. Although it is not the main purpose of this paper to discuss the FMM performance, it should be mentioned that the results of Refs. 15 and 17 are hard (impossible) to believe. A simple analysis of the overheads of the two methods in terms of simple hardware independent operations such as additions and multiplications shows that a breakeven below $N=10\,000$ cannot be correct for a reasonable accuracy requirement. A further indication of which results can be believed is the effort of explaining how the methods were implemented (in particular the Ewald method as uncareful implementations may yield huge slowdowns[6,7]). In Ref. 15 it is admitted that the Ewald method has been implemented very inefficiently, whereas in Ref. 17 there are no details at all about how the Ewald method was implemented, which together with their Ewald timing results suggests that they have also implemented the Ewald method inefficiently. On the other hand, in Ref. 14 discussions of implementational details to ensure efficiency are given for the Ewald method as well as for the FMM. Based on the observations above and own Ewald-FMM comparisons in two dimensions,[13] where breakevens above $N=10\,000$ were found lead us to the conclusion that the correct breakeven will lie in the range stated in Ref. 14. Furthermore, a little discussed disadvantage of the FMM method is that large non-Gaussian values of γ are found. The γ-values are typically in the range of 10–20 or even higher!

A very new method by Darden *et al.*[18] (called the particle mesh Ewald method or PME method) is based on using fast Fourier transform (FFT) techniques to evaluate the reciprocal space part of the Ewald method. The ideas of Hockney and Eastwood[8] is used to assign the charges to a mesh according to their positions and in turn to calculate the reciprocal space contribution to the total energy and the forces by applying FFT's to the mesh. The fast evaluation of the reciprocal space part is used to change the parameters to allow a $\mathcal{O}(N)$ evaluation of the real space part. In Ref. 18 the PME method was successfully applied to three different crystalline ionic systems and the speed of the PME method using charge assignments of order up to 4 was compared to the speed of evaluating the nonbonded interactions using simple cutoffs. However, the PME method was not compared with other methods for accurate evaluation of electrostatic interactions such as the Ewald method. Furthermore, the parameters of the PME is chosen not to optimize the performance of the PME itself, but rather so as to allow an easy incorporation into the rest of a molecular dynamics program by ensuring the range of the real-space PME interactions to be similar to the range of the other (short-ranged) interactions.

The purpose of this article is to study the PME for charge assignment schemes of any order. As the PME method is really a modification of the Ewald method and is to be compared with it, we give a brief review of the Ewald method in Sec. II containing the essential Ewald formulas and the Ewald accuracy estimates. In Sec. III we give a short presentation of the PME method and in Sec. IV an average accuracy estimate for the PME method is derived. In Sec. V, this estimate is compared to average accuracies obtained from molecular dynamics simulations of a molten NaCl for a variety of parameter sets, and it is found that the estimate is reasonably to very precise. Values of γ in the Gaussian range are found experimentally. Furthermore, we measure primitive PME and Ewald time overheads, which are used in Sec. VI to optimize method parameters for any given accuracy. In turn, expressions for the computing time as function of N and the accuracy for the parameter optimized methods are derived. Based on these expressions, we are able to make theoretical comparisons of the performance of the parameter optimized PME and the performance of the parameter optimized Ewald method and we find a breakeven below $N=10\,000$ and that the PME method is several times faster for 100 000 charges and about 10 times faster for 1 000 000 charges. In Sec. VII, we discuss the potential concurrency of the PME method as this issue is certainly also very important. We illustrate, that the PME method is as suitable for a parallel computer as the Ewald method and the FMM. In the derivation of the estimates and the timing models, it is assumed that the particles are approximately uniformly (randomly) distributed. We therefore conclude this article by giving a brief discussion of the expected PME performance for highly nonuniform distributions of charges.

II. THE EWALD METHOD

Consider a system of charges $q_1,\ldots,q_N$ residing at $\mathbf{R}_1,\ldots,\mathbf{R}_N$ in a cubic box of side length L (we shall throughout this paper restrict ourselves to the cubic case, although all results can be extended easily to the noncubic case). To be able to suppress boundary effects and to allow a connection to real world systems one usually consider the system to be embedded in a large three dimensional array of copies of the box having some shape S and being itself embedded in a uniform dielectric with dielectric constant ϵ'. When letting the size of the array go to infinity without changing the shape, the following rapidly computable form for the electrostatic energy can be obtained:[5,22]

$$E=\frac{1}{L}\left[E_r\left(\frac{\mathbf{R}_1}{L},\ldots,\frac{\mathbf{R}_N}{L}\right)+E_k\left(\frac{\mathbf{R}_1}{L},\ldots,\frac{\mathbf{R}_N}{L}\right)+E_D\left(\frac{\mathbf{R}_1}{L},\ldots,\frac{\mathbf{R}_N}{L},\epsilon',S\right)\right], \tag{2}$$

where the real space part E_r and the reciprocal space part E_k are given by

$$\begin{aligned}E_r(\mathbf{r}_1,\ldots,\mathbf{r}_N)&=\sum_{\mathbf{n}\in Z^3}\sum_{1\leqslant i<j\leqslant N}\frac{\operatorname{erfc}(\alpha|\mathbf{r}_i-\mathbf{r}_j+\mathbf{n}|)}{|\mathbf{r}_i-\mathbf{r}_j+\mathbf{n}|},\\ E_k(\mathbf{r}_1,\ldots,\mathbf{r}_N)&=\sum_{\mathbf{k}\in Z^3,\mathbf{k}\neq\mathbf{0}}\frac{\exp(-\pi^2\mathbf{k}^2/\alpha^2)}{2\pi\mathbf{k}^2}\\ &\quad\times\sum_{j=1}^{N}\sum_{l=1}^{N}q_jq_l\exp[2\pi i\mathbf{k}\cdot(\mathbf{r}_j-\mathbf{r}_l)],\end{aligned} \tag{3}$$

and where E_D is a shape dependent term. We shall only consider the case where the shape is spherical. Then E_D can be written as

$$E_D(\mathbf{r}_1,\ldots,\mathbf{r}_N)=\frac{2\pi}{(2\epsilon'+1)}\left|\sum_{j=1}^{N}q_j\mathbf{r}_j\right|^2-\frac{\alpha}{\sqrt{\pi}}Q^2, \tag{4}$$

where $Q=(\Sigma_{j=1}^{N}q_j^2)^{1/2}$. The parameter α can be chosen arbitrarily (positive). It should be chosen so as to optimize performance. We shall return to the choice of α in Sec. VI. The formulas above for the electrostatic energy are the basic formulas for the Ewald method. We first notice, that the computational overhead for computing the energy and corresponding force contribution from E_D is negligible. In a practical calculation of the potential energy (and/or the corresponding forces) it is used that the double sum over particles in E_k actually is a product of two single sums, and the sum over $\mathbf{k}$-vectors is truncated at $|\mathbf{k}|=K$, where K depends on α and the required accuracy. We can thus write E_k as

$$E_k(\mathbf{r}_1,\ldots,\mathbf{r}_N)\simeq\sum_{\mathbf{k}:|\mathbf{k}|<K,\mathbf{k}\neq\mathbf{0}}\frac{\exp(-\pi^2\mathbf{k}^2/\alpha^2)}{\pi\mathbf{k}^2}Z(\mathbf{k})\overline{Z(\mathbf{k})}, \tag{5}$$

where

$$Z(\mathbf{k})=\sum_{j=1}^{N}q_j\exp(2\pi i\mathbf{k}\cdot\mathbf{r}_j), \tag{6}$$

and where the overline indicates complex conjugation. Furthermore, for each pair of charges i,j only the minimum image term $\mathbf{n}_{ij}$ of the $\mathbf{n}$ sum of E_r is used. Terms where $|\mathbf{r}_i-\mathbf{r}_j+\mathbf{n}_{ij}|$ is bigger than an accuracy dependent cutoff distance r_c may be neglected. We can thus write this term as

$$E_r(\mathbf{r}_1,\ldots,\mathbf{r}_N)\simeq\sum_{1\leqslant i<j\leqslant N}^{*}\frac{\operatorname{erfc}(\alpha|\mathbf{r}_i-\mathbf{r}_j+\mathbf{n}_{ij}|)}{|\mathbf{r}_i-\mathbf{r}_j+\mathbf{n}_{ij}|}, \tag{7}$$

where the asterisk indicates that terms where $|\mathbf{r}_i-\mathbf{r}_j+\mathbf{n}_{ij}|>r_c$ may be neglected and where $r_c<\frac{1}{2}$ depends on α and the required accuracy. The corresponding forces are

$$\mathbf{F}_j=-\frac{\partial E}{\partial\mathbf{R}_j}=-\frac{1}{L^2}\left(\frac{\partial E_r}{\partial\mathbf{r}_j}+\frac{\partial E_k}{\partial\mathbf{r}_j}+\frac{\partial E_D}{\partial\mathbf{r}_j}\right). \tag{8}$$

As stated in the Introduction, accurate estimates of Δf due to the truncations at r_c and at K have been found for the Ewald method.[1] Based on the assumption, that the radial distribution function is unity at distances larger than r_c, the following estimate for the error on the E_r forces alone is obtained in Ref. 1:

$$\Delta f_r^{\mathrm{Ew}}\simeq\frac{2Q^2}{(NR_cL^3)^{1/2}}\exp\left(-\frac{\alpha^2R_c^2}{L^2}\right), \tag{9}$$

where we have switched back to absolute coordinates by writing $R_c=Lr_c$. Assuming that the radial distribution function is approximately unity at all distances, the following error estimate for the E_k forces alone are obtained in Ref. 1:

$$\Delta f_k^{\mathrm{Ew}}\simeq\frac{Q^2\alpha}{\pi L^2}\left(\frac{8}{KN}\right)^{1/2}\exp(-\pi^2K^2/\alpha^2). \tag{10}$$

Assuming instead, that the charges are just distributed *randomly* over the simulation box, an error estimate for the E_k force error can be obtained more simply. As a closely related evaluation will be used to obtain the PME errors, we shall briefly go through this evaluation.

Notice first, that the root mean square expectation of the error due to a single (jm) pair interaction can be written as

$$\langle\Delta_{jm}^2\rangle^{1/2}=\frac{2}{L^2}|q_j||q_m|\left[\sum_{\mathbf{k},|\mathbf{k}|>K}\frac{\exp(-2\pi^2|\mathbf{k}|^2/\alpha^2)}{|\mathbf{k}|^2}\right]^{1/2}.$$

We then use the approximation

$$\begin{aligned}&\sum_{\mathbf{k},|\mathbf{k}|>K}\frac{\exp(-2\pi^2|\mathbf{k}|^2/\alpha^2)}{|\mathbf{k}|^2}\\ &\quad\simeq 4\pi\int_{k>K}\exp(-2\pi^2k^2/\alpha^2)dk\\ &\quad\simeq\frac{\alpha^2}{\pi K}\exp(-2\pi^2K^2/\alpha^2),\end{aligned}$$

where the last expression is obtained by using an asymptotic expansion for the complementary error function. Assuming independence of all the Δ_{jm} terms, we then get

$$\Delta f_k^{\mathrm{Ew}} \simeq \sqrt{\frac{1}{N}\sum_{j=1}^{N}\sum_{m=1,m\neq j}^{N}\langle \Delta_{jm}^2\rangle}$$

$$\simeq \frac{2Q^2\alpha}{L^2}\left(\frac{1}{\pi KN}\right)^{1/2}\exp(-\pi^2K^2/\alpha^2). \tag{11}$$

This estimate is a factor $\sqrt{\pi/2} \approx 1.25$ bigger than the estimate in Eq. (10). The authors in Ref. 1 claim that they have a small systematic underestimate, which may indicate that this error estimate should be used instead. Anyway, both estimates are very accurate.

Assuming that the single particle error contributions from E_r and E_k are independent (which is to be expected), we can write Δf as

$$\Delta f = \sqrt{\Delta f_r^2 + \Delta f_k^2}. \tag{12}$$

III. OUTLINE OF THE PME METHOD

Consider $p+1$ different real numbers $x_0,\ldots,x_p$ all belonging to a closed interval I. We can then uniquely define polynomials $\kappa_n^{(p)}(x)$; $n=0,\ldots,p$ of order p such that

$$\kappa_n^{(p)}(x_n)=1;\quad n=0,\ldots,p,$$

$$\kappa_n^{(p)}(x_m)=0;\quad m,n=0,\ldots,p;\quad m\neq n.$$

Now consider a function $f\in C^{m+1}(I)$ (that is, f is continuous and has continuous $m+1$ derivatives on I). The classical Lagrange interpolation of f using the points $x_0,\ldots,x_p$ is then

$$f(x)\simeq\sum_{n=0}^{p} f(x_n)\kappa_n^{(p)}(x)\equiv P(x). \tag{13}$$

When $x\in I$, the accuracy of the Lagrange interpolation can be written as

$$f(x)-P(x)=\frac{f^{(p+1)}(\xi)}{(p+1)!}\prod_{i=0}^{p}(x-x_i) \tag{14}$$

for some $\xi\in I$.

The PME method heavily relies on the usage of the Lagrange interpolation scheme. Assume thus, that we wish to calculate E_k from N charges residing in a cubic box centered at $(\frac{1}{2},\frac{1}{2},\frac{1}{2})$. We then define a mesh on the box

$$\mathbf{s}_{klm}\equiv(kH,lH,mH);\quad 0\leqslant k,l,m\leqslant 2^A-1;$$

where A is an integer and $H=2^{-A}$ is the grid size (we shall view H as a dimensionless number measured in units of the box length throughout this paper). The idea of the PME method is to assign representative charge contributions to the mesh and then use these contributions in a Lagrange interpolation scheme to approximate E_k and the corresponding forces.

We shall first outline how to assign contributions to the mesh from the jth charge q_j residing at $\mathbf{r}_j=(x_j,y_j,z_j)$. Let $S_H^p(\alpha_j)(\alpha\in\{x,y,z\})$ be defined as the subset of the set $\{kH;0\leqslant k\leqslant 2^A-1\}$ containing the $p+1$ numbers which are closest to α_j (periodic boundary conditions are used). We then define the contributions from q_j to be zero for all vectors $\mathbf{s}_{klm}$ in the mesh except for the $(p+1)^3$ vectors satisfying $[\mathbf{s}_{klm}]_\alpha\in S_H^p(\alpha_j)$; $\alpha=x,y,z$. We shall denote the set of nonzero contributions by $\mathbf{S}_H^p(\mathbf{r}_j)$. To find the nonzero contributions for the points in $\mathbf{S}_H^p(\mathbf{r}_j)$, we use a product assignment scheme. Define polynomials $\kappa_k(x)$ such that

$$\kappa_k(s_k)=1;\quad s_k\in S_H^p(x_j),$$

$$\kappa_k(s_{k'})=0;\quad s_{k'}\in S_H^p(x_j);\quad k'\neq k.$$

Define similarly polynomials $\kappa_l(y)$ and $\kappa_m(z)$. (In Appendix A these polynomials are given for $p\leqslant 6$.) We then define the contributions to points in $\mathbf{S}_H^p(\mathbf{r}_j)$ as

$$Q_{H,j}^{(p)}(\mathbf{s}_{klm})=q_j\kappa_k(x_j)\kappa_l(y_j)\kappa_m(z_j). \tag{15}$$

Having made these assignment for all the charges, we can by defining

$$Q_H^{(p)}(\mathbf{s}_{klm})=\sum_{j=1}^{N} Q_{H,j}^{(p)}(\mathbf{s}_{klm}) \tag{16}$$

use the Lagrange interpolation scheme to approximate E_k by

$$E_k(\mathbf{r}_1,\ldots,\mathbf{r}_N)\simeq\sum_{\mathbf{k}\in C,\mathbf{k}\neq\mathbf{0}}\frac{\exp(-\pi^2\mathbf{k}^2/\alpha^2)}{2\pi\mathbf{k}^2}\left|\sum_{k,l,m=0}^{2^A-1} Q_H^{(p)}\times(\mathbf{s}_{klm})\exp(2\pi i\mathbf{k}\cdot\mathbf{s}_{klm})\right|^2, \tag{17}$$

where C denotes the set of integer tuples

$$\{(k_x,k_y,k_z)|-(2^{A-1})\leqslant k_x,k_y,k_z<2^{A-1}\}.$$

Correspondingly, we can approximate the force terms

$$\frac{\partial E_k}{\partial\mathbf{r}_j}(\mathbf{r}_1,\ldots,\mathbf{r}_N)=-2\sum_{\mathbf{k}\in Z^3,\mathbf{k}\neq\mathbf{0}}\frac{\exp(-\pi^2\mathbf{k}^2/\alpha^2)}{\mathbf{k}^2}\mathbf{k}\times\sum_{l=1}^{N} q_jq_l\,\mathrm{Im}\{\exp[2\pi i\mathbf{k}\cdot(\mathbf{r}_j-\mathbf{r}_l)]\} \tag{18}$$

by

$$\frac{\partial E_k}{\partial\mathbf{r}_j}(\mathbf{r}_1,\ldots,\mathbf{r}_N)\simeq-2\sum_{\mathbf{k}\in C,\mathbf{k}\neq\mathbf{0}}\frac{\exp(-\pi^2\mathbf{k}^2/\alpha^2)}{\mathbf{k}^2}\mathbf{k}\times\sum_{\mathbf{s}_{k_1,l_1,m_1}\in\mathbf{S}^{(p)}(\mathbf{R}_j)} Q_{H,j}^{(p)}(\mathbf{s}_{k_1l_1m_1})\sum_{k_2,l_2,m_2=0}^{2^A-1} Q_H^{(p)}\times(\mathbf{s}_{k_2l_2m_2})\mathrm{Im}\{\exp[2\pi i\mathbf{k}\cdot(\mathbf{s}_{k_1l_1m_1}-\mathbf{s}_{k_2l_2m_2})]\}. \tag{19}$$

The important advantage of this way of computing E_k is, that the charge assignment is an $\mathcal{O}(N[p+1]^3)$ procedure and that the evaluation of E_k and the forces by these approximations can be made using three dimensional fast Fourier transforms.

The overall structure of the PME method is the following:

(1) Calculate $Q_{H,j}^{(p)}(\mathbf{s}_{klm})$; $j=1,\ldots,N$ and then calculate $Q_H^{(p)}(\mathbf{s}_{klm})$.
(2) Calculate $\Sigma_{\mathbf{s}_{klm}}Q_H^{(p)}(\mathbf{s}_{klm})\exp(2\pi i\mathbf{k}\cdot\mathbf{s}_{klm})\equiv e(\mathbf{k})$ for all $\mathbf{k}\in C$ by a three dimensional fast Fourier transform.

(3) Calculate $\underline{E_k}(\mathbf{r}_1,\ldots,\mathbf{r}_N)\simeq\Sigma_{\mathbf{k}\in C,\mathbf{k}\neq\mathbf{0}}\exp(-\pi^2\mathbf{k}^2/\alpha^2)/2\pi\mathbf{k}^2e(\mathbf{k})e(\mathbf{k})$.
(4) Calculate $\phi(\mathbf{k})=-2\exp(-\pi^2\mathbf{k}^2/\alpha^2)/\mathbf{k}^2\overline{\mathbf{k}e(\mathbf{k})};\quad \mathbf{k}\neq\mathbf{0}$ and set $\phi(0)=0$.
(5) Calculate all $\mathbf{f}_{klm}\simeq\Sigma_{\mathbf{k}\in C}\phi(\mathbf{k})\exp(2\pi i\mathbf{k}\cdot\mathbf{s}_{klm})$ by another three three-dimensional fast Fourier transforms.
(6) Calculate $(\partial E_k/\partial\mathbf{r}_j)\simeq\Sigma_{\mathbf{s}_{klm}\in\mathbf{S}^{(p)}(\mathbf{r}_j)}Q^{(p)}_{H,j}(\mathbf{s}_{klm})\mathbf{f}_{klm}$.

The real space part of the PME method is evaluated in the same way as for the Ewald method. It may therefore seem, that only the reciprocal space overhead is reduced when using the PME method. We shall however already at this point stress, that the value of α should be chosen differently for the two methods (bigger for the PME method) because of the decrease in time overhead for the reciprocal space part when using the PME method. Therefore, we also expect a reduction in the real-space overhead. We shall return to this issue in much more detail in Sec. VI.

IV. ESTIMATE OF THE ACCURACY

As for the Ewald method, we shall use the estimate

$$\Delta f=\left(\frac{1}{N}\sum_{j=1}^{N}\Delta f_j^2\right) \tag{20}$$

for the average accuracy. Obviously, we can reuse the formula

$$\Delta f_r^{\mathrm{PME}}=\Delta f_r^{\mathrm{Ew}}=\frac{2Q^2}{(NR_cL^3)^{1/2}}\exp\left(-\frac{\alpha^2R_c^2}{L^2}\right) \tag{21}$$

for the real space accuracy. We shall now find an estimate for $\Delta f_k^{\mathrm{PME}}$. That is, the average accuracy for the E_k forces when evaluated using the PME method.

By making the reasonable assumption that contributions to the E_k-forces from $\mathbf{k}$-values not in C are negligible, we can write the error on the E_k-force on particle j due to particle m as

$$\Delta\mathbf{f}_{jm}=\frac{2}{L^2}\,\mathrm{Im}\left[\sum_{\mathbf{k}\in C,\mathbf{k}\neq\mathbf{0}}\frac{\exp(-\pi^2\mathbf{k}^2/\alpha^2)}{\mathbf{k}^2}\mathbf{k}h_{jm}(\mathbf{k})\right], \tag{22}$$

where

$$\begin{aligned}h_{jm}(\mathbf{k})=&-q_jq_m\exp(2\pi i\mathbf{k}\cdot[\mathbf{r}_j-\mathbf{r}_m])\\&+\sum_{\mathbf{s}_{k_1,l_1,m_1}\in\mathbf{S}_H^{(p)}(\mathbf{r}_j)}Q^{(p)}_{H,j}(\mathbf{s}_{k_1l_1m_1})\\&\times\sum_{\mathbf{s}_{k_2,l_2,m_2}\in\mathbf{S}_H^{(p)}(\mathbf{r}_m)}Q^{(p)}_{H,m}(\mathbf{s}_{k_2l_2m_2})\\&\times\exp(2\pi i\mathbf{k}\cdot[\mathbf{s}_{k_1l_1m_1}-\mathbf{s}_{k_2l_2m_2}]).\end{aligned}$$

Using Eq. (15) for $Q^{(p)}_{H,j}$ and $Q^{(p)}_{H,m}$ allow us to write [inspired by Eq. (14)]

$$\begin{aligned}h_{jm}(\mathbf{k})=&\left(\prod_{\beta=x,y,z}\exp(2\pi i[\mathbf{k}]_\beta[\mathbf{r}_j]_\beta)\left[1-\frac{(2\pi iH[\mathbf{k}]_\beta)^{p+1}}{(p+1)!}Z_H^p([\mathbf{k}]_\beta,[\mathbf{r}_j]_\beta)\phi_p([\mathbf{r}_j]_\beta)\right]\prod_{\beta=x,y,z}\exp(-2\pi i[\mathbf{k}]_\beta[\mathbf{r}_m]_\beta)\right.\\&\left.\times\left\{1-\frac{(-2\pi iH[\mathbf{k}]_\beta)^{p+1}}{(p+1)!}Z_H^p(-[\mathbf{k}]_\beta,[\mathbf{r}_m]_\beta)\phi_p([\mathbf{r}_m]_\beta)\right\}\right)-\exp(2\pi i\mathbf{k}\cdot[\mathbf{r}_j-\mathbf{r}_m]),\end{aligned} \tag{23}$$

where $H=2^{-A}$ is the grid size and

$$\phi_p(x)=H^{-(p+1)}\prod_{s\in S_H^{(p)}(x)}(x-s)$$

is a grid size independent function and where $Z_H^p(a,x)$ is the complex number satisfying

$$\begin{aligned}&\frac{(2\pi iaH)^{p+1}}{(p+1)!}\exp(2\pi iax)Z_H^p(a,x)\phi_p(x)\\&\quad=\exp(2\pi iax)-\sum_{s_k\in S_H^p(x)}\kappa_k(s_k)\exp(2\pi ias_k).\end{aligned} \tag{24}$$

Neglecting all but the dominant terms on the right-hand side of Eq. (23) yields

$$\Delta\mathbf{f}_{jm}\simeq\frac{2q_jq_m}{L^2}\,\mathrm{Im}\left[\frac{(2\pi iH)^{p+1}}{(p+1)!}\sum_{\beta=x,y,z}\mathbf{J}_\beta^p(\mathbf{r}_j,\mathbf{r}_m)\right], \tag{25}$$

where

$$\begin{aligned}\mathbf{J}_\beta^p(\mathbf{r}_j,\mathbf{r}_m)=&\sum_{\mathbf{k}\in C,\mathbf{k}\neq\mathbf{0}}\frac{\exp(-\pi^2\mathbf{k}^2/\alpha^2)}{\mathbf{k}^2}\mathbf{k}\\&\times\exp[2\pi i\mathbf{k}\cdot(\mathbf{r}_j-\mathbf{r}_m)][\mathbf{k}]_\beta^{p+1}\\&\times\{Z_H^p(H\mathbf{k}_\beta,[\mathbf{r}_j]_\beta)\phi_p([\mathbf{r}_j]_\beta)\\&+(-1)^{p+1}Z_H^p(-H\mathbf{k}_\beta,[\mathbf{r}_m]_\beta)\phi_p([\mathbf{r}_m]_\beta)\}.\end{aligned} \tag{26}$$

We shall now derive an estimate for $\langle|\mathbf{J}_\beta^p|^2\rangle$. This involves the term $|Z_H^p(a,x)|^2$. As the right-hand side of Eq. (24) is bounded for any value of a, we get that $Z_H^p(a,x)\to0$ as $a\to\infty$. From Eq. (14) we obtain that $Z_H^p(a,x)\to1$ as $a\to0$. It is therefore reasonable to assume that $|Z_H^p(a,x)|$ *on average* will decay monotonically from one to zero as a goes from zero to infinity. A closer study of those a-values which are of importance indicates that we may use $Z_H^p=1$ in the following estimation of $\langle|\mathbf{J}_\beta^p|^2\rangle$ and expect to obtain a reasonably good estimate (we shall give a discussion of the validity of this choice of Z_H^p in Sec. V). We then get

$$\langle|\mathbf{J}_\beta^p|^2\rangle = 2[\langle\phi_p^2\rangle + \langle\phi_p\rangle^2] \times \sum_{\mathbf{k}\in C,\mathbf{k}\neq\mathbf{0}} \frac{\exp(-2\pi^2\mathbf{k}^2/\alpha^2)}{\mathbf{k}^2}[\mathbf{k}]_\beta^{2(p+1)}. \quad (27)$$

Assume now that p is even. For symmetry reasons we then have $\langle\phi_p\rangle^2=0$. We then easily see that $\langle J_\beta^p \overline{J_\gamma^p}\rangle = 0$ when $\beta\neq\gamma$ and by using that $\langle|\mathbf{J}_x^p|^2\rangle=\langle|\mathbf{J}_y^p|^2\rangle=\langle|\mathbf{J}_z^p|^2\rangle\equiv\langle|\mathbf{J}^p|^2\rangle$, we get

$$\langle|\Delta\mathbf{f}_{jm}|^2\rangle^{1/2} = 2\sqrt{3}\,\frac{|q_j||q_m|}{L^2}\,\frac{(2\pi H)^{p+1}}{(p+1)!}\,\langle|\mathbf{J}^p|^2\rangle^{1/2}, \quad (28)$$

where

$$\langle|\mathbf{J}^p|^2\rangle = 2\langle\phi_p^2\rangle \sum_{\mathbf{k}\in C,\mathbf{k}\neq\mathbf{0}} \frac{\exp(-2\pi^2\mathbf{k}^2/\alpha^2)}{\mathbf{k}^2}[\mathbf{k}]_\beta^{2(p+1)}. \quad (29)$$

The value of $\langle\phi_p^2\rangle$ can be obtained by direct integration. We thus only need an estimate for

$$\sum_{\mathbf{k}\in C,\mathbf{k}\neq\mathbf{0}} \frac{\exp(-2\pi^2\mathbf{k}^2/\alpha^2)}{\mathbf{k}^2}[\mathbf{k}]_\beta^{2(p+1)}.$$

Replacing the sum by an integral and using polar coordinates allow us to estimate this term by

$$2\pi\left(\frac{\alpha}{\sqrt{2}\pi}\right)^{2p+3}\int_{y=0}^{\infty} y^{2(p+1)}\exp(-y^2)dy \times \int_{\theta=0}^{\pi}\sin\theta\cos^{2(p+1)}\theta d\theta.$$

The θ-integration is performed directly whereas the y-integral is accurately approximated by an expansion at the maximum of $y^{2(p+1)}\exp(-y^2)$. We then get

$$\langle|\mathbf{J}^p|^2\rangle^{1/2} \simeq \sqrt{\frac{2\alpha}{2p+3}}\,\pi^{1/4}\left(\frac{\alpha}{\pi}\right)^{p+1} \times\exp\{(p+1)[\log(p+1) -\log 2-1]/2\}\langle\phi_p^2\rangle. \quad (30)$$

Inserting Eq. (30) into Eq. (28) allows us to estimate $\langle|\Delta\mathbf{f}_{jm}|^2\rangle^{1/2}$ by

$$\langle|\Delta\mathbf{f}_{jm}|^2\rangle^{1/2} = 2\pi^{1/4}\sqrt{\frac{6\alpha}{2p+3}}\frac{|q_j||q_m|}{L^2}\frac{(2\alpha H)^{p+1}}{(p+1)!} \times\exp\left\{\frac{(p+1)[\log(p+1)-\log 2-1]}{2}\right\}\langle\phi_p^2\rangle. \quad (31)$$

In turn, we get

$$\Delta f_k^{\mathrm{PME}} = \left(\frac{1}{N}\sum_{j=1}^{N}\sum_{m=1,m\neq j}^{N}\langle|\Delta\mathbf{f}_{jm}|^2\rangle\right)^{1/2} = 2\pi^{1/4}\sqrt{\frac{6\alpha}{N(2p+3)}}\frac{Q^2}{L^2}\frac{(2\alpha H)^{p+1}}{(p+1)!} \times\exp\left\{\frac{(p+1)[\log(p+1)-\log 2-1]}{2}\right\}\langle\phi_p^2\rangle^{1/2}. \quad (32)$$

Consider now the case where p is odd. We then easily observe that $\langle\phi_p\rangle^2>0$ and that $\langle J_\beta^p \overline{J_\gamma^p}\rangle>0$ also when $\gamma\neq\beta$. This give rise to additional contributions to the error, which we shall not estimate.

TABLE I. Lennard-Jones parameters.

Interaction	σ (Å)	ϵ (J/mol)
Na–Na	2.744	5.03
Na–Cl	3.047	4.12
Cl–Cl	3.350	5.03

V. EXPERIMENTAL MEASUREMENTS

A. Accuracy measurements: Validity of the error estimates

The aim of this section is to use experimental measurements to check the validity of the error estimate found in the previous sections.

We have applied the method to a molten NaCl consisting of $N=10\,000$ point charges (5000 Na^+ and 5000 Cl^- ions). The van der Waals interactions were modeled by the Lennard-Jones pair potential

$$u(r)=4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12}-\left(\frac{\sigma}{r}\right)^{6}\right],$$

where σ and ϵ were derived from the CHARMM parameters[9] and is shown in Table I.

The system was equilibrated at a temperature of $T=2000$ K and at a density of $\rho=0.0495A^{-3}$ before the error measurements took place. In Tables II and III, we have listed theoretical and experimental absolute values of the Δf_k-values for the Ewald method and the PME method. The errors are divided by the root mean square force

$$\langle f^2\rangle^{1/2}\equiv\left(\frac{1}{N}\sum_{j=1}^{N} f_j^2\right)=0.583q^2(A^{-2}).$$

As the error behavior of the Ewald method has been studied in detail in Ref. 1, we only list results from a few runs to

TABLE II. Errors obtained from Ewald simulations and the estimates of Eq. (9) and Eq. (11) for different cutoff distances and maximum k-values. The value of α is 10 everywhere.

$R_c(A)$	$\Delta f_r^{\mathrm{Ew}}/\langle f^2\rangle$ (simul.)	$\Delta f_r^{\mathrm{Ew}}/\langle f^2\rangle$ (estim.)	γ
9.78	1.52×10^{-2}	1.73×10^{-2}	2.7
11.7	4.44×10^{-3}	4.08×10^{-3}	2.7
14.7	4.28×10^{-4}	4.05×10^{-4}	2.8
K	$\Delta f_k^{\mathrm{Ew}}/\langle f^2\rangle$ (simul.)	$\Delta f_k^{\mathrm{Ew}}/\langle f^2\rangle$ (estim.)	γ
6	5.74×10^{-3}	6.57×10^{-3}	2.8
8	3.71×10^{-4}	3.59×10^{-4}	2.6
10	9.02×10^{-6}	9.22×10^{-6}	2.6

TABLE III. (a) Errors obtained from PME simulations and the estimates of Eq. (32) for different values of α and p. The value of H is $L/32$ everywhere. (b) Errors obtained from PME simulations and the estimates of Eq. (32) for different values of α and p. The value of H is $L/64$ everywhere.

	p	$\langle\phi_p^2\rangle^{1/2}$	α	$\Delta f_k^{\mathrm{PME}}/\langle f^2\rangle$ (simul.)	$\Delta f_k^{\mathrm{PME}}/\langle f^2\rangle$ (estim.)	γ
(a)	2	0.246	10	2.86×10^{-3}	2.25×10^{-3}	3.0
			15	8.65×10^{-3}	9.32×10^{-3}	2.9
	3	0.404	10	9.87×10^{-4}	6.74×10^{-4}	2.9
			15	5.19×10^{-3}	4.18×10^{-3}	3.1
	4	0.950	10	2.48×10^{-4}	2.68×10^{-4}	3.0
			15	1.94×10^{-3}	2.50×10^{-3}	3.0
	5	2.51	10	1.65×10^{-4}	1.13×10^{-4}	3.1
			15	1.43×10^{-3}	1.57×10^{-3}	3.0
	6	8.42	10	4.58×10^{-5}	5.66×10^{-5}	3.1
			15	7.00×10^{-4}	11.8×10^{-5}	3.0
(b)	2	0.246	10	2.86×10^{-4}	2.82×10^{-4}	3.1
			15	1.12×10^{-3}	1.17×10^{-3}	2.9
			20	2.87×10^{-3}	3.19×10^{-3}	3.0
	3	0.404	10	6.43×10^{-5}	4.21×10^{-5}	3.1
			15	3.63×10^{-4}	2.61×10^{-4}	3.0
			20	12.0×10^{-4}	9.53×10^{-4}	3.0
	4	0.950	10	8.93×10^{-6}	8.39×10^{-6}	2.9
			15	7.08×10^{-5}	7.80×10^{-5}	3.0
			20	3.18×10^{-4}	3.80×10^{-4}	2.9
	5	2.51	10	3.32×10^{-6}	1.76×10^{-6}	3.0
			15	2.97×10^{-5}	2.45×10^{-5}	3.0
			20	1.61×10^{-4}	1.59×10^{-4}	3.0
	6	8.42	10	4.41×10^{-7}	4.42×10^{-7}	2.9
			15	7.81×10^{-6}	9.25×10^{-6}	3.2
			20	5.94×10^{-5}	8.01×10^{-5}	3.1

verify the validity of the Ewald estimates. The results are shown in Table II. As can be seen, the estimates for both Δf_r and Δf_k are reasonably to very accurate.

In Tables III(a) and III(b), we show the estimated and experimental values of $\Delta f_k^{\mathrm{PME}}$ for $H=L/32$ and $H=L/64$, respectively. Consider first the case, where p is even. For small values of α, the estimates are all very accurate, but for large α-values, the estimates are too high. This was to be expected as the important contributions in the k-summation of Eq. (29) come from k-values in the neighborhood of the peak of the summand, that is near $k \simeq (\alpha\sqrt{p+1}/\pi)$. At large values of α, the peak thus occurs at large values of k, and thus the average value of Z_H^p becomes small. As the peak also grows with p, it is also to be expected that the discrepancy should be bigger for large p-values, which is also (at least vaguely) seen from the experimental results.

The results for odd p-values show that the experimental errors are in general bigger than the "estimate." This was also to be expected because of the nonzero $\langle\phi_p\rangle$ values. We shall therefore only consider simulations where p is even throughout the remaining part of this article.

When the single particle force errors are independent Gaussian random variables, the (N-dependent) expected γ value is (for $N=10\,000$),

$$\gamma=2.72\pm0.16.$$

The γ-values found in the Ewald simulations are all close to this Gaussian value. This was also found in Refs. 1 and 13.

The values of γ found in the PME-simulations are a little higher. Anyway, they are close to the Gaussian value, and thus the PME method will not suffer from large maximum error disadvantages occurring in for example the fast multipole method as stated in the Introduction.

B. Timing experiments

To compare the PME-method with the Ewald method, we need to set up equations for the overall computation time for the two methods. These equations will contain hardware and implementation dependent constants. We have estimated these constants by time experiments.

For the reason of simplification, we only include the most significant overheads in the model. The overall computation time for the Ewald method is thus

$$T_{\mathrm{Ew}}=a_{SR}N^2\left(\frac{R_c}{L}\right)^3+a_FNK^3, \tag{33}$$

where r_c is the cutoff distance and K is the maximum length of the Fourier-space $\mathbf{k}$ vector. The overall computation time for the PME-method may be written as

$$T_{\mathrm{PME}}=a_{SR}N^2\left(\frac{R_c}{L}\right)^3+a_p\left(\frac{1}{H}\right)^3\left[\frac{\log_2(1/H)}{3}\right]^3+a_AN(p+1)^3, \tag{34}$$

where the second term denotes the fast Fourier transform overhead and the third term denotes the overhead for assigning charges to the mesh and for evaluating the forces due to the charges from the mesh values. The primitive overheads a_{SR}, a_F, a_p, and a_A are to be found from experiments. The relative performance of the Ewald method and the PME method is of course highly dependent on these primitive overheads. It is therefore important to use comparable effort in the optimization of the code for the two methods. In this implementation the short-ranged forces were calculated using the linked-cell method[8] and the complementary error function and its derivative was calculated using table lookup. The Fourier-space sum of the Ewald method was calculated using a highly optimized code[6,7] where only one cosine and one sine evaluation is needed for each particle. The fast Fourier transform method in Ref. 21 was applied unchanged in the PME program. The time experiments were carried out on a SPARC station ELC (based on a 33 MHz CMOS SPARC RISC processor with integrated floating point unit) using a standard Fortran 77 compiler with no optimization options set (setting these yields a speedup of about a factor 2). The following values for the primitive times were found:

$$\begin{aligned} a_{SR}&=125\ \mu\mathrm{s},\\ a_F&=32\ \mu\mathrm{s},\\ a_p&=64\ \mu\mathrm{s},\\ a_A&=104\ \mu\mathrm{s}, \end{aligned} \tag{35}$$

where a_p and a_A have been multiplied by 4 to account for the four necessary FFT's and assignments (see end of Sec. III).

VI. SCALING PROJECTIONS

In this section, we shall use the models for the truncation errors and the execution times derived in the previous sections to analyze and compare the performance of the Ewald method and the PME method. What we need is, to find the parameters of the two methods that for given N and required accuracy δ satisfies the accuracy constraints and minimizes the computation time.

We first derive a rather qualitative theoretical model for the optimal parameters and the performance showing essentially only the N-dependence.

A. The Ewald method

We need to find the parameters α, K, and R_c that minimizes

$$T_{\mathrm{Ew}}=a_{SR}N\rho R_c^3+a_F NK^3 \tag{36}$$

with respect to the constraints of the error bounds [Eqs. (9) and (11)], which we restate as

$$\begin{aligned}\frac{\delta}{\sqrt{2}}&=\frac{2Q^2}{N}\left(\frac{\rho}{R_c}\right)^{1/2}\exp(-\alpha^2R_c^2[\rho/N]^{2/3}),\\ \frac{\delta}{\sqrt{2}}&=2\,\frac{Q^2}{N^{7/6}}\,\alpha\rho^{2/3}\left(\frac{1}{\pi K}\right)^{1/2}\exp(-\pi^2K^2/\alpha^2).\end{aligned} \tag{37}$$

For given δ and N this optimization problem can be viewed as optimizing the time as function of α by obtaining the functions $R_c(\alpha)$ and $K(\alpha)$ and inserting into Eq. (36). We can make a qualitative study of the behavior of the method as function of N similar to the one in Ref. 6 by noticing that $R_c(\alpha)\simeq(A/\alpha)(N/\rho)^{1/3}$ and $K(\alpha)\simeq B\alpha$. Inserting into Eq. (36) and differentiating yields $\alpha\propto N^{1/6}$ and thus $R_c\propto N^{1/6}$ and the optimal computation time is then proportional to $\sqrt{a_{SR}a_F}N^{3/2}$, where the proportionality constant depends on the accuracy.

B. The PME method

We here need to find the parameters α, H, R_c, and p that minimizes

$$T_{\mathrm{PME}}=a_{SR}N\rho R_c^3+a_p\left(\frac{1}{H}\right)^3\left[\frac{\log_2(1/H)}{3}\right]^3+a_AN(p+1)^3 \tag{38}$$

with respect to the constraints of the error bounds [Eqs. (9) and (31)], which we restate as

$$\begin{aligned}\frac{\delta}{\sqrt{2}}&=\frac{2Q^2}{N}\left(\frac{\rho}{R_c}\right)^{1/2}\exp(-\alpha^2R_c^2[\rho/N]^{2/3}),\\ \frac{\delta}{\sqrt{2}}&=2\,\frac{Q^2}{N^{7/6}}\,\pi^{1/4}\rho^{2/3}\sqrt{\frac{6\alpha}{(2p+3)}}\frac{(2\alpha H)^{p+1}}{(p+1)!}\\ &\quad\times\exp\left\{\frac{(p+1)[\log(p+1)-\log 2-1]}{2}\right\}\langle\phi_p^2\rangle^{1/2}.\end{aligned} \tag{39}$$

A qualitative analysis similar to the one above for the Ewald method yields $R_c(\alpha)\simeq(A/\alpha)(N/\rho)^{1/3}$ and $H(\alpha)\simeq C/\alpha$. Inserting into Eq. (38) and differentiating with respect to α yields the optimum, where $\alpha\propto[N^{1/3}/(\log_2 N)^{1/2}]$ and thus $R_c\propto(\log_2 N)^{1/2}$ and the optimal time should then be expected to grow as roughly $N(\log_2 N)^{3/2}$.

These functional dependencies are not sufficient to analyze the absolute performances of the two methods and to compare these because we have not estimated the proportionality constants and we have use a rather crude approximations for the functional dependencies of the parameters. Therefore we shall use the primitive time overheads derived in the previous section together with Eqs. (36)–(39) to *numerically* estimate the optimal parameters and in turn the overall computation time for a variety of accuracies and N-values. We have chosen a range of accuracies which should cover most applications ($\delta\simeq10^{-2}$ to $\delta\simeq10^{-5}$ measured in A^{-2}) and we have considered N-values ranging from $N=10^4$ to $N=10^6$ as this will be the range of charge numbers that can be simulated today or in the near future. We consider again a molten NaCl consisting of N charges at density $\rho=0.0495A^{-3}$.

We first discuss the resulting reciprocal space parameters. The optimal values of K for the Ewald method are ranging between 7 and 25 depending very little on the accuracy, but depending on N as about $N^{1/6}$ as expected. The ''corresponding'' values $1/H$ for the PME method was 64, 128, or 256 also depending mainly on the system size and is thus (as expected) much larger. Due to the strong requirement that $1/H$ has to be a power of two, it can hardly be concluded that $1/H$ is having the expected behavior. The optimal values of p was mainly depending on the accuracy and was $p=2$ or $p=4$ for $\delta=10^{-2}$; $p=4$ for $\delta=10^{-3}$; $p=4$ or $p=6$ for $\delta=10^{-4}$ and $p=6$ for $\delta=10^{-5}$. This clearly illustrates the importance of considering not only H but also p when optimizing performance at a given accuracy.

The remaining parameters of the two methods can be compared directly and we shall therefore show these graphically. In Figs. 1(a) and 1(b) we show the optimal values of α for the two methods and in Figs. 2(a) and 2(b) we show the corresponding optimal values R_c. As can be seen, we obtain the expected behavior of these parameters for the two methods. That is, the N dependence of the optimal α-values can be seen to be approximately $N^{1/6}$ for the Ewald method and $N^{1/3}$ for the PME method and correspondingly. In both cases the α-values hardly depends on the accuracies (the picture is somewhat obscured for the PME method because of the requirement that $1/H$ must be a power of 2). We get corresponding R_c-values which grow as approximately $N^{1/3}$ for the Ewald method and being hardly N-dependent over two decades for the PME method {not surprising as $[\log_2(10^6)/\log_2(10^4)]\simeq1.22$}. This illustrates clearly the power of the parameter optimized PME for large systems; the fast evaluation of the reciprocal space part allow optimal values of α bigger than in the Ewald method and in turn values of R_c that are smaller at least for $N>10\,000$. Therefore, not only the overhead from the reciprocal space part, but also the overhead from the real space part is reduced.

The corresponding overall computation time per particle

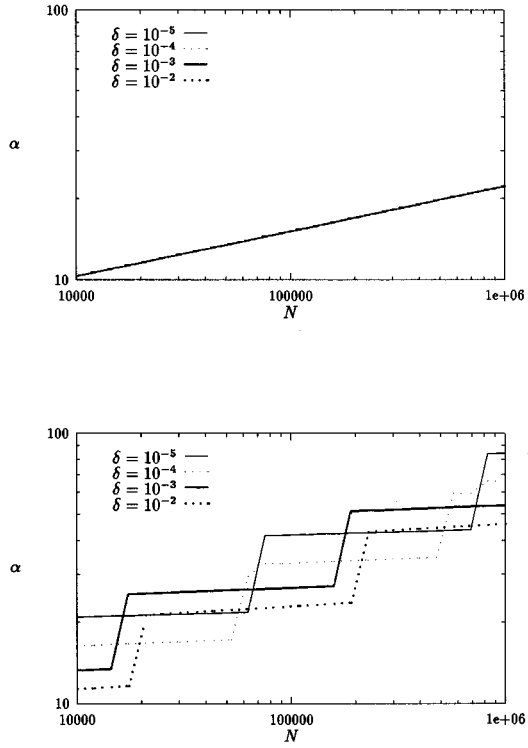

FIG. 1. (a) Optimal Ewald α-values. (b) Optimal PME α-values.

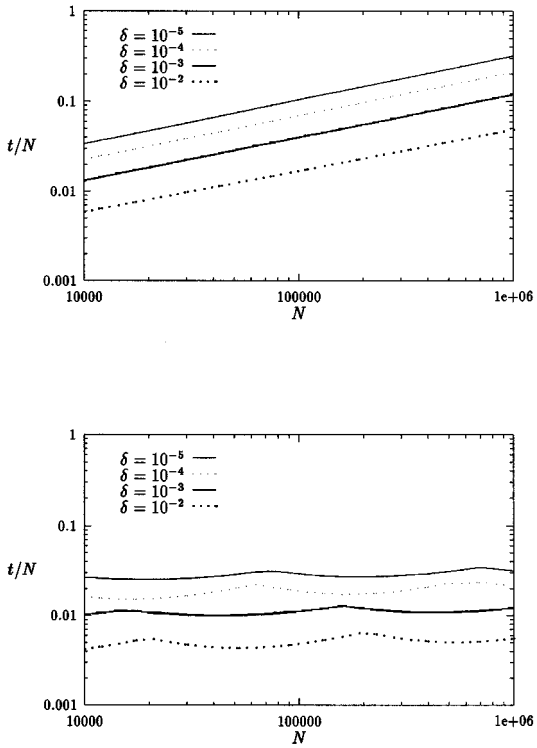

FIG. 3. (a) Optimal Ewald time overheads. (b) Optimal PME time overheads.

is shown as function of N in Figs. 3(a) and 3(b). As expected, the Ewald time grows as $N^{3/2}$. The PME time is growing much slower and although the N-dependence is again somewhat obscured by the requirement that $1/H$ has to be a power of 2 (notice that the changes in $1/H$ leads to first order discontinuities for the PME performance curves), it is at least vaguely indicated that the N-dependence is $N[\log_2(N)]^{3/2}$ as expected. It is however more important, that we can now directly compare the performances of the two methods when optimized. As can be seen, the PME method is already a little faster at 10 000 particles. For $N=100\,000$ the PME is several times faster than Ewald and for $N=1\,000\,000$ the PME outperforms Ewald with a factor of about 10 depending very little on the required accuracy.

Although the comparison was made for a specific density, it should be noticed that the reciprocal space errors for the two methods has the same density dependence and thus no big differences should be expected for other densities. We have however based the discussion on even distributions of charges. We shall therefore give a short discussion of the case of uneven distributions in the conclusion.

Someone may claim, that the performance results was very depending on the primitive time overheads a_{SR}, a_F, a_p, and a_A. Notice however, that ''errors'' in these values are partly suppressed by the optimizing procedure as the overall time when optimized typically is proportional only to the square root of these numbers. Furthermore, as stated in the last section, care was taken to write efficient code for both methods.

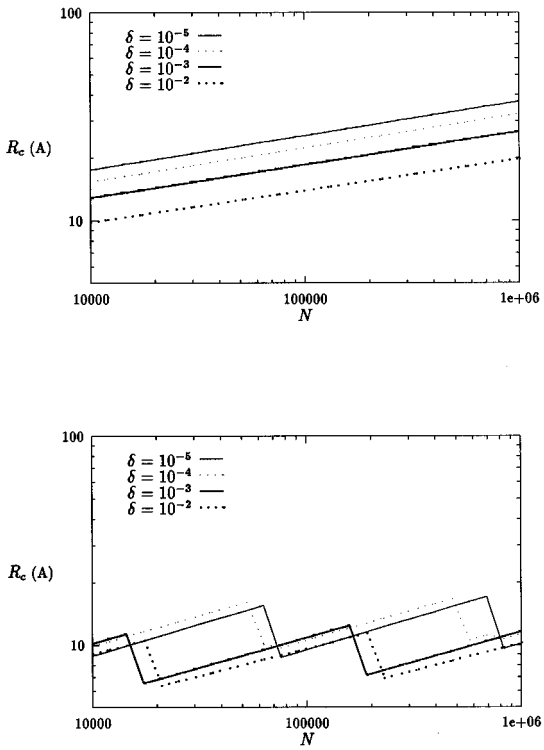

FIG. 2. (a) Optimal Ewald r_c-values. (b) Optimal PME r_c-values.

VII. SHORT DISCUSSION OF CONCURRENCY

Long accurate simulations of macromolecules or other systems of $N \simeq 10^5$–10^6 charges will only be realistic on a parallel computer. In a discussion of the efficiency of the

PME method for systems of that size, it is therefore important to discuss the potential concurrency of the PME method, that is, to discuss whether it can be applied efficiently to a parallel computer. Due to the limited access speed (and thus limited usability) when having one big shared memory storage for all processors, we shall only discuss the case where the memory is distributed among the processors. That is, when designing a parallel algorithm for doing the computation, not only the computation itself but also the related data should be distributed. The distribution of the computation and the data should obviously be made so as to balance the load on the processors and so as to minimize communication costs.

When searching for a scheme for the data distribution, one has to bear in mind that a complicated computation such as the computation of interactions in a molecular dynamics program may consist of inherently different independent subcomputational parts (an Ewald computation consists for example of the independent, but very different, computations of E_r and E_k). Considering each of the subcomputations individually may thus lead to optimal decomposition schemes which are very different. Using these different optimal decomposition schemes for each subcomputation will necessarily lead to inefficient overall performance due to large data communication overheads when switching between subcomputations. Thus, instead of searching for an optimal parallel algorithm for each subcomputation, one should rather search for one and only one distribution scheme for the data that somehow allow all subcomputations to be performed reasonably efficiently. For many computational problems, this is an extremely difficult task, especially when the optimal distribution scheme changes dynamically during the calculation.

However, for a molecular dynamics simulation of a system of approximately uniformly distributed charged atoms, a well-known distribution scheme for evaluating van der Waals (or other short-ranged) interactions is to decompose the charges geometrically according to their positions[20] (see Fig. 4). The advantage of such a decomposition is, that due to the small cutoff distance of such interactions, data on a given processor only need to be communicated to a limited number of "neighbor processors" to enable computation of all interactions. Furthermore, when the charges are evenly distributed, good balance of the load is obtained as there will be approximately the same number of charges on each processor. It should be noticed that such a decomposition scheme is dynamical in the sense that charge information has to be moved from one processor to another when the charge move across the corresponding boundary of the processor domain in the simulation box. This physically obvious geometrical data decomposition scheme has been shown to be well suited for the Ewald method[6] and for the fast multipole method.[13] We shall now briefly illustrate, that this geometrical decomposition scheme is also well suited for the PME method.

The computation of E_r and the corresponding forces can be viewed as a short-ranged interaction and thus fits obviously into the decomposition scheme. Concurrency in the E_r calculation is actually improved compared to the Ewald method because of the shorter range of the interactions due to the higher optimal α-values for the PME method. The

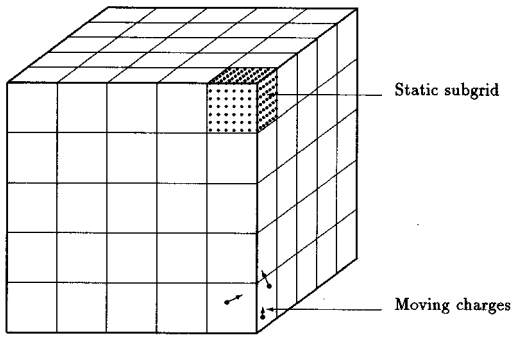

FIG. 4. Geometrical decomposition of a molecular dynamics computation. Geometrical decomposition of data on a 125 processor parallel computer. Each subbox illustrates the domain in space of the corresponding processor. When using the PME method, the grid (in this case $H=L/32$) is decomposed statically so that a given processor contains data concerning the subgrid which is inside the spatial domain of the processor. The charged particles are decomposed similarly. As the charges are moving during the simulation the information concerning a given charge has to be moved between processors when the charge crosses the boundary of a processor domain.

evaluation of charge assignments and the corresponding extractions of the energy and/or the forces can be distributed by statically decomposing the grid geometrically as well (see Fig. 4). Again charge information is then needed on a limited set of neighboring processors only and the load is well balanced. The final (and most difficult) task to consider is to decompose the fast Fourier transforms. A reasonably efficient distributed memory parallel algorithm for the FFT calculation using the same decomposition of the grid as for the assignments can be found in for example.[20] It should finally be noticed that as for the Ewald method the E_r and van der Waals computations may be evaluated completely concurrently with the reciprocal space evaluations (including the FFT and the real space charge assignments and corresponding energy/force extractions). Thus, an application to a parallel computer should not be expected to lead to any significant reduction in the PME performance with respect to the Ewald and/or FMM performances.

VIII. NONUNIFORM CHARGE DISTRIBUTIONS: CONCLUSION

The accuracy estimates and time overhead estimates of both the Ewald method and the PME method derived and used in the previous sections were based on the assumption, that the charges are almost randomly respectively uniformly distributed. For highly nonuniform distributions, these estimates can no longer be expected to be as precise. As many molecular dynamics simulations will be dealing with such systems, we shall give a brief study of these. Consider a model system, which still has an overall charge density ρ, but consists of large clouds of charges where the charge density inside the clouds is a constant ρ_C, where ρ_C is significantly bigger than ρ. By studying the derivation of the error estimates, it can be seen that the errors will grow by approximately a factor $\sqrt{\rho_C/\rho}$. Thus, the results of the previous sections may be reused by replacing ρ by ρ_C everywhere. Furthermore, the model for the time overheads of the real

space (and van der Waals) computations are no longer valid. These overheads should be expected to increase by a factor of roughly ρ_C/ρ.

As this behavior of the error occurs in all the methods discussed here (the PME method, the Ewald method and the FMM) and since short ranged interactions with similar overheads are present in other methods as well (in the FMM the direct computations from charges in the same cell or in neighbor cells are of this type), nonuniform distributions of this type should not be expected to yield any big change in the relative performances of these methods. We believe, that even for a more general nonrandom distribution the relative performances are approximately unchanged.

We have derived an accuracy estimate for the PME method for systems of randomly distributed charges. It was checked against accuracies obtained from simulations and found to be reasonably to very precise. Assuming a uniform charge distribution, a model of the overall time requirement of the PME method and the Ewald method as function of the method parameters and primitive times was given. Values for the primitive times were found from simulations and in turn the accuracy estimates were used to find the optimal parameters for a variety of accuracies. The PME method was compared to the Ewald method and a breakeven of $N \simeq 10\,000$ was found. As stated in the introduction it seems that the Ewald-FMM breakeven is at $N \simeq 100\,000$. Thus, although these breakevens are very sensitive to implementation efficiency and may be subject to small changes for nonuniform charge distributions and/or when using a parallel computer, it seems that the PME method outperforms both methods in the important range $N \simeq 10^4$–10^5. Together with the fact that the PME method is very easy to implement (charge assignments formulas can be found in the Appendix and efficient FFT packages are available on the market) and easy to optimize and control, we conclude that the PME method is extremely important and should be considered much more when choosing methods for large scale molecular dynamics or similar computations containing long ranged interactions.

APPENDIX: CHARGE ASSIGNMENT SCHEMES

We shall now provide the charge assignment schemes. We assume that the grid size is H and define x to be the relative position of a unit charge to a point $W_0 \equiv P$ which is defined as, the nearest grid point of the charge when p is even; the midpoint of the two nearest gridpoints of the charge when p is odd, $\{x \in [-(H/2),(H/2)]\}$. For a pth order charge assignment scheme we get the contributions from a unit charge to gridpoints W_γ residing at γ with respect to P given below.

$p=2$:

$$W_{-H}=\frac{x^2-Hx}{2H^2},$$

$$W_0=\frac{-2x^2+2H^2}{2H^2},$$

$$W_H=\frac{x^2+Hx}{2H^2}.$$

$p=3$:

$$W_{-(3H/2)}=\frac{-8x^3+12Hx^2+2H^2x-3H^3}{48H^3},$$

$$W_{-(H/2)}=\frac{24x^3-12Hx^2-54H^2x+27H^3}{48H^3},$$

$$W_{(H/2)}=\frac{-24x^3-12Hx^2+54H^2x+27H^3}{48H^3},$$

$$W_{(3H/2)}=\frac{8x^3+12H^2-2H^2x-3H^3}{48H^3}.$$

$p=4$:

$$W_{-2H}=\frac{x^4-2Hx^3-H^2x^2+2H^3x}{24H^4},$$

$$W_{-H}=\frac{-4x^4+4Hx^3+16H^2x^2-16H^3x}{24H^4},$$

$$W_0=\frac{6x^4-30H^2x^2+24H^4}{24H^4},$$

$$W_H=\frac{-4x^4-4Hx^3+16H^2x^2+16H^3x}{24H^4},$$

$$W_{2H}=\frac{x^4+2Hx^3-H^2x^2-2H^3x}{24H^4}.$$

$p=5$:

$$W_{-(5H/2)}=\frac{-32x^5+80Hx^4+80H^2x^3-200H^3x^2-18H^4x+45H^5}{3840H^5},$$

$$W_{-(3H/2)}=\frac{160x^5-240Hx^4-1040H^2x^3+1560H^3x^2+250H^4x-375H^5}{3840H^5},$$

$$W_{-(H/2)}=\frac{-320x^5+160Hx^4+2720H^2x^3-1360H^3x^2-4500H^4x+2250H^5}{3840H^5},$$

$$W_{(H/2)}=\frac{320x^5+160Hx^4-2720H^2x^3-1360H^3x^2+4500H^4x+2250H^5}{3840H^5},$$

$$W_{(3H/2)}=\frac{-160x^5-240Hx^4+1040H^2x^3+1560H^3x^2-250H^4x-375H^5}{3840H^5},$$

$$W_{(5H/2)}=\frac{32x^5+80Hx^4-80H^2x^3-200H^3x^2+18H^4x+45H^5}{3840H^5}.$$

$p=6$:

$$W_{-3H}=\frac{x^6-3Hx^5-5H^2x^4+15H^3x^3+4H^4x^2-12H^5x}{720H^6},$$

$$W_{-2H}=\frac{-6x^6+12Hx^5+60H^2x^4-120H^3x^3-54H^4x^2+108H^5x}{720H^6},$$

$$W_{-H}=\frac{15x^6-15Hx^5-195H^2x^4+195H^3x^3+540H^4x^2-540H^5x}{720H^6},$$

$$W_0=\frac{-20x^6+280H^2x^4-980H^4x^2+720H^6}{720H^6},$$

$$W_H=\frac{15x^6+15Hx^5-195H^2x^4-195H^3x^3+540H^4x^2+540H^5x}{720H^6},$$

$$W_{2H}=\frac{-6x^6-12Hx^5+60H^2x^4+120H^3x^3-54H^4x^2-108H^5x}{720H^6},$$

$$W_{3H}=\frac{x^6+3Hx^5-5H^2x^4-15H^3x^3+4H^4x^2+12H^5x}{720H^6}.$$

[1] J. Kolafa and J. W. Perram, Mol. Simul. **9**, 351 (1992).
[2] M. P. Allen and D. J. Tildesley, *Computer Simulations of Liquids* (Clarendon, Oxford, 1987).
[3] L. Greengard and V. Rokhlin, J. Comput. Phys. **73**, 325 (1987).
[4] P. P. Ewald, Ann. Phys. **64**, 253 (1921).
[5] S. W. de Leeuw, J. W. Perram, and E. R. Smith, Proc. R. Soc. A **373**, 27 (1980).
[6] J. W. Perram, H. G. Petersen, and S. W. de Leeuw, Mol. Phys. **65**, 875 (1988).
[7] D. Fincham, Info. Q. Comput. Simul. Condensed Phases **38**, 17 (1993).
[8] R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles* (McGraw–Hill, New York, 1981).
[9] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, J. Comp. Chem. **4**, 187 (1983).
[10] H. G. Petersen, D. Soelvason, J. W. Perram, and E. R. Smith, Proc. R. Soc. **448**, 389 (1995).
[11] H. G. Petersen, D. Soelvason, J. W. Perram, and E. R. Smith, Proc. R. Soc. **448**, 401 (1995).
[12] H. G. Petersen, D. Soelvason, J. W. Perram, and E. R. Smith, J. Chem. Phys. **101**, 8870 (1994).
[13] D. Soelvason, J. Kolafa, H. G. Petersen, and J. W. Perram, Comput. Phys. Commun. **87**, 307 (1995).
[14] K. Esselink (submitted).
[15] K. E. Schmidt and M. A. Lee, J. Stat. Phys. **63**, 1223 (1991).
[16] H-Q. Ding, N. Karasawa, and W. A. Goddard III, J. Chem. Phys. **97**, 4309 (1992).
[17] H-Q. Ding, N. Karasawa, and W. A. Goddard III, Chem. Phys. Lett. **196**, 6 (1992).
[18] T. Darden, D. York, and L. Pedersen, J. Chem. Phys. **98**, 10 089 (1993).
[19] H. G. Petersen and J. W. Perram, Mol. Phys. **67**, 849 (1989).
[20] G. Fox *et al.*, *Solving Problems on Concurrent Processors* (Prentice–Hall, Englewood Cliffs, 1988), Vol. I.
[21] W. H. Press *et al.*, *Numerical Recipes* (Cambridge University, Cambridge, 1986).
[22] E. R. Smith (in preparation).