JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Optimizing the Accuracy and Efficiency of Fast Hierarchical Multipole Expansions for MD Simulations

Konstantin Lorenzen, Magnus Schwörer, Philipp Tröster, Simon Mates, and Paul Tavan*

Lehrstuhl für Biomolekulare Optik, Ludwig-Maximilians-Universität, Oettingenstrasse 67, 80538 München, Germany

Supporting Information

ABSTRACT: Based on p'th order Cartesian Taylor expansions of Coulomb interactions and on hierarchical decompositions of macromolecular simulation systems into hierarchies of nested, structure-adapted, and adaptively formed clusters of increasing size, fast multipole methods are constructed for rapid and accurate calculations of electrostatic interactions. These so-called SAMM$_p$ algorithms are formulated through totally symmetric and traceless tensors describing the multipole moments and the coefficients of local Taylor expansions. Simple recursions for the efficient evaluation and shifting of multipole moments are given. The required tensors are explicitly given up to order $p = 4$. The SAMM$_p$ algorithms are shown to guarantee the reaction principle. For systems with periodic boundaries, a reaction field (RF) correction is applied, which introduces at distances beyond the "minimum image convention" boundary a dielectric continuum surrounding each cluster at the top level of coarse graining. The correctness of the present SAMM$_p$ implementation is demonstrated by analyzing the scaling of the residuals and by checking the numerical accuracy of the reaction principle for a pair of distant molecular ions in vacuum. Molecular dynamics simulations of pure water and aqueous solutions containing artificial ions, which are enclosed by periodic boundaries, demonstrate the stability and low-noise behavior of SAMM$_p$/RF.

1. INTRODUCTION

Molecular dynamics (MD) all-atom simulations of large biomolecular systems[1,2] pose a challenging computational problem mainly due to the long-range nature of electrostatic interactions. If one applies one of the common molecular mechanics (MM) force fields[3−5] for the description of such systems, the electrostatic signatures of the molecular components are specified by partial charges q_i localized at essentially all N atoms i in the simulated system, where N is typically in the range between 10^4−10^6. Because the electrostatic interactions cannot be truncated at the corresponding system sizes L of about 5−20 nm without introducing sizable artifacts,[6−8] the effort for the exact computation of the electrostatic forces scales like N^2 and, thus, rapidly becomes intractable with increasing N. Therefore, various schemes for the approximate and more efficient computation of the electrostatic forces were designed.

The most popular approximation schemes are the so-called lattice sum (LS) methods.[9−11] These methods take advantage of periodic boundary conditions (PBC), which avoid surface artifacts and, thus, enable the control of the density or of the pressure within the simulated system. LS methods enable for PBC systems the approximate computation of the correspondingly periodic electrostatic potential at a computational effort scaling with N log N. On the other hand, the assumption of a periodic electrostatic potential can introduce periodicity artifacts into the description of nonperiodic systems, such as liquids and proteins in solution.[12]

As an alternative, a combination of a fast structure-adapted multipole method (SAMM)[13−15] with a moving boundary reaction field (RF) approach[8] has been suggested, which applies the minimum image convention (MIC)[16] to the explicitly computed electrostatic forces and approximates the electrostatic forces for distances larger than the MIC distance $R_{MIC} = L/2$ through the Kirkwood RF.[17] Thus, this approach adds to toroidal boundary conditions (TBCs)[16] a suitable RF correction and, therefore, avoids the use of an artificially periodic potential as well as corresponding artifacts. Because the SAMM scheme is applied to the electrostatics computation within the MIC sphere surrounding every charge and because its hierarchically nested ternary tree structure is efficiently exploited in a top-down fashion for the generation of the interaction lists at all hierarchy levels, this TBC/RF approach scales linearly with N.[8,15] A corresponding parallelized MD code called EGO has been successfully applied in various large-scale biomolecular simulations (see refs 18 and 19 for recent examples).

The linear scaling achieved by SAMM is typical for such fast multipole methods (FMM),[20] which employ hierarchical decompositions of the total system into a tree of nested subsystems[21,22] of decreasing sizes. Whereas most FMM methods employ regular and nested real-space grids for the construction of this tree (see, e.g., refs 20 and 23−31), SAMM applies a partially adaptive bottom-up clustering of atoms into a nested hierarchy of molecular groups and clusters of such groups, which are formed by neural clustering algorithms.[32,33] A key difference between the real-space grid approaches and the SAMM is the structure of the resulting trees onto which the respective nested hierarchies of charge groups are mapped. The grid approaches employ octal trees, whereas SAMM uses ternary trees, within which the cluster sizes increase much more

Special Issue: Wilfred F. van Gunsteren Festschrift

Received: January 31, 2012
Published: March 21, 2012

slowly from a lower to the next higher level.[8,15] This leads to more efficient but more complex algorithms.

In the past, a variety of other FMM algorithms were suggested for the efficient simulation of periodic biomolecular systems, which usually added LS sum methods to the FMM description of the local electrostatics with the aim of including the periodic images into the computation of the periodic electrostatic potential.[23−25,27,28] Markedly different is the recent combination[34] of a FMM tree code[26] with the isotropic periodic sum approach,[35] which includes the effects of the periodic images in a mean-field sense and, just like SAMM/RF, avoids the generation of an artificially periodic electrostatic potential.

FMM algorithms employ either spherical[20,23−25] or Cartesian[14,26−31,36] coordinates for the computation of the required multipole and Taylor expansions. While the expansions based on spherical harmonics generally were extended to relatively high orders p, the Cartesian methods usually truncate the expansions of the electrostatic potential at lower orders, e.g., $p = 2$ or 3 (quadrupole or octopole, respectively), and choose the depth of the grid hierarchy accordingly for a reasonable compromise between the conflicting aims of accuracy and efficiency. In the case of SAMM, e.g., the multipole and local Taylor expansions were both truncated at second order.[8]

Extending an earlier analysis of Cartesian FMM approaches by Warren and Salmon,[29] Dehnen[31] made in 2002 the important observation that the resulting approximate FMM forces obey Newton's reaction principle, if the multipole and local Taylor expansions of the electrostatic potential are truncated at those levels m and n, respectively, which obey the sum rule $p = m + n$, where p is the highest multipole order considered. Within the thus specified FMM design, the order p defines for the energy of two interacting charges q_i and q_j the approximate magnitude of the expected FMM error through $O[(q_iq_j/r)(d/r)^{p+1}]$. Here, r is the distance between the respective clusters A and B, to which the charges q_i and q_j belong, and d is the typical radius of these clusters.

As an important consequence of this FMM design, the local Taylor expansion of the potential Φ^m, which is generated by a multipole moment of order m, has to be carried out only up to the order $n = p - m$ to reach a desired accuracy (as defined by the highest multipole order p included in the treatment). Accounting for higher order terms $n > p - m$ in this Taylor expansion, like in the previous implementation of SAMM/RF,[8] considerably increases the computational effort without adding a substantial gain of accuracy.

Therefore the quoted result of Dehnen[31] represents a guideline for the construction of Cartesian FMM methods, which not only yield dynamically reasonable interatomic forces but also additionally represent optimal compromises between accuracy and efficiency. In both respects the existing SAMM/RF version,[8] which we will call $SAMM_{2003}$/RF from now on, was clearly suboptimal. Note that $SAMM_{2003}$/RF extended the even simpler predecessor[15] $SAMM_{1997}$, which had been solely applicable to systems under fixed boundary conditions.

Because the molecular polarizability has to be included into MM-MD simulations[2] and into quantum-classical hybrid simulations[37] and because this inclusion requires an enhanced accuracy of the electrostatics calculations, a revision of our parallelized MD code became necessary. In view of the quoted result of Dehnen,[31] we decided to revise $SAMM_{2003}$/RF accordingly. For a most efficient and flexible description of polarization effects we decided to employ atomic polarization dipoles, which are easily integrated into Cartesian FMM approaches.

It is the purpose of this contribution, to sketch the resulting SAMM algorithms which will be called $SAMM_p$, $p = 2−4$, where p is the highest multipole order employed for the Cartesian FMM expansion of the electrostatic potential. We start with a sketch of the theory in a form that closely matches the actual implementation and, therefore, can serve as a guideline for future users of our revised MD code, which will be called IPHIGENIE from now on. Using two simple sample systems, i.e., a pair of the molecular ions H_3O^+ and $H_2PO_4^-$ at varying distances and a periodic box filled with 1500 MM water molecules at ambient temperature and pressure, we study the properties of the new $SAMM_p$ and $SAMM_p$/RF algorithms.

2. THEORY

The revised $SAMM_p$ algorithms announced above, which have been implemented in our parallelized MM-MD simulation program IPHIGENIE, are based on Cartesian multipole and Taylor expansions approximating the electrostatic potentials and fields caused by distant charge distributions.

2.1. FMM from Taylor Expansions. Figure 1 identifies the geometry for the derivation of the $SAMM_p$ approximations to

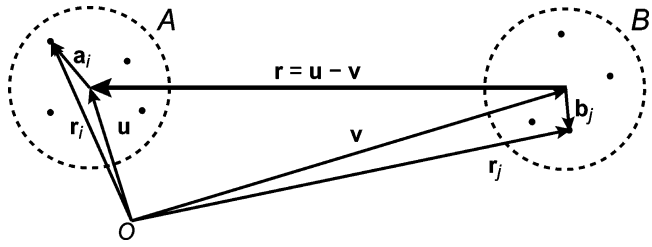

Figure 1. SAMM geometry for two interacting clusters A and B (dashed spheres) of charges $q_i \in A$ and $q_j \in B$ (dots). The interaction of q_i and q_j at $\mathbf{r}_i$ and $\mathbf{r}_j$, respectively, depends on the connecting vector $\mathbf{r}_i - \mathbf{r}_j$ and is evaluated by a Taylor expansion around the vector $\mathbf{r}$ linking the two cluster centers. The positions of these centers are denoted by $\mathbf{u}$ and $\mathbf{v}$, respectively, those of the charges within the respective clusters by $\mathbf{a}_i$ and $\mathbf{b}_j$.

the electrostatic potential and field. The Coulomb potential Φ^B at the position r_i of the charge $q_i \in A$, which is generated by all charges $q_j \in B$, is given (in Gaussian CGS units) by

$$\Phi^B(\mathbf{r}_i) = \sum_{j \in B} \frac{q_j}{|\mathbf{r}_i - \mathbf{r}_j|} \tag{1}$$

For the geometry displayed in Figure 1 this expression can be rewritten as

$$\Phi^B(\mathbf{r}_i) = \sum_{j \in B} \frac{q_j}{|\mathbf{r} + (\mathbf{a}_i - \mathbf{b}_j)|} \tag{2}$$

where we have used the cluster-local coordinates $\mathbf{a}_i = \mathbf{r}_i - \mathbf{u}$ and $\mathbf{b}_j = \mathbf{r}_j - \mathbf{v}$ together with the cluster−cluster connection vector $\mathbf{r} = \mathbf{u} - \mathbf{v}$.

Employing the tensor notation of Warren and Salmon,[29] which is explained in Section 1 of the Supporting Information by providing relevant examples for inner and outer tensor

dx.doi.org/10.1021/ct300080n | J. Chem. Theory Comput. 2012, 8, 3628−3636

products, the p'th order Taylor expansion of eq 2 around $\mathbf{r}$ reads

$$\Phi^{B}(\mathbf{r}_i) = \Phi^{B,p}(\mathbf{r}_i) + R^{B,p}(\mathbf{r}_i) \tag{3}$$

with the residual $R^{B,p}(\mathbf{r}_i)$ and the expansion

$$\Phi^{B,p}(\mathbf{r}_i) = \sum_{j\in B} q_j \sum_{n=0}^{p} \frac{1}{n!}\left(\partial_{(n)}\frac{1}{r}\right) \odot (\mathbf{a}_i - \mathbf{b}_j)^{(n)} \tag{4}$$

Here $\partial_{(n)}(1/r)$ is a tensor of rank n composed of the n'th order partial derivatives of $1/r$. Section 2 of the Supporting Information explicitly lists its components for $n \leq 4$. The symbol $\odot$ denotes the inner contraction product of two tensors. The tensor $(\mathbf{a}_i - \mathbf{b}_j)^{(n)}$ of rank n is the n-fold outer product of the vector $\mathbf{a}_i - \mathbf{b}_j$ with itself. In the notation of these outer products $\partial_{(n)}(1/r)$ may be equivalently written as $\nabla_{\mathbf{r}}^{(n)}(1/|\mathbf{r}|)$.

Denoting the outer tensor product by $\otimes$, applying the binomial law

$$(\mathbf{a}_i - \mathbf{b}_j)^{(n)} = \sum_{m=0}^{n} (-1)^m \binom{n}{m} \mathbf{a}_i^{(n-m)} \otimes \mathbf{b}_j^{(m)} \tag{5}$$

to the n-fold outer product in eq 4, sorting the resulting linear combination according to increasing powers of $\mathbf{a}_i$ and exchanging the order of summations, the p'th order Taylor expansion eq 4 becomes

$$\Phi^{B,p}(\mathbf{r}_i) = \sum_{n=0}^{p} \frac{1}{n!}\mathbf{a}_i^{(n)} \odot \partial_{(n)} \sum_{m=0}^{p-n} \Phi^{m,B}(\mathbf{u}) \tag{6}$$

where

$$\Phi^{m,B}(\mathbf{u}) = \frac{(-1)^m}{m!}\left(\partial_{(m)}\frac{1}{r}\right) \odot \sum_{j\in B} q_j \mathbf{b}_j^{(m)} \tag{7}$$

are the potentials $\Phi^{m,B}(\mathbf{u})$ of m'th order multipole moments localized at the reference point $\mathbf{v}$ of cluster B (cf., Figure 1). Equation 6 is one of the basic FMM equations, because it describes a p'th order Taylor expansion of multipole potentials generated by cluster B around the center $\mathbf{u}$ of cluster A.

However, in eq 7 the representation of the m'th order multipole moments through the outer products $\mathbf{b}_j^{(m)}$ is suboptimal, because the associated totally symmetric tensors have $(m + 2)(m + 1)/2$ independent components.[38] A more compact representation is achieved, if one employs the reduced totally symmetric multipole tensors

$$\mathbf{M}^{m,\mathbf{v}} = \sum_{j\in B} q_j (-1)^m b_j^{2m+1}\left(\partial_{(m)}\frac{1}{b_j}\right) \tag{8}$$

which have only $2m + 1$ independent components, because they are traceless with respect to every pair of tensor components.[39,38] Here, the symbols b_j denote the absolute values of the local coordinates $\mathbf{b}_j$ of the charges q_j making up cluster B localized around $\mathbf{v}$. Section 3 of the Supporting Information lists explicit expressions for the components of these multipole moments up to order $m = 4$. With the reduced moments, eq 8, the multipole potentials, eq 7, can be alternatively expressed as

$$\Phi^{m,B}(\mathbf{u}) = \frac{(-2)^m}{(2m)!}\left(\partial_{(m)}\frac{1}{r}\right) \odot \mathbf{M}^{m,\mathbf{v}} \tag{9}$$

The local Taylor expansion eq 6 of the multipole potentials within cluster A can be compactly rewritten by introducing the expansion coefficient tensors

$$\mathbf{T}^{B,n,p}(\mathbf{u}) \equiv \partial_{(n)} \sum_{m=0}^{p-n} \Phi^{m,B}(\mathbf{u}), \qquad n = 0, ..., p \tag{10}$$

which account for the contributions of all multipole moments $\mathbf{M}^{m,\mathbf{v}}$ localized at the center $\mathbf{v}$ of cluster B to a given order n of the local Taylor expansion in cluster A in such a way that the maximal order p of the original Taylor expansion eq 4 is preserved. Correspondingly, all multipole moments $\mathbf{M}^{m,\mathbf{v}}$ up to the maximal order $m = p$ contribute to the zeroth order term of the FMM potential eq 6, whereas only the potential of cluster B's total charge $\mathbf{M}^{0,\mathbf{v}}$ contributes to the p'th order term in eq 6 through its p'th partial derivatives.

We note that the expansion coefficient tensors $\mathbf{T}^{B,n,p}(\mathbf{u})$ have only $2n + 1$ independent components [just like the reduced multipole moment tensors eq 8]. With eq 10 the local Taylor expansion eq 6 of the potential reads

$$\Phi^{B,p}(\mathbf{r}_i) = \sum_{n=0}^{p} \frac{1}{n!}\mathbf{a}_i^{(n)} \odot \mathbf{T}^{B,n,p}(\mathbf{u}) \tag{11}$$

For dynamics simulations one needs the electrostatic forces $\mathbf{f}_i = q_i\mathbf{E}(r_i)$ acting on the charges q_i. FMM offers two ways to calculate the required fields $\mathbf{E}(\mathbf{r}_i)$. One can either use the negative gradient

$$\mathbf{E}^{B,p}(\mathbf{r}_i) = -\nabla_i \Phi^{B,p}(\mathbf{r}_i) \tag{12}$$

of the FMM potential eq 11, which gives the p'th order field $\mathbf{E}^{B,p}(\mathbf{r}_i)$ generated by the charge cluster B. On the other hand we can start with the field

$$\mathbf{E}^{B}(\mathbf{r}_i) = \sum_{j\in B} \frac{q_j(\mathbf{r} + (\mathbf{a}_i - \mathbf{b}_j))}{|\mathbf{r} + (\mathbf{a}_i - \mathbf{b}_j)|^3} \tag{13}$$

associated to the original electrostatic potential eq 1 and apply a FMM Taylor expansion analogous to that in eq 4 but limited to order $p - 1$, i.e.

$$\mathbf{E}^{B,p}(\mathbf{r}_i) = \sum_{j\in B} q_j \sum_{n=0}^{p-1} \frac{1}{n!}\left(\partial_{(n)}\frac{\mathbf{r}}{r^3}\right) \odot (\mathbf{a}_i - \mathbf{b}_j)^{(n)} \tag{14}$$

to find with eqs 8 and 10 the same final expression

$$\mathbf{E}^{B,p}(\mathbf{r}_i) = -\sum_{n=0}^{p-1} \frac{1}{n!}\mathbf{a}_i^{(n)} \odot \mathbf{T}^{B,n+1,p}(\mathbf{u}) \tag{15}$$

Like eq 11 for the potential, this expansion also accounts for all terms up to order $(1/r)^{p+1}$.

2.2. FMM Forces Fulfill the Reaction Principle. As mentioned in the Introduction and pointed out by Dehnen,[30,31] electric fields calculated by eqs 10 and 15 ensure Newton's

third law. This is seen by considering the field, eq 14, at the location $\mathbf{r}_i$ in A for a single generating charge q_j in B, i.e.

$$\mathbf{E}^{j,p}(\mathbf{r}_i) = -q_j \sum_{n=0}^{p-1} \frac{1}{n!}\left(\partial_{(n+1)}\frac{1}{r}\right) \odot (\mathbf{a}_i - \mathbf{b}_j)^{(n)} \tag{16}$$

and the field $\mathbf{E}^{i,p}(\mathbf{r}_j)$ generated by q_i in A at the location $\mathbf{r}_j$ in B. With the geometry in Figure 1 and the vector $\mathbf{r}' \equiv -\mathbf{r}$, this FMM field is

$$\mathbf{E}^{i,p}(\mathbf{r}_j) = -q_i \sum_{n=0}^{p-1} \frac{1}{n!}\left(\partial_{(n+1)}\frac{1}{r'}\right) \odot (\mathbf{b}_j - \mathbf{a}_i)^{(n)} \tag{17}$$

With $\partial_{(n+1)}(1/r') = (-1)^{n+1}\partial_{(n+1)}(1/r)$ and $(\mathbf{b}_j - \mathbf{a}_i)^{(n)} = (-1)^n(\mathbf{a}_i - \mathbf{b}_j)^{(n)}$, one immediately arrives at the reaction principle $\mathbf{f}_{ij} = q_i\mathbf{E}^{j,p}(\mathbf{r}_i) = -q_j\mathbf{E}^{i,p}(\mathbf{r}_j) = -\mathbf{f}_{ji}$ for the pair forces generated by the FMM field eq 15.

2.3. The SAMM Algorithm. A characteristic feature of FMM methods, like SAMM, is the decomposition of the whole system into a hierarchically nested tree of charge clusters. In SAMM, 3−5 partially charged atoms are combined into clusters at the lowest level $l = 0$ of the hierarchy following local chemical motifs.[13] Applying neural clustering algorithms[32,33] (in predefined time intervals, which are large compared to the integration step), on average, 4 of these clusters are combined into compact clusters at the next higher level $l = 1$, and this procedure is repeated until a certain top-level t is reached.[14,15]

Next, interaction lists L_c^l are calculated from distance criteria for each cluster c in every level l of the tree in a top-down fashion.[14] If the system is enclosed by periodic boundaries, only those clusters c' are included into L_c^t, whose surface-to-surface distance $d_{c'c}$ complies with the MIC ($d_{c'c} < R_{\mathrm{MIC}}$) and is larger than a predefined threshold d_t. Here, the top-level t is chosen in such a way that $L_c^t \neq 0$. Clusters c and c' with $d_{c'c} < d_t$ are decomposed into their children. Pairs of these children are included into the interaction lists of level $t-1$, if their distances exceed the threshold $d_{t-1} < d_t$ belonging to this level. This procedure is repeated until the lowest cluster level $l = 0$ is reached. Clusters violating the distance threshold d_0 are finally decomposed into individual atoms, whose interactions are treated with the usual Coulomb expressions. The interaction lists are updated in predefined time intervals. Note that the accuracy of SAMM can be steered by the choice of the distance class boundaries d_l.

2.4. Calculation of Multipole Moments. Using the tree structure in a bottom-up fashion the multipole moments $\mathbf{M}_C^{m,\mathbf{0}}$ of parent clusters C are calculated from those of their children $c(C)$ by a simple iterative procedure:

At level $l = 0$, the multipole moments $\mathbf{M}_c^{m,\mathbf{0}}$ of all clusters c are calculated directly by eq 8 from the partial charges of the embedded atoms taking the origin $\mathbf{0}$ as the reference point. Collecting all clusters $c \in C$, which belong to a parent cluster C at the next higher level, the multipole moments

$$\mathbf{M}_C^{m,\mathbf{0}} = \sum_{c\in C} \mathbf{M}_c^{m,\mathbf{0}} \tag{18}$$

of parent C with respect to $\mathbf{0}$ are simply the sums of the moments $\mathbf{M}_c^{m,\mathbf{0}}$ of its children c. This procedure is repeated until the top-level t is reached.

As is apparent from Figure 1 and eqs 9 and 11, a computation of FMM interactions between two clusters A and B belonging to a given level l requires multipole moments localized at the respective cluster centers $\mathbf{u}$ and $\mathbf{v}$. Thus, at all hierarchy levels the multipole moments $\mathbf{M}^{m,\mathbf{0}}$ of the various clusters have to be shifted from the origin $\mathbf{0}$ to the associated cluster centers $\mathbf{c}$. Using auxiliary tensors $\mathbf{H}_{m,\mathbf{c}}^i$ of rank $i \in \{0, ..., m\}$, which are recursively calculated for $k \in \{0, ..., m-1\}$ through

$$\mathbf{H}_{m,\mathbf{c}}^{k+1} = \mathbf{M}^{k+1,\mathbf{0}} - \frac{(k+1)}{m-k}\hat{S}_{k+1}[(2k+1)(\mathbf{c} \otimes \mathbf{H}_{m,c}^k) - k(\mathbf{c} \odot \mathbf{H}_{m,\mathbf{c}}^k) \otimes \mathbf{I}] \tag{19}$$

from the multipole moments $\mathbf{M}^{k+1,\mathbf{0}}$ at the origin, where the starting point $k = 0$ of the recursion

$$\mathbf{H}_{m,\mathbf{c}}^0 = \mathbf{M}^{0,\mathbf{0}} \tag{20}$$

is the total charge $\mathbf{M}^{0,\mathbf{c}} = \mathbf{M}^{0,\mathbf{0}}$ of the cluster, the shifted multipole tensors

$$\mathbf{M}^{m,\mathbf{c}} = \mathbf{H}_{m,\mathbf{c}}^m \tag{21}$$

are given by the auxiliary tensors of rank m. In eq 19 the operator $\hat{S}_n$ is the symmetrizer

$$\hat{S}_n(A_{1,2,...,n}^n) = \frac{1}{n!}\sum_{p\in\mathcal{I}^n} A_{p(1,2,...,n)}^n \tag{22}$$

for the components $A_{1,2,...,n}^n$ of a tensor of rank n, where $\mathcal{I}^n$ denotes the symmetric group of permutations p for n objects. Note here that the recursion eq 19 enables a sequential and highly efficient evaluation of the multipole moments $\mathbf{M}^{m,\mathbf{c}}$ for $m = 1, ..., p$, in which only the $2(m+1)$ nonredundant tensor components have to be considered at each rank m.

2.5. FMM Interactions in SAMM. Starting at the top-level t, the Taylor expansion coefficients eq 10, which are eventually required to compute the electrostatic fields and potentials at the positions $\mathbf{r}_i$ of the atoms, are computed by descending the levels of the tree. At each level l of the hierarchy, the multipole moments $\mathbf{M}^{m,\mathbf{v}}$ of all clusters B centered at the positions $\mathbf{v}$, which belong to the interaction list L_A^l of a cluster A, contribute up to the rank $m = p - n$ to the tensor $\mathbf{T}^{B,n,p}(\mathbf{u})$ of coefficients in the Taylor expansion eq 11 around the center $\mathbf{u}$ of A. As is illustrated in Figure 2 by a dashed arrow, this action of B on A is inherited to the children $c \in A$ at level $l-1$, i.e., to local Taylor expansions centered at the reference points $\mathbf{c}$, by a shifting operation. The shifting can be carried out without loss of information. The children have then, of course, additional direct contributions to their local Taylor expansions from all

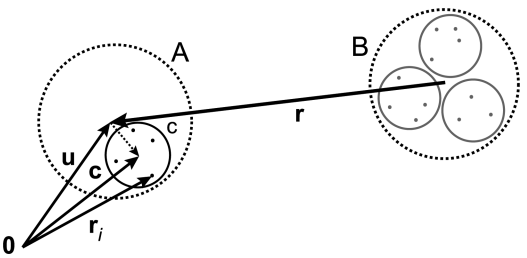

Figure 2. At the parent level l, the potential, which is generated by the charge distribution B, is given at the point r_i by the p'th order Taylor expansion eq 11 around the center $\mathbf{u}$ of cluster A. This potential can be equivalently expressed by a p'th order Taylor expansion around the center $\mathbf{c}$ of the child cluster c using the shifted expansion coefficient eq 24.

dx.doi.org/10.1021/ct300080n | J. Chem. Theory Comput. 2012, 8, 3628−3636

clusters in their respective interaction lists L_c^{l-1}, and all these contributions are shifted to the next lower level.

For a proof that the shifting operation sketched above preserves information, one considers the approximate potential $\Phi^{B,p}(\mathbf{r}_i)$ in eq 11, which is generated by a distant charge distribution B and is given as a p'th order Taylor expansion around a center $\mathbf{u}$ with the coefficients $\mathbf{T}^{B,n,p}(\mathbf{u})$ calculated by means of eq 10. Then one replaces in eq 11 the local atomic coordinates $\mathbf{a}_i$, which refer to the cluster center $\mathbf{u}$, by the vectors $\mathbf{d} + \tilde{\mathbf{a}}_i$, where the $\tilde{\mathbf{a}}_i$ is the local atomic coordinate with respect to the new reference point $\mathbf{c}$ and where $\mathbf{d} = \mathbf{c} - \mathbf{u}$ is the translation from $\mathbf{u}$ to $\mathbf{c}$. In Figure 2, which illustrates the geometry, $\mathbf{d}$ is drawn as a dashed arrow. Using, as in eq 5, the binomial law to evaluate the powers $(\mathbf{d} + \tilde{\mathbf{a}}_i)^{(n)}$ appearing in the resulting expression and sorting according to ascending powers of $\tilde{\mathbf{a}}_i$ yields

$$\Phi^{B,p}(\mathbf{r}_i) = \sum_{n=0}^{p} \frac{1}{n!}\tilde{\mathbf{a}}_i^{(n)} \odot \tilde{\mathbf{T}}^{B,n,p}(\mathbf{c}) \tag{23}$$

with the shifted Taylor expansion coefficients

$$\tilde{\mathbf{T}}^{B,n,p}(\mathbf{c}) = \sum_{l=0}^{p-n} \frac{1}{l!}\mathbf{d}^{(l)} \odot \mathbf{T}^{B,l+n,p}(\mathbf{u}) \tag{24}$$

In this way, electrostatic interactions between higher level clusters are inherited to the lowest level, where the resulting Taylor expansions are used to compute the contributions of distant charges to the electrostatic potential and field acting on the individual atoms.

For systems in periodic boundaries, also the top-level clusters inherit electrostatic interactions from a higher level, which is a dielectric continuum starting at the MIC distance R_{MIC} from the center of every top-level cluster and is modeled by the Kirkwood RF[17] as described in ref 8. In such cases the electrostatics treatment will be called SAMM$_p$/RF. We strongly advice readers interested in the concepts and details of this highly efficient combination of FMM with a RF correction, which include, e.g., provisions for smooth transitions of distant top-level clusters into and out of the "Kirkwood sphere" surrounding each of these clusters, to study the original paper.[8]

2.6. Aspects of the Current Implementation. As mentioned in the Introduction, the SAMM$_p$/RF algorithm sketched above has been implemented in a parallelized fashion in the C-program IPHIGENIE for the expansion orders $p = 2-4$. This program represents a thorough revision and extension of an earlier MD simulation code called EGO,[8,15] which employed the FMM approach SAMM$_{2003}$/RF. As opposed to the SAMM$_p$ algorithms, SAMM$_{2003}$ generally violated the reaction principle, because the orders of the multipole and Taylor expansions were not properly balanced. For electrostatic forces originating from charged clusters, the residual error scaled as $1/r^4$, just like that of SAMM$_2$. For neutral clusters, however, the residual of the SAMM$_{2003}$ force calculation scaled as $1/r^5$ (like SAMM$_3$) with a computational effort somewhat larger than that of SAMM$_3$. Note that SAMM$_{1997}$ employed a more balanced combination of multipole and Taylor expansions for the computation of electric fields; therefore, it fulfilled the reaction principle. Like in SAMM$_2$, its residual force error scaled with $1/r^4$ despite a larger computational effort.

Beyond a systematic, balanced, and more accurate combination of multipole and Taylor expansions achieved particularly through SAMM$_4$, the extensions of the code include the use of polarizable force fields, in which the individual atoms can carry inducible dipoles in addition to the static partial charges. Furthermore, they involve the interface[40] to the density functional theory program CPMD,[41] which meanwhile enables fully Hamiltonian MD simulations[42] through the use of Hellmann–Feynman forces (evaluated by FMM at larger distances). For the polarizable degrees of freedom, a separate FMM tree and self-consistent field iterations are available.

However some features are still inherited from its predecessor EGO. For instance, IPHIGENIE still employs the same scheme of SAMM distances d_l ($l = 0, 1, ..., t$), which define the various distance classes and associated cluster levels l, as EGO.[8] This choice is most certainly suboptimal because SAMM$_4$ enables, as we will show now, a much more accurate treatment of larger clusters at even much smaller distances.

3. METHODS

To check the accuracy of the SAMM$_p$ electrostatics calculation and the dynamic stability of SAMM$_p$/RF MD simulations of condensed phase systems subject to toroidal boundary conditions,[16] we chose three model systems. The first consists of the molecular ion pair $H_2PO_4^-\cdots H_3O^+$ in vacuum and the other two of polar and ionic liquids enclosed by periodic boundaries.

3.1. Accuracy Checks. For the SAMM$_p$ accuracy checks we chose, like in Figure 1, two charge clusters A and B separated by a center-to-center distance r. Because the charge distributions should carry multipole moments of all orders, we chose the two nontrivial molecular ions $H_2PO_4^-$ and H_3O^+ as representatives for A and B. Figure 3 illustrates one of the many relative arrangements of these ions.

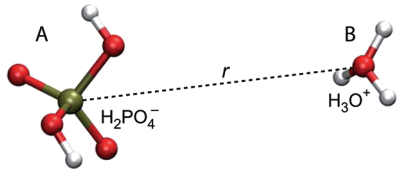

Figure 3. The molecular ions $H_2PO_4^-$ and H_3O^+ at a distance r as representatives for the charge clusters in Figure 1.

Rigid and purely electrostatic MM models were derived for each of these ions by density functional theory (DFT) calculations with the program Gaussian[43] using the B3LYP functional and the 6-31G** basis set. The resulting electrostatic potential (ESP)[44] partial charges and optimized geometries are listed in Section 5 of the Supporting Information. Like in Figure 1, we denote the atomic coordinates of $H_2PO_4^-$ by $\mathbf{r}_i$ and those of H_3O^+ by $\mathbf{r}_j$. Furthermore, the geometric centers $\mathbf{u}$ of $H_2PO_4^-$ and $\mathbf{v}$ of H_3O^+ were chosen as reference points such that the cluster–cluster distance is $r = |\mathbf{u} - \mathbf{v}|$.

The accuracy, by which the SAMM$_p$ algorithms describe the electrostatic interactions, depends on r. To characterize this dependence, r was varied in the range 5–15 Å in steps of 0.2 Å. At each step each of the rigid molecules was randomly rotated and an ensemble $\mathcal{A}$ of 10 000 relative arrangements was generated. For each arrangement $a \in \mathcal{A}$ the total electrostatic energy of the system U and the forces $\mathbf{f}_n$ on the various atoms $n \in A \cup B$ were calculated approximately by SAMM$_p$, $p = 2-4$ and exactly through the Coulomb expressions.

Because the error $O[(1/r)(d/r)^{p+1}]$ of a p'th order FMM expansion does not only depend on the cluster–cluster distance

dx.doi.org/10.1021/ct300080n | J. Chem. Theory Comput. 2012, 8, 3628–3636

r but also on the typical radius d of the interacting clusters, we designed a second test, focusing on the influence of d. Choosing r fixed at 10 Å, the local coordinates $\mathbf{a}_i = \mathbf{r}_i - \mathbf{u}$ and $\mathbf{b}_j = \mathbf{r}_j - \mathbf{v}$ were scaled by factors $g \in [1.0, 2.0]$, which were varied in steps $\Delta g = 0.05$.

For each value of the parameters r and g, which we jointly denote by x, the absolute SAMM$_p$ errors $\xi(x|p)$ of the energies U and forces $\mathbf{f}_n$ were measured by the root-mean-square deviations

$$\xi_U(x|p) = \sqrt{\langle (U - U^p)^2 \rangle_{\mathcal{A}}} \tag{25}$$

and

$$\xi_{\mathbf{f}}(x|p) = \sqrt{\left\langle \frac{1}{3|A \cup B|} \sum_{n \in A \cup B} (\mathbf{f}_n - \mathbf{f}_n^p)^2 \right\rangle_{\mathcal{A}}} \tag{26}$$

where the brackets $\langle ... \rangle_{\mathcal{A}}$ denote the arithmetic mean over the structural ensemble $\mathcal{A}$ at a given value of the respective parameter $x \in \{r,g\}$ and where $(U,\mathbf{f}_n)$ denote the exact quantities and $(U^p,\mathbf{f}_n^p)$ the respective SAMM$_p$ approximations. Relative errors $\rho(x|p)$ are then defined by

$$\rho_U(x|p) = \sqrt{\frac{\langle (U - U^p)^2 \rangle_{\mathcal{A}}}{\langle U^2 \rangle_{\mathcal{A}}}} \tag{27}$$

and

$$\rho_{\mathbf{f}}(x|p) = \sqrt{\frac{\langle \sum_{n \in A \cup B} (\mathbf{f}_n - \mathbf{f}_n^p)^2 \rangle_{\mathcal{A}}}{\langle \sum_{n \in A \cup B} \mathbf{f}_n^2 \rangle_{\mathcal{A}}}} \tag{28}$$

3.2. Dynamics Checks. In SAMM$_p$/RF, the algorithm for electrostatics calculations discontinuously changes with distance switching, e.g., at about $D_0 = 10$ Å from pairwise Coulomb interactions to the level $l = 0$ of the FMM treatment, at a larger distance of about $D_1 = 16$ Å to level $l = 1$, etc., until the computation eventually smoothly switches[8] to the RF description at about R_{MIC}. In a liquid, the diffusive dynamics induces constant distance boundary crossings of clusters. The discontinuities resulting for the associated forces cause algorithmic noise, which heats the simulated system. Note here that the symbols D_l are center-to-center distances of clusters or atoms, whereas the symbols d_l used in Section 2.3 for the construction of the interaction lists are surface-to-surface distances.

To estimate the decrease of algorithmic noise with increasing SAMM order p, we have collected ensembles $\mathcal{T}$ of 200 short ($\Delta t = 10$ ps) MD trajectories at constant volume V, number of molecules N, and total energy E (i.e., in the NVE setting) for two periodic liquid systems ($\mathcal{L}_1$, $\mathcal{L}_2$) and have measured heating rates per molecule

$$\dot{Q}_p = \frac{1}{N} \left\langle \frac{E_p(\Delta t) - E_p(0)}{\Delta t} \right\rangle_{\mathcal{T}} \tag{29}$$

as ensemble averages $\langle ... \rangle_{\mathcal{T}}$ from the total energies $E(t)$ at $t = 0$ and $t = \Delta t$. Here, statistically independent initial conditions had been drawn every 5 ps from 1 ns NVT trajectories of the systems in which a Berendsen thermostat[45] was used for control of the temperature T ($\tau = 100$ fs, $T_0 = 295$ K).

The test system $\mathcal{L}_1$ consisted of 1500 TIP3P[46] water models, which were kept rigid using MSHAKE[47] with a relative tolerance of 10^{-6}. To generate a solution $\mathcal{L}_2$ with a few ions, we changed the partial charges, which are assigned to the oxygen atoms, at two molecules by $+1e$ and at another two by $-1e$. After an initial embedding into cubic boxes with periodic boundaries, $\mathcal{L}_1$ and $\mathcal{L}_2$ were equilibrated by MD for 1 ns in the NpT ensemble using a Berendsen thermostat ($T_0 = 295$ K, $\tau = 100$ fs) and barostat ($p_0 = 1$ atm, $\tau = 5$ ps, $\kappa_T = 4.6 \times 10^{-5}$ atm^{-1}).[45] Equations of motion were integrated by the Verlet algorithm[48] with a time step of 1 fs. For both systems, R_{MIC} was about 18 Å, and the dielectric constant of the surrounding continuum was set to $\varepsilon = 78$. Because the boundary D_1 to the cluster level $l = 1$ was too close[8] to R_{MIC}, only predefined level 0 clusters (i.e., the molecules) were used for SAMM.

Beside the SAMM treatment of the electrostatics, the heating $\dot{Q}_p$ can have further sources like, e.g., the cutoff of the Lennard-Jones interactions at D_0, the disappearance of molecules into and their reappearance from the RF continuum or delayed updates of the interaction lists. To uniquely identify the SAMM contribution to $\dot{Q}_p$, we updated the interaction lists at every integration step and implemented an "exact" reference, which calculates all interactions within R_{MIC} by the Coulomb expressions and treats the RF boundary in exactly the same way as SAMM$_p$/RF. Hence, the reference method mimics SAMM$_p$/RF for $p \to \infty$ and, therefore, is denoted as SAMM$_\infty$/RF. The associated heating rate is called $\dot{Q}_\infty$. The differences $\dot{Q}_p - \dot{Q}_\infty$ then uniquely characterize the contribution of the SAMM$_p$ electrostatics approximation to the algorithmic noise.

4. RESULTS AND DISCUSSION

For a first verification of our implementation, the model calculations with the charge clusters $H_2PO_4^-$ and H_3O^+ depicted in Figure 3 serve us to check whether the residuals $R^{B,p}(\mathbf{r})$ [cf. eq 3] and $-\nabla R^{B,p}(\mathbf{r})$ of the SAMM$_p$ approximations, eqs 11 and eq 15, for the potential and field, respectively, show the expected scaling behaviors with the distance r and with the scaling factor g of the cluster sizes (cf. Section 3.1). From $|R^{B,p}(\mathbf{r})| = O[(1/r)(gd/r)^{p+1}]$ one expects that the scalings are $r^{-(p+2)}$ and g^{p+1}. Similarly, from $|-\nabla R^{B,p}(\mathbf{r})| = O[(1/r)^2(gd/r)^p]$, one gets $r^{-(p+2)}$ and g^p.

4.1. Actual Scalings. The log–log scale chosen in Figure 4 for the presentation of the absolute errors eq 26, which are connected with the SAMM$_p$ force calculations for our sample clusters (cf. Figure 3) at varying $r \in [5, 15$ Å$]$, allows us to extract the exponents of distance-scaling from the slopes of the shown linear regressions. Instead of the expected values

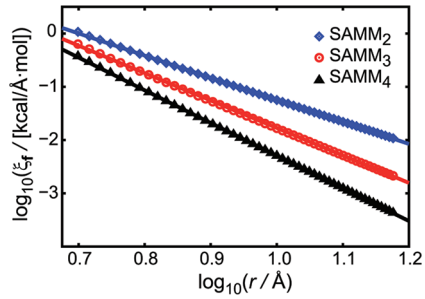

Figure 4. Log–log plot for the absolute errors eq 26 generated by SAMM$_p$ calculations of atomic forces $\mathbf{f}_n$ for the charge clusters $H_2PO_4^-$ and H_3O^+ shown in Figure 3 at cluster–cluster distances $r \in [5, 15$ Å$]$.

dx.doi.org/10.1021/ct300080n | J. Chem. Theory Comput. 2012, 8, 3628–3636

$-(p+2)$, $p = 2-4$, the regressions have slightly different slopes of −4.15, −5.15, and −6.13, respectively. For the absolute errors eq 26, which are associated with the SAMM$_p$ computation of the electrostatic energies U (data not shown), one obtains similar slopes of −4.07, −5.13, and −6.14. The small deviations from the expected values indicate that not only the neglected terms of order $(1/r)^{p+2}$ but also higher order and, thus, more rapidly vanishing terms contribute to the absolute SAMM$_p$ errors in the studied distance range.

Similarly, Figure 5 serves to check whether the observed absolute SAMM$_p$ errors eqs 25 and 26 show the expected

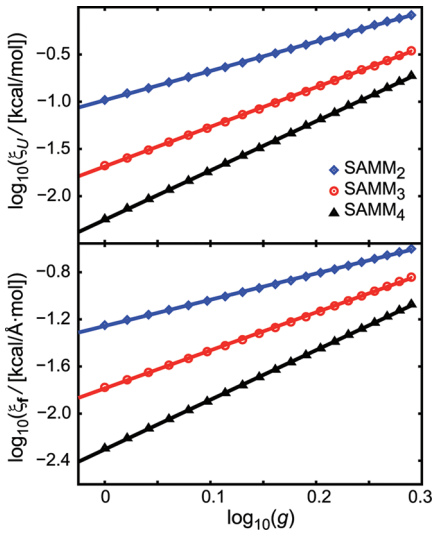

Figure 5. Log−log plots for the absolute errors (top) eq 25 and (bottom) eq 26 of potential energies U and atomic forces $\mathbf{f}_n$, respectively, at a cluster−cluster distance $r = 10.0$ Å when the cluster sizes are scaled by factors $g \in [1.0, 2.0]$; see the caption to Figure 4 for further information.

scaling with the cluster size. For the energy error one expects the slopes $p + 1$, whereas the regressions in Figure 5 (top) yield 3.10, 4.19, and 5.21 for $p = 2-4$, respectively. The force error depends less critically than that of the electrostatic energy on the cluster size. Here one expects that the slopes have the values p and finds from the linear regressions in Figure 5 (bottom) the values 2.23, 3.23, and 4.21, which are again very close to the expectations. As a result, the scaling behavior of our implementation complies with theory strongly suggesting that it is correct (note that we have successfully scrutinized the correctness of the implementation by a series of further tests including, for instance, the check whether Newton's reaction principle is obeyed at numerical accuracy).

However, the absolute SAMM$_p$ errors eqs 25 and 26 considered above are less indicative for the quality of a SAMM$_p$ electrostatics treatment in a condensed phase MD simulation than the relative errors, eq 28 because the relative errors allow us to estimate the relative sizes of the inevitable algorithmic discontinuities at distance class boundaries. In the original SAMM implementation,[15] the first distance class boundary is at $D_0 \approx 10$ Å, where the computation switches from exact Coulomb interactions to a FFM approximation for predefined molecular groups comprising 3−5 atoms. The size of the phosphate ion $H_2PO_4^-$ employed for our error estimates thus represents an upper limit for the typical size of molecular groups used in SAMM at cluster level $l = 0$. Correspondingly, the relative errors eq 28 shown in Figure 6 for our sample clusters at $r = 10$ Å (dashed lines) represent upper limits for the

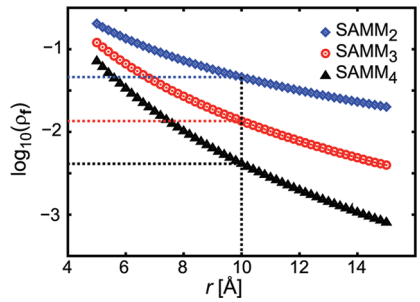

Figure 6. The relative errors eq 28 of the SAMM$_p$ force computations are represented on a logarithmic scale as functions of the cluster−cluster distance r; the dashed lines associated to the distance $r = 10$ Å give upper limits for relative discontinuities of force computation at the distance class boundary $D_0 = 10$ Å.

relative discontinuities of calculated forces, which are encountered during a MD simulation whenever a molecular group crosses D_0.

According to the data presented in Figure 6, the relative discontinuities of force computation at D_0 are smaller than 5, 1.4, and 0.4% for $p = 2-4$, respectively. Here, the relative energy errors eq 27 are about 1 order of magnitude smaller (data not shown). The noted discontinuities lead to the expectation that algorithmic noise will be a serious issue for SAMM$_2$, may be tolerable for SAMM$_3$, and should be small for SAMM$_4$.

In fact, SAMM$_{1997}$ and SAMM$_{2003}$ were plagued with substantial algorithmic noise, whenever the system contained charged clusters, because the residual force error scaled in this case as $1/r^4$ just like that of SAMM$_2$ (cf. Section 2.6). Furthermore, the reaction principle was generally violated in SAMM$_{2003}$. Hence, algorithmic noise should be reduced already when using SAMM$_3$ with its 1.4% relative errors near D_0 and even more at SAMM$_4$. Note that the relative errors of the SAMM$_4$ computation become comparable to those of SAMM$_3$ only at the very small center−center distance $r = 7.5$ Å, which approximately corresponds to a distance of about 1.8 Å between the van der Waals surfaces of the two molecular ions.

4.2. Algorithmic Noise of SAMM$_p$/RF. As explained in Section 3 we have set up two periodic cubic simulation systems filled either with 1500 simple TIP3P water models ($\mathcal{L}_1$) or with a variant, in which four of the water models were artificially converted into ions ($\mathcal{L}_2$). Using eq 29 we have measured method-specific heating rates $\dot{Q}_p$, $p = 2-4, \infty$ from many short *NVT* MD simulations on these systems. Here, $p = \infty$ labels the reference simulations with SAMM$_\infty$/RF, in which the SAMM$_p$ approximations were replaced by the exact Coulomb expressions. Table 1 lists for our sample liquid models $\mathcal{L}_1$ and $\mathcal{L}_2$ the differences between the various $\dot{Q}_p$ and the associated reference heating rates $\dot{Q}_\infty$ together with the statistical standard deviations measured for the SAMM$_p$/RF and SAMM$_\infty$/RF methods.

According to the values shown in Table 1, the heating rates observed in the two systems $\mathcal{L}_1$ and $\mathcal{L}_2$ are statistically indistinguishable at all levels p of approximation. The very crude approximations of SAMM$_2$/RF yield heating rates that are by about 57 kcal/(mol ns) larger than the heating rates $\dot{Q}_\infty \approx 1.4$ kcal/(mol ns) of the reference method SAMM$_\infty$/RF.

dx.doi.org/10.1021/ct300080n | J. Chem. Theory Comput. 2012, 8, 3628−3636

Table 1. SAMM$_p$/RF Heating Rates, $\dot{Q}_p$, $p = 2-4$, Relative to the Reference Rate, $\dot{Q}_\infty$, Evaluated for the Two Test Systems $\mathcal{L}_1$ (polar solvent) and $\mathcal{L}_2$ (ionic solution)[a]

	$\mathcal{L}_1$	$\mathcal{L}_2$
$\dot{Q}_2 - \dot{Q}_\infty$	56.87 ± 0.22	57.36 ± 0.22
$\dot{Q}_3 - \dot{Q}_\infty$	2.67 ± 0.05	2.59 ± 0.05
$\dot{Q}_4 - \dot{Q}_\infty$	0.05 ± 0.04	0.01 ± 0.04

[a]Rates are given per molecule in kcal/(mol ns); for explanations see the text.

This difference is strongly reduced to about 2.6 kcal/(mol ns) by the transition to SAMM$_3$/RF and essentially vanishes for SAMM$_4$/RF, whose heating rates $\dot{Q}_4$ are statistically indistinguishable from the reference rates $\dot{Q}_\infty$ despite the large statistical ensembles of 200 MD trajectories (Δt = 10 ps) spent for the computation of each $\dot{Q}$ in each of the two systems.

On the one hand, the heating rate data largely confirm the expectations derived from the relative force computation errors analyzed in connection with Figure 6. Like the discontinuities of the force computation at the distance class boundary D_0, also the heating rates become strongly suppressed with increasing order p of the SAMM expansions. On the other hand they surprisingly demonstrate that the relative discontinuities (<0.4%) remaining with SAMM$_4$ actually lead to a negligible heating.

The latter finding suggests that with SAMM$_4$/RF the distance class boundaries[8,15] D_l, l = 0, 1, ..., can be chosen considerably smaller without introducing large algorithmic artifacts. Such a change will then cause large efficiency gains, because the numbers of interaction partners will be greatly reduced at each level l of the cluster hierarchy. However, such a change requires that the van der Waals dispersion interaction is additionally included into the FMM scheme of computing nonbonded interactions for the following reason: Currently, the dispersion is truncated at the distance $D_0 \approx 10$ Å, at which the electrostatics computation switches from the Coulomb expressions to the SAMM$_p$ approximation. A choice of a smaller D_0 (e.g., 7 Å), which is compatible with a quite accurate SAMM$_4$ electrostatics computation, would cause with the current code a correspondingly short-range truncation of the dispersion interaction and, therefore, a sizable amount of algorithmic noise.

This noise can be avoided and quite small values D_l can be used as soon as the dispersion will be included into our FMM scheme, which is ongoing work.[26,36] Its completion will subsequently enable systematic studies of how the conflicting aims of accuracy and efficiency can be optimally reached with SAMM$_4$/RF. Already at the present stage of the implementation SAMM$_4$/RF is, for the sample systems $\mathcal{L}_1$ and $\mathcal{L}_2$, only by 24% less efficient than SAMM$_3$/RF, whereas the extremely crude SAMM$_2$/RF is only by 17% more efficient than SAMM$_3$/RF. As a result, increasing the order of p to p = 4 should be capable of shifting the compromise between accuracy and efficiency to a higher level.

Finally we would like to stress that the use of the Kirkwood RF correction[8] in connection with SAMM$_p$ is necessary for low-noise MD simulations. If one switches off the RF correction (e.g., by choosing the dielectric constant ε = 1 for the continuum surrounding each top-level cluster beyond R_{MIC}, thereby[8] implementing a smooth electrostatics cutoff at R_{MIC}), then the heating rates become very large for all p (data not shown). On the other hand, one observes an effective cooling for our sample systems with SAMM$_4$/RF and SAMM$_\infty$/RF characterized, e.g., by a rate $\dot{Q}_4$ = −3.11 kcal/(mol ns), if interaction lists are updated only every 64 time steps. As we have checked by switching off the electrostatics, this cooling is caused by the fact that particles are always diffusively spreading within the Lennard-Jones cutoff spheres and, due to the rare updates, even beyond ,while the new particles entering these spheres are identified only with a delay. As a result, deceleration by the dispersion attraction on average dominates the acceleration. Therefore, rare interaction list updates imply a "dispersion cooling".

4.3. Summary. We have presented a careful revision of the SAMM/RF algorithm.[8] This revision has been designed for rapid and accurate MD simulations of periodic condensed phase systems. IPHIGENIE, the new SAMM$_p$/RF implementation, employs systematic p'th order Cartesian FMM expansions, which, by construction, guarantee the reaction principle and scale linearly with the system size. Through the use of the Cartesian, totally symmetric and traceless multipole tensors eq 8 featuring at rank m only $2m + 1$ independent components and through the use of the coefficient tensors eq 10, which have analogous properties, the computational and storage costs are kept minimal. In particular, the shifting of the multipole moment tensors to new reference points can be effected through the efficient recursion, eq 19. As a corollary, this recursion reduces to an algorithm for the sequential computation of rank m multipole tensors for distributions composed of point charges, point dipoles, etc., as is immediately clear if one assumes that all these electrostatic point objects are initially located at the origin (cf. the Supporting Information).

SAMM$_4$ yields very accurate forces and electrostatic energies showing the theoretically expected scaling of the residuals. With SAMM$_4$/RF the algorithmic noise turned out to be negligible even for a relatively small sample simulation system with an inner radius R_i of only 18 Å. Because previous and much less accurate implementations of SAMM/RF showed a strong reduction of algorithmic noise with increasing R_i, we expect further improvements also for SAMM$_4$/RF at larger R_i.

The computational scenario outlined above is easily generalized toward the efficient treatment of polarizable force fields and of DFT/MM hybrid simulations (ongoing work). On the other hand, not all options for algorithmic optimizations have been exhausted so far such that there is, as always, ample room for further optimizations. Several of these issues have been identified in this work like, e.g., the choice of smaller distance classes with SAMM$_4$ and the inclusion of the dispersion into the FMM scheme, the use of polarizable force fields (PFF) in this setting, or the efficient electrostatics computation in fully Hamiltonian DFT/PFF simulations.

■ ASSOCIATED CONTENT

Supporting Information

Explanation of the tensorial notation, explicit expressions and recursion relations for the n'th derivatives of $1/r$, for the m'th order totally symmetric and traceless multipole moments and for the electrostatic potentials generated by these moments for $n, m \leq 4$. Furthermore parameters for the charge distributions $H_2PO_4^-$ and H_3O^+ are given. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: tavan@physik.uni-muenchen.de.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft (SFB749/C). Like all our work it has been guided by the desire to approach with our own contributions those standards, which have been set by Wilfred and his group in the field of macromolecular simulation through decades of outstanding scientific efforts and achievements.

REFERENCES

(1) van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glättli, A.; Hünenberger, P. H.; Kastenholz, M. A.; Oostenbrink, C.; Schenk, M.; Trzesniak, D.; van der Vegt, N. F. A.; Yu, H. B. *Angew. Chem., Int. Ed.* **2006**, *45*, 4064 −4092.

(2) Tavan, P.; Carstens, H.; Mathias, G. In *Protein Folding Handbook*; Buchner, J., Kiefhaber, T., Eds.; Wiley-VCH: Weinheim,Germany, 2005; Vol. *1*; pp 1170−1195.

(3) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.

(4) Ponder, J.; Case, D. *Adv. Protein Chem.* **2003**, *66*, 27−85.

(5) Oostenbrink, C.; Villa, A.; Mark, A.; Van Gunsteren, W. *J. Comput. Chem. B* **2004**, *25*, 1656−1676.

(6) Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. F. *J. Chem. Phys.* **1995**, *102*, 5451−5459.

(7) Hünenberger, P. H.; van Gunsteren, W. F. *J. Chem. Phys.* **1998**, *108*, 6117−6134.

(8) Mathias, G.; Egwolf, B.; Nonella, M.; Tavan, P. *J. Chem. Phys.* **2003**, *118*, 10847−10860.

(9) Darden, T. A.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

(10) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577−8593.

(11) Luty, B. A.; Tironi, I. G.; van Gunsteren, W. F. *J. Chem. Phys.* **1995**, *103*, 3014−3021.

(12) Hünenberger, P. H.; McCammon, J. A. *Biophys. Chem.* **1999**, *78*, 69−88.

(13) Niedermeier, C.; Tavan, P. *J. Chem. Phys.* **1994**, *101*, 734−748.

(14) Niedermeier, C.; Tavan, P. *Mol. Simul.* **1996**, *17*, 57−66.

(15) Eichinger, M.; Grubmüller, H.; Heller, H.; Tavan, P. *J. Comput. Chem.* **1997**, *18*, 1729−1749.

(16) Allen, M. P.; Tildesley, D. *Computer Simulations of Liquids*; Clarendon: Oxford, U.K., 1987.

(17) Kirkwood, J. G. *J. Chem. Phys.* **1934**, *2*, 351−361.

(18) Lingenheil, M.; Denschlag, R.; Tavan, P. *Eur. Biophys. J.* **2010**, *39*, 1177−1192.

(19) Denschlag, R.; Schreier, W. J.; Rieff, B.; Schrader, T. E.; Koller, F. O.; Moroder, L.; Zinth, W.; Tavan, P. *Phys. Chem. Chem. Phys.* **2010**, *12*, 6204−6218.

(20) Greengard, L.; Rokhlin, V. *J. Comput. Phys.* **1987**, *73*, 325 −348.

(21) Appel, A. A. *SIAM J. Sci. Stat. Comput.* **1985**, *6*, 85−103.

(22) Barnes, J.; Hut, P. *Nature* **1986**, *324*, 446−449.

(23) Figueirido, F.; Levy, R. M.; Zhuo, R.; Berne, B. J. *J. Chem. Phys.* **1997**, *106*, 9835−9849.

(24) Challacombe, M.; White, C.; Head-Gordon, M. *J. Chem. Phys.* **1997**, *107*, 10131−10139.

(25) Amisaki, T. *J. Comput. Chem.* **2000**, *21*, 1075−1087.

(26) Ding, H.-Q.; Karasawa, N.; Goddard, W. A. III *J. Chem. Phys.* **1992**, *97*, 4309−4315.

(27) Ding, H.-Q.; Karasawa, N.; Goddard, W. A. III *Chem. Phys. Lett.* **1992**, *196*, 6−10.

(28) Shimada, J.; Kaneko, H.; Takada, T. *J. Comput. Chem.* **1994**, *15*, 28−43.

(29) Warren, M. S.; Salmon, J. K. *Comput. Phys. Commun.* **1995**, *87*, 266 −290.

(30) Dehnen, W. *Astrophys. J.* **2000**, *536*, L39 −L42.

(31) Dehnen, W. *J. Comput. Phys.* **2002**, *179*, 27 −42.

(32) Martinetz, T.; Berkovich, S.; Schulten, K. *IEEE Trans. Neural Networks* **1993**, *4*, 558−569.

(33) Dersch, D. R.; Tavan, P. In Proceedings of the International Conference on Artificial Neural Networks 1994, (ICANN'94), , Sorrento, Italy, May 26-29, 1994; Moreno, M., Morasso, P., Eds.; Springer: London, 1994; pp 1067−1070.

(34) Takahahi, K. Z.; Narumi, T.; Yasuoka, K. *J. Chem. Phys.* **2011**, *135*, 174108.

(35) Wu, X.; Brooks, B. R. *J. Chem. Phys.* **2005**, *122*, 044107.

(36) Shanker, B.; Huang, H. *J. Comput. Phys.* **2007**, *226*, 732−753.

(37) Schmitz, M.; Tavan, P. In *Modern methods for theoretical physical chemistry of biopolymers*; Tanaka, S., Lewis, J., Eds.; Elsevier: Amsterdam, The Netherlands, 2006; Chapter 8, pp 157−177.

(38) Hinsen, K.; Felderhof, B. U. *J. Math. Phys.* **1992**, *33*, 3731−3735.

(39) Buckingham, A. D. *Adv. Chem. Phys.* **1967**, *12*, 107−147.

(40) Eichinger, M.; Tavan, P.; Hutter, J.; Parrinello, M. *J. Chem. Phys.* **1999**, *110*, 10452−10467.

(41) Hutter, J.; Alavi, A.; Deutsch, T.; Bernasconi, M.; Goedecker, S.; Marx, D.; Tuckerman, M.; Parrinello, M. *CPMD: Car−Parinello Molecular Dynamics*, version 3.10; IBM Corporation and Max-Planck Institut, Stuttgart: Armonk, NY and Stuttgart, Germany, 1997; www.cpmd.org.

(42) Laio, A.; VandeVondele, J.; Rothlisberger, U. *J. Chem. Phys.* **2002**, *116*, 6941−6947.

(43) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.

(44) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129−145.

(45) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.

(46) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.

(47) Kräutler, V.; van Gunsteren, W. F.; Hünenberger, P. H. *J. Comput. Chem.* **2001**, *22*, 501−508.

(48) Verlet, L. *Phys. Rev.* **1967**, *159*, 98−103.