The Journal of Chemical Physics

RESEARCH ARTICLE | OCTOBER 20 2023

Smooth particle mesh Ewald-integrated stochastic Lanczos many-body dispersion algorithm

Pier Paolo Poier ; Louis Lagardère ; Jean-Philip Piquemal

Check for updates

J. Chem. Phys. 159, 154109 (2023)
https://doi.org/10.1063/5.0166476

View Online Export Citation

Articles You May Be Interested In

Optimizing working parameters of the smooth particle mesh Ewald algorithm in terms of accuracy and efficiency

J. Chem. Phys. (July 2010)

Tuning the smooth particle mesh Ewald sum: Application on ionic solutions and dipolar fluids

J. Chem. Phys. (November 2014)

Gaussian split Ewald: A fast Ewald mesh method for molecular simulation

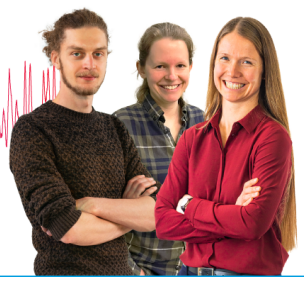
J. Chem. Phys. (January 2005)

11 November 2025 13:46:34

Webinar From Noise to Knowledge

May 13th – Register now

Zurich Instruments Universität Konstanz



AIP Publishing

The Journal of Chemical Physics ARTICLE pubs.aip.org/aip/jcp

Smooth particle mesh Ewald-integrated stochastic Lanczos many-body dispersion algorithm

Cite as: J. Chem. Phys. 159, 154109 (2023); doi: 10.1063/5.0166476
Submitted: 5 July 2023 • Accepted: 2 October 2023 •
Published Online: 20 October 2023

View Online Export Citation CrossMark

Pier Paolo Poier,[1,a] Louis Lagardère,[1,2] and Jean-Philip Piquemal[1,3,b]

AFFILIATIONS

[1] Laboratoire de Chimie Théorique, Sorbonne Université, 75005 Paris, France
[2] Sorbonne Université, IP2CT, FR 2622 CNRS, Paris, France
[3] Department of Biomedical Engineering, The University of Texas at Austin, Austin, Texas 78712, USA

[a] **Electronic mail:** pier.poier@sorbonne-universite.fr
[b] **Author to whom correspondence should be addressed:** jean-philip.piquemal@sorbonne-universite.fr

ABSTRACT

We derive and implement an alternative formulation of the Stochastic Lanczos algorithm to be employed in connection with the Many-Body Dispersion model (MBD). Indeed, this formulation, which is only possible due to the Stochastic Lanczos' reliance on matrix-vector products, introduces generalized dipoles and fields. These key quantities allow for a state-of-the-art treatment of periodic boundary conditions via the $\mathcal{O}(N \log(N))$ Smooth Particle Mesh Ewald (SPME) approach which uses efficient fast Fourier transforms. This SPME-Lanczos algorithm drastically outperforms the standard replica method which is affected by a slow and conditionally convergence rate that limits an efficient and reliable inclusion of long-range periodic boundary conditions interactions in many-body dispersion modelling. The proposed algorithm inherits the embarrassingly parallelism of the original Stochastic Lanczos scheme, thus opening up for a fully converged and efficient periodic boundary conditions treatment of MBD approaches.

Published under an exclusive license by AIP Publishing. https://doi.org/10.1063/5.0166476

11 November 2025 13:46:34

I. INTRODUCTION

Electron correlation is one of the most fascinating and difficult phenomenon to model. Dispersion in particular originates from the long-range electronic correlation among distant electron densities and represents the purely attractive contribution in van der Waals interactions. These are ubiquitous in nature: they can be for example observed in milk as they drive the formation of lipid droplets that, through light scattering, give to milk its typical white color. Geckos and spiders, on the other hand, also take advantage of dispersion for supporting their entire weight on smooth vertical surfaces.

From the microscopic point of view, dispersion interactions are crucial in many processes driven by non-covalent phenomena such as protein folding, protein-protein interactions, supra molecular and inter-molecular interactions in general.

An exact modelization of dispersion requires the analytical solution of the electronic Schrödinger equation, which is unfortunately impossible for practical cases. In the past decades, very accurate numerical wave function-based quantum chemical methods have been developed to tackle electron correlation, thus implicitly capable of describing dispersion and intermolecular interactions.[1,2] These methodologies, however, can only be applied to molecules composed of very few atoms, thus preventing the study of chemically and biologically relevant systems.

The advent of Density Functional Theory (DFT) represents a milestone in quantum chemistry as it provides a cheap way of including electronic correlation, as its computational cost is similar to that of the Hartree–Fock method. Nevertheless, the intrinsic local nature of common exchange-correlation functionals, makes DFT inadequate for describing long-range correlation effects, thus dispersion. To retain the DFT scaling benefits, extensive efforts have been spent in the past years in developing dispersion corrections able to improve the DFT capability of describing intermolecular interactions, crucial in material design and molecular modelling in general.

Many of these correction techniques rely on simple empirical pairwise treatments of dispersion, similar to those embraced in force fields. Their simplicity, together with the negligible computa-

Published under an exclusive license by AIP Publishing

tional cost and the good accuracy improvement, made possible for these methods to be included in most of the quantum chemistry softwares.[3]

Despite their large diffusion, these pairwise corrections completely neglect the many-body nature of dispersion interactions inherited from the long-range electronic correlation on the basis of these phenomena.

In recent years, the interest towards Many-Body Dispersion correction models has risen.[4] In particular the MBD@rSCS model by Tkatchenko, Di Stasio and Ambrosetti, together with its variations, has become especially popular by virtue of its high accuracy obtained despite of the absence of empirical parameters except for a single range-separation parameter for the coupling between the long-range Many-Body Dispersion model (MBD) energy and the chosen DFT functional.[5–7]

The MBD@rsSCS model can be summarized as follows. First, a set of atomic dipole polarizabilities are obtained from the partitioning of the molecular electron density or, alternatively, retrieved from a deep-neuronal network as recently proposed.[8,9] Secondly, the polarizabilities are made frequency-dependent via Padé approximation and subsequently a Dyson-like self-consistent screening linear equation is solved for a selected set of frequencies. Lastly, the set of screened frequency-dependent polarizabilities are used as key quantities in building the MBD interaction matrix which spectrum is used to express the final many-body dispersion energy.

Compared to the $\mathcal{O}(N^4)$ scaling of Kohn–Sham equations' resolution, the MBD@rsSCS model involves a small additional computational cost. However, for increasingly large systems, the $\mathcal{O}(N^3)$ scaling of the diagonalization procedure becomes no longer negligible and, it can even become a burden if coupled to $\mathcal{O}(N)$ DFT methods.

Recently, we have proposed and implemented an alternative resolution of the MBD key equations that overcomes this scaling issue that is based on the state-of-art Stochastic Lanczos (SL) trace estimation.[10] Due to the the sparsity of the matrices involved, it exhibits linear-scaling with the system size. The proposed stochastic Lanczos MBD approach (SL-MBD) further benefits from an embarrassingly parallel implementation arising from its stochastic nature and this allows for reaching system sizes of hundred thousands atoms within a few minutes' time.[11]

Compared to a simple pairwise description, this many-body treatment of dispersion interactions in systems such as solvated proteins has revealed a higher degree of delocalization as well as a collective solute-solvent character leading to remarkable long-range interactions.[12] The potentially longer-range of MBD interactions stresses the importance of the inclusion of a coherent full periodic boundary condition (PBC) treatment, especially in highly ordered and periodic systems. In this direction, recent efforts have been spent in past years. Bucko and co-workers have provided a method expanding over the Brillouin cell that introduced consistent improvements compared to the standard replica method used to include long-range periodic boundary conditions effects.[13]

By virtue of the above mentioned long-range nature of MBD interactions, it is of broad interest to generalize the SL-MBD approach to a full PBC treatment. However, the quadratic-scaling approaches typically employed in connection to MBD models are clearly not suitable to be integrated in the SL-MBD methodology for both memory requirements and computational efficiency due to the large systems targeted. A more sophisticated approach has therefore to be developed.

In the context of long-range electrostatics modelling, this scaling limitation was addressed via Ewald summation techniques, as they formally scale as $\mathcal{O}(N^2)$ but a proper optimization lowers the factor to $\mathcal{O}(N^{3/2})$. Ewald summation techniques replace the original conditionally convergent energy summation with a direct and reciprocal space absolutely convergent ones consisting of a real and reciprocal summations as well as a self interaction term. The Particle Mesh Ewald (PME) method proposed by Darden *et al.*, drastically improved Ewald summation technique's associated performance.[14] Its idea relies on the efficient calculation of the reciprocal space energy contribution thanks to fast Fourier transforms scaling as $\mathcal{O}(N\log(N))$. The PME method with its different variants especially the Smooth Particle Mesh Ewald (SPME),[15] has become the standard algorithm implemented in nearly all the most efficient Molecular Dynamics packages thanks to its scaling features although alternative but related methods also exist.[16]

In this work, we derive and present a modification of the SL-MBD method based on a PME treatment of periodic boundary conditions. The resulting Smooth Particle Mesh Ewald stochastic Lanczocz (SPME-SL) MBD approach is suitable for large systems as it exhibits the typical $\mathcal{O}(N\log(N))$ scaling inherited from the PME method.

In Sec. II, we review the MBD model as well as the stochastic Lanczos method in its standard form. A theory section is then dedicated to the derivation of the modified SPME-based Lanczos quadrature scheme followed by a section dedicated to numerical results where the computational performances of the method are discussed and compared to the ones of the standard replica method.

II. REVIEW OF THE MBD AND SL-MBD

The MBD model is based on the idea that a molecule is described as a set of interacting quantum harmonic oscillators, which Hamiltonian is shown in Eq. (1), $\mathbf{d}_i = \sqrt{m_i}\boldsymbol{\xi}_i$ being the mass-weighted dipole moment displaced by the vector $\boldsymbol{\xi}_i$ from its equilibrium position. $\alpha_i(0)$ and ω_i represent the model's key parameters and correspond to the static dipole polarizability and characteristic excitation frequency respectively.

$$\begin{aligned}\hat{H}_{\mathrm{MBD}} &= \frac{1}{2}\sum_{i=1}^{N}\left(-\hat{\nabla}^2_{\mathbf{d}_i} + \mathbf{d}_i^\dagger \mathbf{V}_{ii}\mathbf{d}_i\right) + \sum_{i>j}\mathbf{d}_i^\dagger \mathbf{V}_{ij}\mathbf{d}_j \\ \mathbf{V}_{ij} &= \mathbf{I}_3\delta_{ij}\omega_i^2 + (1-\delta_{ij})\omega_i\omega_j\sqrt{\alpha_i(0)\alpha_j(0)}\mathbf{T}'_{ij}(\beta)\end{aligned} \tag{1}$$

These parameters are obtained from *ab initio* data as the atom-in-molecule (AIM) polarizability is typically retrieved via partitioning of the electron density while ω_i is defined in terms of accurate free atoms quantities.[17–19]

We note that these AIM polarizability parameters can be screened by solving a Dyson-like equation[5] that can be solved extremely efficiently,[11] however, we will not discuss this in the present work as the presented algorithm is general and does not depend neither on the choice of AIM polarizabilities nor on their screening. We further mention that recently Johnson and co-workers have analyzed the sensitivity of the screening procedure for selected systems.[20]

11 November 2025 13:46:34

Published under an exclusive license by AIP Publishing

The $\mathbf{T}'_{ij}(\beta)$ term is built from the pure point dipole-dipole interaction tensor for the ij atom pair that is screened via a damping function $s(R_{ij};\ \beta)$ depending on the interactomic distance R_{ij} and the single range-separation parameter β typically optimized for the correspondent DFT functional to be dispersion-corrected, $\mathbf{T}'_{ij}(\beta) = s(R_{ij};\beta)\mathbf{T}_{ij}$. Recently the MBD model was generalized to higher than dipole interactions,[9,21] however, here we will only consider the dipole-dipole interaction case. For the explicit expression of $\mathbf{T}'$ we refer to the work in Ref. 11 The eigenvalues (λ_i) of the MBD interaction matrix $\mathbf{V}$, shown for the ij block in Eq. (1), are required to obtain the MBD energy $\mathscr{E}_{\mathrm{MBD}}$ via the plasmonic formula shown in Eq. (2) that represents the correlation energy of the interacting fluctuating dipoles.

$$\mathscr{E}_{\mathrm{MBD}} = \frac{1}{2}\sum_{i=1}^{3N}\sqrt{\lambda_i} - \frac{3}{2}\sum_{i=1}^{N}\omega_i \tag{2}$$

The solution of Eq. (2) is bound to the $\mathcal{O}(N^3)$ scaling of the diagonalization step that, as mentioned earlier, strongly limits the applicability of the method to large systems.

The SL-MBD method bypasses the diagonalization of $\mathbf{V}$ by exploiting the alternative but equivalent expression of the plasmonic formula, Eq. (3), where the sum over the whole spectrum of $\mathbf{V}$ is rewritten in term of its trace, that is invariant under any change of basis, namely $\sum_{i=1}^{3N}\sqrt{\lambda_i} = \mathrm{Tr}\,(\sqrt{\mathbf{\Lambda}}) = \mathrm{Tr}\,(\sqrt{\mathbf{V}})$ where $\mathbf{\Lambda}$ is the diagonal form of $\mathbf{V}$ obtained via the unitary transformation $\mathbf{\Lambda} = \mathbf{W}^{\dagger}\mathbf{V}\mathbf{W}$.

$$\mathscr{E}_{\mathrm{MBD}} = \frac{1}{2}\,\mathrm{Tr}\,(\sqrt{\mathbf{V}}) - \frac{3}{2}\sum_{i=1}^{N}\omega_i \tag{3}$$

The evaluation of the trace of a symmetric matrix function such as $\mathrm{Tr}\,[\sqrt{\mathbf{V}}]$ is, in the proposed SL-MBD, based on two main assumptions.

First, the stochastic Hutchinson trace estimator (HTE)[22] is invoked, Eq. (4), $\mathbf{v}_l$ being one of the R normalized random vectors of dimension D (in our case $D = 3N$), which entries follow a Rademacher distribution, i.e. they can assume values of either 1 or −1 with the same probability.

$$\mathrm{Tr}\,[\sqrt{\mathbf{V}}] \approx \frac{D}{R}\sum_{l=1}^{R}\mathbf{v}_l^{\dagger}\sqrt{\mathbf{V}}\mathbf{v}_l \tag{4}$$

$$\mathbf{v}_l = \frac{\mathbf{u}_l}{\|\mathbf{u}_l\|}$$
$$u_{l,i} = \begin{cases} 1, & \Pr = 1/2 \\ -1, & \Pr = 1/2 \end{cases} \tag{5}$$

Second, each of the R scalar expectation values in Eq. (4) can be expressed in terms of $\mathrm{Tr}\,[\sqrt{\mathbf{\Lambda}}]$ and the unitary transformation $\mathbf{W}$ as reported in Eq. (6) where we introduced $\boldsymbol{\mu}_l = \mathbf{W}^{\dagger}\mathbf{v}_l$.

$$\mathbf{v}_l^{\dagger}\sqrt{\mathbf{V}}\mathbf{v}_l = \mathbf{v}_l^{\dagger}\mathbf{W}\sqrt{\mathbf{\Lambda}}\mathbf{W}^{\dagger}\mathbf{v}_l = \sum_{i}^{D}\mu_{l,i}^2\sqrt{\lambda_i} \tag{6}$$

The last equality in Eq. (6) corresponds to the Riemann–Stieltjes integral[23] defined in Eq. (7) which is approximated via the general $(M+1)$-points quadrature shown in Eq. (8), $\{\tau_k\}$ and $\{\theta_k\}$ representing the unknown weights and nodes respectively.

$$\sum_{i}^{D}\mu_{l,i}^2\sqrt{\lambda_i} = \int_a^b\sqrt{t}\,d\mu(t)$$
$$\mu(t) = \begin{cases} 0, & t < a = \lambda_1 \\ \sum_{j=1}^{i-1}\mu_j^2, & \lambda_{i-1} \le t < \lambda_i \\ \sum_{j=1}^{D}\mu_j^2, & b = \lambda_n < t \end{cases} \tag{7}$$

$$\mathbf{v}_l^{\dagger}\sqrt{\mathbf{V}}\mathbf{v}_l = \int_a^b\sqrt{t}\,d\mu(t) \approx \sum_{k=1}^{M+1}\tau_k^{(l)}\sqrt{\theta_k} \tag{8}$$

By inserting Eq. (8) in Eq. (4), one can identify the complete expression for the stochastic trace estimation, Eq. (9).

$$\mathrm{Tr}\,[\sqrt{\mathbf{V}}] \approx \frac{D}{R}\sum_{l=1}^{R}\sum_{k=1}^{M+1}\tau_k^{(l)}\sqrt{\theta_k^{(l)}} \tag{9}$$

In the stochastic Lanczos algorithm, the nodes and weights for the quadrature relative to each of the l-th terms in the first summation, are identified as the eigenvalues $\{\widetilde{\lambda}_k^{(l)}\}$ and the first entry (squared) of the eigenvectors $\{[U_{1,k}^{(l)}]^2\}$ of the tridiagonal $\mathbf{\Delta}^{(l)}$ matrix which is the representation of the original MBD potential matrix $\mathbf{V}$ in the $M+1$ Krylov subspace $\mathscr{K}_{M+1} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{M+1}\}$ where the basis vectors are gathered as the $\mathbf{Y}^{(l)}$ matrix's columns.

$$\mathbf{\Delta}^{(l)} = \mathbf{Y}^{\dagger(l)}\mathbf{V}\mathbf{Y}^{(l)} \tag{10}$$

$$\widetilde{\mathbf{\Lambda}}^{(l)} = \mathbf{U}^{(l)\dagger}\mathbf{\Delta}^{(l)}\mathbf{U}^{(l)} \tag{11}$$

The solution of Eq. (10) represents the crucial part of the algorithm in terms of efficiency while Eq. (11), by virtue of the small matrices involved (Krylov subspace dimension rarely exceeding 15), is inexpensive and it is solved by means of standard libraries.

Equation (10) is practically solved as follows: For each of the R terms employed in the HTE, $\mathbf{v}$ [from now on the upperscript (l) is dropped for simplicity] is taken as the first basis vector of the Krylov subspace $(\mathbf{y}_1)$ while the remaining basis vectors $\{\mathbf{y}_k\}$ (columns of $\mathbf{Y}$) and the diagonal (Δ_{kk}) and out-of diagonal $(\Delta_{(k-1)k} = \Delta_{k(k-1)})$ elements of $\mathbf{\Delta}$ are retrieved recursively as shown in Eq. (12) where the asterisk denotes the unnormalized k-th basis vector.

$$\begin{gathered} \mathbf{y}_1 = \mathbf{v} \\ b_k\mathbf{y}_k = \mathbf{y}_k^* = \mathbf{l}_{k-1} - a_{k-1}\mathbf{y}_{k-1} - b_{k-1}\mathbf{y}_{k-2} \\ \mathbf{l}_k = \mathbf{V}\mathbf{y}_k \\ b_k = \sqrt{\mathbf{y}_k^{*\dagger}\mathbf{y}_k^*} = \Delta_{(k-1)k} = \Delta_{k(k-1)} \\ a_k = \mathbf{y}_k^{\dagger}\mathbf{V}\mathbf{y}_k = \mathbf{y}_k^{\dagger}\mathbf{l}_k = \Delta_{kk} \end{gathered} \tag{12}$$

$$\mathbf{\Delta}^{(l)} = \begin{pmatrix} \Delta_{11}^{(l)} & \Delta_{12}^{(l)} & 0 & 0 & 0 \\ \Delta_{21}^{(l)} & \ddots & \ddots & 0 & 0 \\ 0 & \ddots & \Delta_{kk}^{(l)} & \ddots & 0 \\ 0 & 0 & \ddots & \ddots & \Delta_{(M)(M+1)}^{(l)} \\ 0 & 0 & 0 & \Delta_{(M+1)(M)}^{(l)} & \Delta_{(M+1)(M+1)}^{(l)} \end{pmatrix}$$

11 November 2025 13:46:34

Published under an exclusive license by AIP Publishing

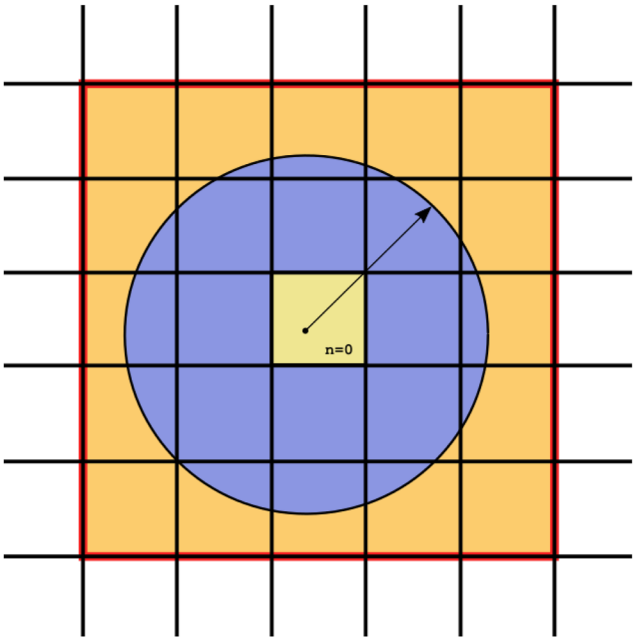


FIG. 1. Pictorial representation of the replica method for a 2-D squared box of side L where the chosen cutoff radius is R_{cut}. Given the ratio $x = R_{cut}/L$, a supercell (yellow ochre delimited by red boundary) with vertices identified from all the four possible pairs of integers $(\pm n_{max}, \pm n_{max})$ is built. n_{max} represents the smallest integer value that is greater than or equal to x. A given particle belonging to the central cell ($\mathbf{n} = \mathbf{0}$) will therefore interact with other particles in supercell placed at a distance smaller than R_{cut}, i.e. within the blue circle.

In general the k-th iteration retrieves the Δ_{kk} diagonal element as well as the contiguous upper/lower $\Delta_{(k-1)k}$ and $\Delta_{k(k-1)}$ ones. In Sec. III, expressions for $\mathbf{y}$, a_k and b_k in the case of full PBC enforced via PME method will be derived.

III. THEORY

The easiest strategy for including PBC in the MBD model consists in looping over a selected number of cell vectors $\mathbf{n}$, each of which denoting the periodic image of the central simulation cell U defined by its edges $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ and with volume $V = \mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)$. This would result in the modified dipole-dipole interaction matrix $\mathbf{T}^{pbc}$ shown in Eq. (13) where $\mathbf{T}'_{ij}(j \in \mathbf{0})$ represents the ij interaction block belonging to the central simulation cell while $\mathbf{T}'_{ij}(j \in \mathbf{n})$ the interaction between the particle i and the particle j this time belonging to the cell's periodic replica identified with $\mathbf{n}$. In particular, the list of cells (and therefore their associated $\mathbf{n}$ vectors) are chosen according to a cutoff radius as pictorially represented in Fig. 1.

$$\begin{aligned} \mathbf{T}^{pbc}_{ij} &= \mathbf{T}'_{ij}(j \in \mathbf{0}) + \sum_{\mathbf{n}\neq\mathbf{0}} \mathbf{T}'_{ij}(j \in \mathbf{n}) \\ \mathbf{n} &= n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3 \qquad n_1, n_2, n_3 \in \mathbb{Z}^3 \end{aligned} \tag{13}$$

The substitution of $\mathbf{T}'_{ij}$ with $\mathbf{T}^{pbc}_{ij}$ inside $\mathbf{V}$ (often referred as to the replica method) and the subsequent use of its eigenvalues in Eq. (2) was discussed in Ref. 24. However, the use of truncated methos based on Eq. (13) involves the problematics listed and discussed below.

First, the summation in Eq. (13) represents a slowly and conditionally convergent series that characterizes not only dipole-dipole interactions, but also charge-charge, charge-dipole and charge-quadrupole Coulomb interactions kernels.[25]

Consequently, the slow convergence of Eq. (13) strongly limits the applicability of the SL-MBD algorithm where the efficient "on-the-fly" computation of each $\mathbf{V}_{ij}$ block is crucial for the evaluation of the $\mathbf{V}\mathbf{y}_k$ products discussed in connection to Eq. (12).

The Ewald summation (ES) method, as well as its more efficient PME variants, was design to improve over Eq. (13) since the conditionally convergent features of long-range electrostatic interactions of periodic systems are replaced by an absolutely convergent treatment.

Let's consider a set of N interacting dipoles belonging to the central simulation cell U and gathered into the $3N$-dimensional array $\mathbf{d}$. The correspondent electric field array $\mathbf{E} = \mathbf{T}^{pbc}\mathbf{d}$ arising from the dipoles in both the central simulation cell and all its periodic images is, in the ES method, expressed as the sum of three component, Eq. (14).

$$\mathbf{E} = \mathbf{T}^{pbc}\mathbf{d} \rightarrow \mathbf{E}^{\star} = \mathbf{E}^{dir} + \mathbf{E}^{rec} + \mathbf{E}^{self} \tag{14}$$

$\mathbf{E}^{dir}$ represents the direct space contribution to the Ewald electric field, the $\mathbf{E}^{rec}$ is the long-range term computed in Fourier (reciprocal) space while $\mathbf{E}^{self}$ represents the so called self-interaction term. The explicit expressions for each of these terms will be given later in the discussion, however, it is important to stress that each of these field components consist of absolutely convergent contributions as the resulting $\mathbf{E}^{\star}$ field.

Our strategy is thus to identify and isolate from the SL-MBD equations, Eq. (12), an electric field-like term that can be then evaluated according to the three absolute convergent contributions in Eq. (14), thus allowing us to include PBC in a robust and efficient manner.

To do so, we will now start by partitioning $\mathbf{V}$ into its diagonal and out-of-diagonal contributions given below, where $\mathbf{I}_3$ is a (3, 3) identity matrix.

$$\begin{aligned} \mathbf{V}_{ij} &= \omega_i\omega_j\sqrt{\alpha_i(0)\alpha_j(0)}\mathbf{T}'_{ij} \\ \mathbf{V}_{ii} &= \mathbf{I}_3\omega_i^2 \end{aligned} \tag{15}$$

Due to the fact that the diagonal blocks $\mathbf{V}_{ii}$ are themselves diagonal, we introduce the identity in Eq. (16), where $\mathbf{\Omega}$ is the diagonal matrix defined below and $\widetilde{\mathbf{V}}$ is the hollow matrix coposed of the off-diagonal entries of $\mathbf{V}$. These quantities will turn useful later in the discussion.

$$\begin{aligned} \mathbf{V} &= \mathbf{\Omega} + \widetilde{\mathbf{V}} \\ \mathbf{\Omega} &= \bigoplus_i^N \mathbf{V}_{ii} \end{aligned} \tag{16}$$

We further introduce the $\mathbf{g}$ vector (of dimension $3N$) defined as the concatenation of N three-dimensional vectors-of-ones $(\mathbf{1}_3)$ as shown in Eq. (17).

$$\mathbf{g} = \bigoplus_i^N \omega_i\sqrt{\alpha_i(0)}\mathbf{1}_3 \tag{17}$$

11 November 2025 13:46:34

Published under an exclusive license by AIP Publishing

At this point, we use the newly introduced quantities defined in Eq. (16) to rewrite the diagonal a_k term as shown in Eq. (18).

$$a_k = \mathbf{y}_k^\dagger \boldsymbol{\Omega} \mathbf{y}_k + \mathbf{y}_k^\dagger \widetilde{\mathbf{V}} \mathbf{y}_k \tag{18}$$

One can now easily prove that the second term on the right hand side of Eq. (18) can be rewritten in terms of the $\mathbf{g}$, Eq. (19), where $\odot$ denotes the Hadamard product.

$$\mathbf{y}_k^\dagger \widetilde{\mathbf{V}} \mathbf{y}_k = (\mathbf{y}_k \odot \mathbf{g})^\dagger \mathbf{T}' (\mathbf{g} \odot \mathbf{y}_k) \tag{19}$$

By inserting Eq. (19) into (18), we obtain an expression for a_k which will soon turn crucial for the discussion.

$$a_k = \mathbf{y}_k^\dagger \boldsymbol{\Omega} \mathbf{y}_k + (\mathbf{y}_k \odot \mathbf{g})^\dagger \mathbf{T}' (\mathbf{g} \odot \mathbf{y}_k) \tag{20}$$

The $3N$-dimensional term $(\mathbf{g} \odot \mathbf{y}_k)$ can be thought as a generalized dipole array $\mathbf{d}_k$ that, via the interaction tensor $\mathbf{T}$ originates the generalized field $\mathbf{E}_k = \mathbf{T}\mathbf{d}_k$ that can be then eventually computed according to Eq. (14).

$$a_k = \mathbf{y}_k^\dagger \boldsymbol{\Omega} \mathbf{y}_k + \mathbf{d}_k^\dagger \mathbf{E}_k^\star \tag{21}$$

We note in passing that the introduction of this generalized field can be used in different situations as it allows us to couple our system with en external perturbation that, as discussed in references, could arise from implicit solvent contribution.[26,27]

At this point we note from Eq. (12) (last equality) that a_k is related to $\mathbf{l}_k$ via a differentiation with respect to the basis vector $\mathbf{y}_k$. We can therefore differentiate Eq. (20) to finally obtain Eq. (23) where the rule for the differentiation of a commuting Hadamard product has been applied, Eq. (22). We note that a similar approach based on differentiation was adopted by Stamm and co-workers in deriving Ewald summation for arbitrary orders of multipoles with particular emphasis on the self term, for which different expressions can be found in literature.[28]

$$\frac{\partial(\mathbf{y}_k \odot \mathbf{g})}{\partial \mathbf{y}_k} = \frac{\partial \mathrm{Diag}(\mathbf{g})}{\partial \mathbf{y}_k}\mathbf{y}_k + \mathrm{Diag}(\mathbf{g})\frac{\partial \mathbf{y}_k}{\partial \mathbf{y}_k} = \mathrm{Diag}(\mathbf{g}) \tag{22}$$

$$\mathbf{l}_k = \frac{1}{2}\frac{\partial a_k}{\partial \mathbf{y}_k} = \boldsymbol{\Omega}\mathbf{y}_k + \mathrm{Diag}(\mathbf{g})\mathbf{T}'(\mathbf{g} \odot \mathbf{y}_k) \tag{23}$$

Once again we use the definition of the generalized dipole and field to finally obtain Eq. (24).

$$\begin{aligned} \mathbf{l}_k &= \boldsymbol{\Omega}\mathbf{y}_k + \mathrm{Diag}(\mathbf{g})\mathbf{T}'\mathbf{d}_k \\ &= \boldsymbol{\Omega}\mathbf{y}_k + \mathrm{Diag}(\mathbf{g})\mathbf{E}_k^\star \end{aligned} \tag{24}$$

Equation (12) can therefore be rewritten in terms of the generalized electric field $\mathbf{E}_k^\star$ through the above derived quantities, Eq. (25).

$$\begin{aligned} \mathbf{y}_1 &= \mathbf{v} \\ b_k \mathbf{y}_k = \mathbf{y}^*_k &= \mathbf{l}_{k-1} - a_{k-1}\mathbf{y}_{k-1} - b_{k-1}\mathbf{y}_{k-2} \\ \mathbf{l}_k &= \boldsymbol{\Omega}\mathbf{y}_k + \mathrm{Diag}(\mathbf{g})\mathbf{E}_k^\star \\ b_k &= \sqrt{\mathbf{y}_k^{*\dagger}\mathbf{y}_k^*} = \Delta_{(k-1)k} = \Delta_{k(k-1)} \\ a_k &= \mathbf{y}_k^\dagger \boldsymbol{\Omega}\mathbf{y}_k + \mathbf{d}_k^\dagger \mathbf{E}_k^\star = \Delta_{kk} \end{aligned} \tag{25}$$

$\mathbf{E}_k^\star$ can be evaluated by ES and the explicit expressions for $\mathbf{E}^{\mathrm{dir}}$, $\mathbf{E}^{\mathrm{self}}$ and $\mathbf{E}^{\mathrm{rec}}$ are shown below, however, for a broader discussion and derivation we refer to the following Refs. 25, 29, and 30.

Starting from the direct component, we identify the three dimensional electric field $\vec{\mathbf{E}}^{\mathrm{dir}}_{i,k}$ at the atomic position $\mathbf{R}_i$ arising from the generalized dipole array $\mathbf{d}_k$, where its three-dimensional contribution related to the j-th atom is denoted $\vec{\mathbf{d}}_{j,k}$, as shown in Eq. (26).

$$\begin{aligned} L_{j,k} &= \vec{\mathbf{d}}_{j,k}\nabla_j \\ \vec{\mathbf{E}}^{\mathrm{dir}}_{i,k} &= -\sum_{\mathbf{n}}\sum_{j=1}^{N^*} L_{j,k}\frac{\partial}{\partial \mathbf{R}_i}\left(\frac{\mathrm{erfc}(\tau\,|\,\mathbf{R}_j - \mathbf{R}_i + \mathbf{n}\,|)}{|\,\mathbf{R}_j - \mathbf{R}_i + \mathbf{n}\,|}\right. \\ &\quad \left. + \sum_{j=1}^{N^*} L_{j,k}(1 - s_{ij})\mathbf{T}_{ij}\vec{\mathbf{d}}_{j,k}\right) \end{aligned} \tag{26}$$

In the above, τ represents a real parameter governing the balance between the direct and reciprocal contributions. For a cubic cell of side h, it is typically taken to be $5/h$.[31] τ is commonly chosen so that the direct term convergence is fast as the reciprocal contribution can be efficiently computed via FFT. This makes the summation over $\mathbf{n}$ fastly converging, and only particles belonging to neighboring periodic images are therefore usually considered. $\mathbf{E}^{\mathrm{dir}}$ is practically computed by means of neighbor lists based on the choice of τ determining the suitable cutoff and this ensures an efficient and linear-scaling evaluation.

The self term $\vec{\mathbf{E}}^{\mathrm{self}}_{i,k}$ consists in the single term shown in Eq. (27) which evaluation involves a negligible computational effort.

$$\vec{\mathbf{E}}^{\mathrm{self}}_{i,k} = \frac{2\tau^3}{3\sqrt{\pi}}\vec{\mathbf{d}}_{i,k} \tag{27}$$

From a computational point of view, with standard τ parameters, the most expensive and thus crucial term to evaluate is represented by the $\vec{\mathbf{E}}^{\mathrm{rec}}_{i,k}$ contribution. In order to discuss its explicit expression, we introduce the reciprocal conjugate vectors $(\mathbf{a}_1^*, \mathbf{a}_2^*, \mathbf{a}_3^*)$ which are related to their dual set by $\mathbf{a}_\alpha^* \cdot \mathbf{a}_\beta = \delta_{\alpha\beta}$, with $\alpha, \beta = \{1, 2, 3\}$ and $\delta_{\alpha\beta}$ being the Kronecker delta. In analogy to what done for $\mathbf{n}$, we define $\mathbf{m}$.

$$\mathbf{m} = m_1\mathbf{a}_1^* + m_2\mathbf{a}_2^* + m_3\mathbf{a}_3^* \quad m_1, m_2, m_3 \in \mathbb{Z}^3 \tag{28}$$

We further introduce the structure factor $S(\mathbf{m})$, defined in Eq. (29) for a given $\mathbf{m}$ is defined in.

$$S(\mathbf{m}) = \sum_{j=1}^{N} \vec{\mathbf{d}}_{j,k} \cdot \mathbf{m}\, \exp(2i\pi \mathbf{m} \cdot \mathbf{R}_j) \tag{29}$$

In the Ewald summation method the reciprocal component of the field is finally given in Eq. (30)

$$\begin{aligned} \vec{\mathbf{E}}^{\mathrm{rec}}_{i,k} &= -\frac{1}{\pi V}\sum_{\mathbf{m}\neq\mathbf{0}}\frac{\partial}{\partial \mathbf{R}_i}\left(\frac{\exp(-\pi^2\mathbf{m}^2/\tau^2)}{\mathbf{m}^2}S(\mathbf{m})\right. \\ &\quad \left. \times \exp(-2i\pi \mathbf{m} \cdot \mathbf{R}_i)\right) \end{aligned} \tag{30}$$

The optimal choice of τ makes the evaluation of Eq. (30) (and therefore of the whole ES method) $\mathcal{O}(N^{3/2})$ scaling, however, the PME

11 November 2025 13:46:34

Published under an exclusive license by AIP Publishing

method sensibly improves the scaling by approximating the complex exponentials via interpolation. In the Smooth PME method (SPME) in particular, the complex exponentials are first rewritten in terms of the scaled fractional coordinates $u_{\alpha j}$, Eq. (31), and then interpolated by a p-degree B-spline function $\theta_p(u_{\alpha j} - n_\alpha)$ on a grid of size $K_1 \times K_2 \times K_3$ and the final contribution due to the reciprocal space is given in Eq. (32)

$$\begin{aligned} u_{\alpha j} &= K_\alpha \mathbf{a}_\alpha^* \cdot \mathbf{R}_j \qquad \alpha = \{1,2,3\} \quad , \quad K_\alpha \in \mathbb{N}^+ \\ \exp(2i\mathbf{m}\cdot\mathbf{R}_j) &= \prod_{\alpha=1}^{3} \exp\left(2i\pi m_\alpha \frac{u_{\alpha_j}}{K_\alpha}\right) \end{aligned} \tag{31}$$

ALGORITHM 1. Schematic general representation of the SPME-based stochastic Lanczos algorithm.

1: SPME Grid allocation (K_1, K_2, K_3) and initialization
2: Neighbor list (direct space) generation
3: **for** $(l = 1, R)$ **do**
4: Generate $\mathbf{v}_l$ from a Rademacher distribution
5: $\mathbf{y}_1^{(l)} = \mathbf{v}_l$
6: call SPME-Lanczos $(k = 1)$: $\Delta_{11}^{(l)}$
7: **for** $(k = 2, M + 1)$ **do**
8: call SPME-Lanczos (general k): $\mathbf{y}_k^{(l)}$, $a_k^{(l)}$, $b_k^{(l)}$
9: **end for**
10: Eigendecomposition of $\mathbf{\Delta}^{(l)} : \mathbf{U}^{\dagger(l)}\mathbf{\Delta}^{(l)}\mathbf{U}^{(l)} = \widetilde{\mathbf{\Lambda}}^{(l)}$
11: **for** $(k = 1, M + 1)$ **do**
12: $E_{\text{sum}} = \text{E}_{\text{sum}} + [U_{1,k}^{(l)}]^2 \sqrt{\widetilde{\lambda}_k^{(l)}}$
13: **end for**
14: **end for**
15: Calculate the average over samples: $\text{Tr}\left[\sqrt{\mathbf{V}}\right] \approx \frac{3}{R} N E_{sum}$

ALGORITHM 2. Schematic representation of the k-th iteration SPME-Lanczos routine.

1: **Input quantities** $\longleftarrow \mathbf{y}_{k-1}, \mathbf{y}_{k-2}, \mathbf{l}_{k-1}, a_{k-1}, b_{k-1}$
2: build $\mathbf{y}_k^*$ from input quantities
3: get b_k from the normalization of $\mathbf{y}_k^*$
4: build the generalized dipole vector $\mathbf{d}_k$ from $\mathbf{y}_k$
5: distribute $\mathbf{d}_k$ in the (K_1, K_2, K_3) grid (necessary for $\mathbf{E}_k^{\text{rec}}$)
6: compute the generalized field $\mathbf{E}_k^\star = \mathbf{E}_k^{\text{dir}} + \mathbf{E}_k^{\text{self}} + \mathbf{E}_k^{\text{rec}}$
7: compute $\mathbf{l}_k$ from $\mathbf{E}_k^\star$
8: compute a_k
9: **Output quantities** $\rightarrow \mathbf{y}_k, b_k, \mathbf{l}_k, a_k$

The $\vec{\mathbf{E}}_{i,k}^{\text{rec}}$ is finally given by Eq. (32).

$$\vec{\mathbf{E}}_{i,k}^{\text{rec}} \approx -\frac{\partial}{\partial \mathbf{R}_i} \sum_{\mathbf{n}} \prod_{\alpha=1}^{3} \theta_p(u_{\alpha,i} - n_\alpha)(G^R * D^R)(\mathbf{n}) \tag{32}$$

The $(G^R * Q^R)$ term is the convolution between the pair potential G^R discussed by Sagui *et al.* and the real space dipole array D^R defined below.[32] The use of fast Fourier transtorms in the evaluation of (32) ensures an overall $\mathcal{O}(N \log(N))$ scaling.

$$\begin{aligned} D_k^R(k_1,k_2,k_3) = \sum_{\mathbf{n}} \sum_j L_{j,k} \theta_p(u_{1,j-k_1-K_1 n_1}) \theta_p \\ \times (u_{2,j-k_2-K_2 n_2}) \theta_p(u_{3,j-k_3-K_3 n_3}) \end{aligned} \tag{33}$$

The discussed algorithm is schematically reported in Algorithm 1 in its global structure while a more detailed scheme relative to the specific k-th iteration is given in Algorithm 2.

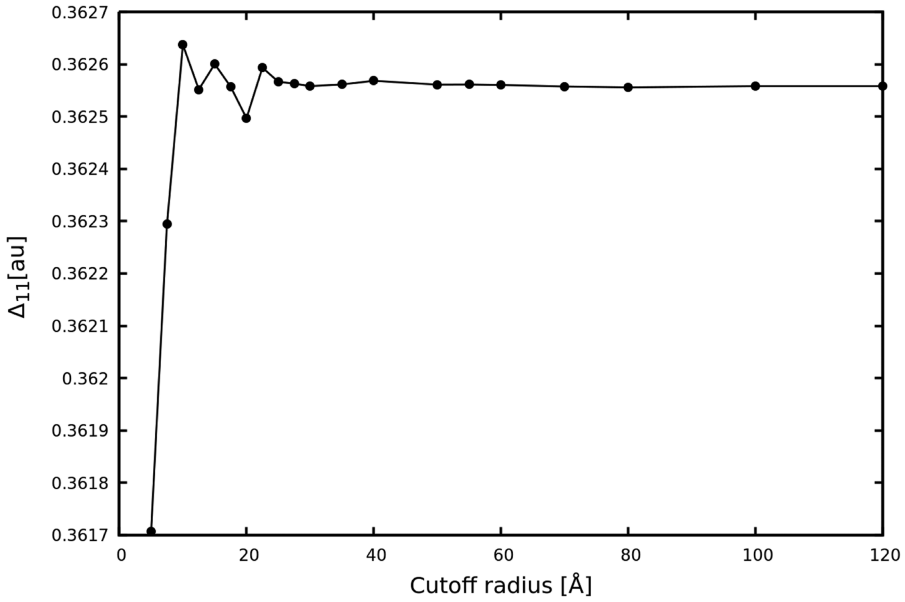


FIG. 2. First diagonal element of Δ computed via the replica method as a function of the cutoff radius for the cubic water box taken as test system.

11 November 2025 13:46:34

Published under an exclusive license by AIP Publishing

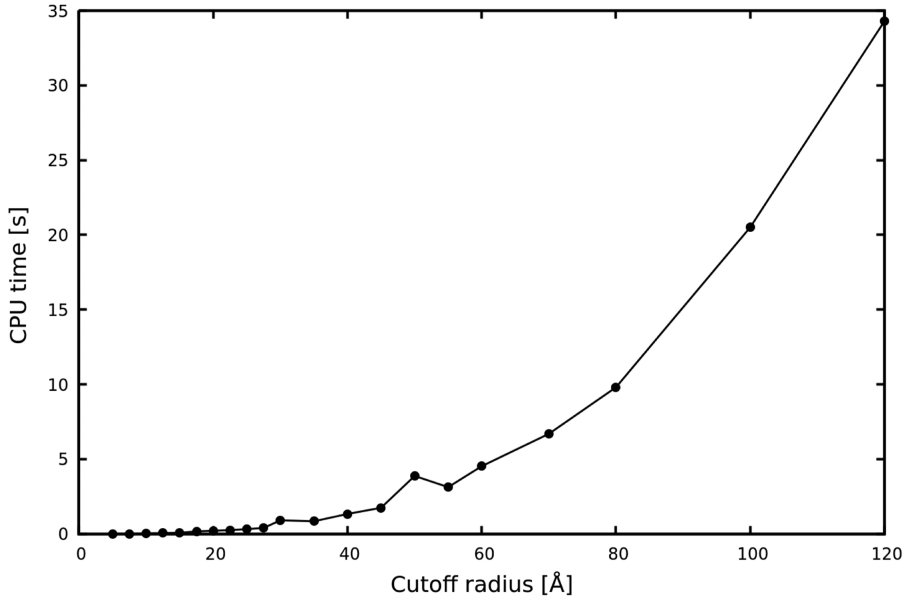


FIG. 3. CPU time as a function of the cutoff radius relative to Fig. 2, i.e. waterbox.

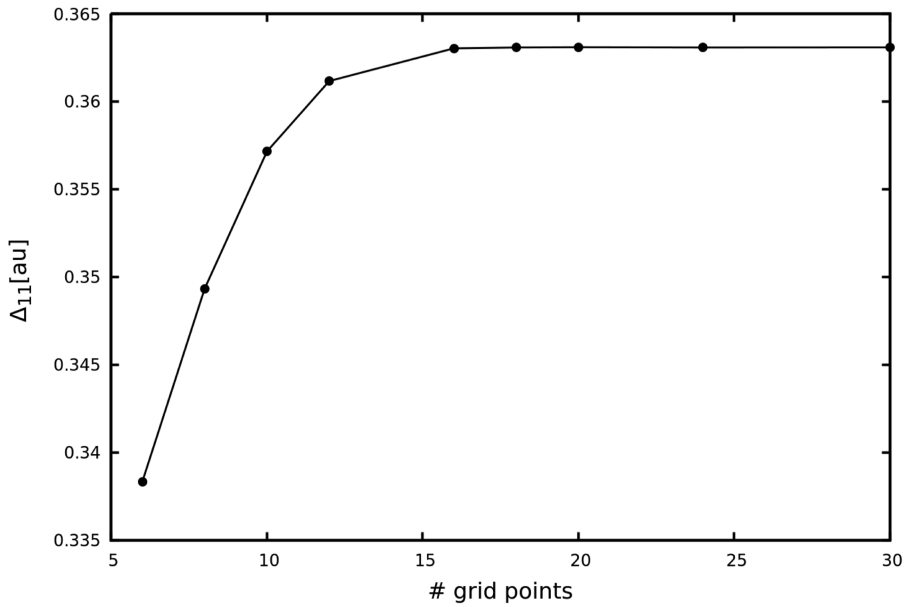


FIG. 4. First diagonal element of Δ computed via the SPME-Lanczos method as a function of the grid points for the waterbox considered (only K_1 is reported as the box is cubic). The initial Krylov subspace basis vector $\mathbf{y}_1$ is the same as for this figure.

Section IV will report numerical results relative to the algorithm's implementation in the Tinker-HP molecular dynamics package.[33,34] The replica method [Eq. (13)] has also been implemented and coupled to the SL-MBD method as this will allow us to perform a direct comparison for a few test cases with the newly proposed SPME version which numerical results will always refer to a fixed Ewald's τ parameter ($\tau = 0.544\,590$) corresponding to a real space cutoff of 7 Å.

IV. NUMERICAL RESULTS

We start by considering results related to the simple replica method based on Eq. (13). In particular, for all the results we choose as a measure the first diagonal element of the Δ matrix calculated from the same fixed initial vector $\mathbf{y}_1 = \mathbf{v}$, chosen as usual from a Rademacher distribution. This choice will allow us to eliminate the stochastic noise from the computed Δ_{11} values that otherwise would

11 November 2025 13:46:34

Published under an exclusive license by AIP Publishing

make harder the interpretation and comparison of the effects arising from long-range interactions introduced via both the replica and SPME methods. The first system analyzed is a small cubic box of dimension 18.64 Å containing 216 water molecules in the liquid phase. Figure 2 shows the evolution of Δ_{11} as a function of the cutoff radius R_{cut} that is used to determine the replicas identified by a set of $\{\mathbf{n}\}$ vectors to be included in Eq. (13). Even for a not highly symmetric system such as bulk water, the convergence is reached for a cutoff radius of nearly 30 Å thus confirming the slow (and conditional) convergence rate that characterizes the replica method. The large cutoff radius required by the replica method, because of its consequent quadratic scaling, has a direct impact on the computational time as shown in Fig. 3. In particular, for a 30 Å cutoff the CPU-time required for the computation of the diagonal element chosen as observable reaches 1 s.

The situation if quite different if the SPME-based algorithm is employed since in this case the overall convergence is determined by the number of grid points to be used in the solution of the reciprocal field contribution $[K_1, K_2, K_3$ in Eq. (31)] that also represents the computationally most expensive part of the algorithm as the direct summation part is computed very efficiently in a linear-scaling fashion. Figure 4 shows the convergence of our target quantity Δ_{11} as a function of the number of grid points for the box of water undertaken as test system. We stress that, given the choice that we made to fix Ewald's τ parameter, the only quantity governing the convergence is thus the grid size. We first note that the convergence has a monotonic behavior as a smaller grid size does not involve a physical truncation of the space and thus of the interactions as for the replica case that in fact shows an oscillatory behavior.

It is now interesting to compare the computational cost required by the SPME-based approach to that of the replica method. In particular for a 18 point sized grid for which convergence is observed, the CPU time is 10^{-2}, therefore a factor 100 faster than the cumbersome replica method.

The slow convergence rate observed for the replica method is further exacerbated when highly symmetrical systems are taken under consideration. Figure 5 shows the evolution of Δ_{11} as a function of the cutoff radius, this time for a 14.2 Å sided cubic box of diamond. In this case the cutoff radius reaches the extremely large value of 60 Å before the convergence is reached, with a huge impact on outcoming computational cost as shown in Fig. 6. The system dependency affecting the choice for a proper cutoff radius observed for the replica method is not suffered by the SPME-Lanczos method as it can be seen in Fig. 7 showing the Δ_{11} convergence as a function of the number of grid points. Even in this case, convergence is observed starting from circa 20 points, similarly to that observed for water as both boxes have quite similar size. In fact, convergence is ensured when a certain density of grid points is provided, independently of the system. In general a density of 1.2 points/Å (for each of the three box dimensions) is enough to ensure convergence, and this is the default value chosen in our implementation.

For highly periodic systems for which the replica method is particularly slow to converge, the computational gain provided by the SPME alternative becomes even more marked. For a 18 points grid, the Δ_{11} computation via the SPME-Lanczos is nearly 350 times faster than its replica counterpart.

Although our analysis focused, for the sake of clarity, on the Δ_{11}, the same results hold for the convergence of out of diagonal terms $\Delta_{(k-1)(k)}$. Moreover, we note that the solution of the SPME-Lanczos equations, Eq. (12), does not spoil the orthogonality of the Krylov subspace basis vectors $\mathbf{y}_j^\dagger \mathbf{y}_k = \delta_{jk}$ as the set of vectors remain orthogonal by construction as in the original algorithm. Furthermore, we stress that for an accurate resolution of Eq. (7), the number of quadrature points i.e. the dimension $(M+1)$ of the Krylov

11 November 2025 13:46:34

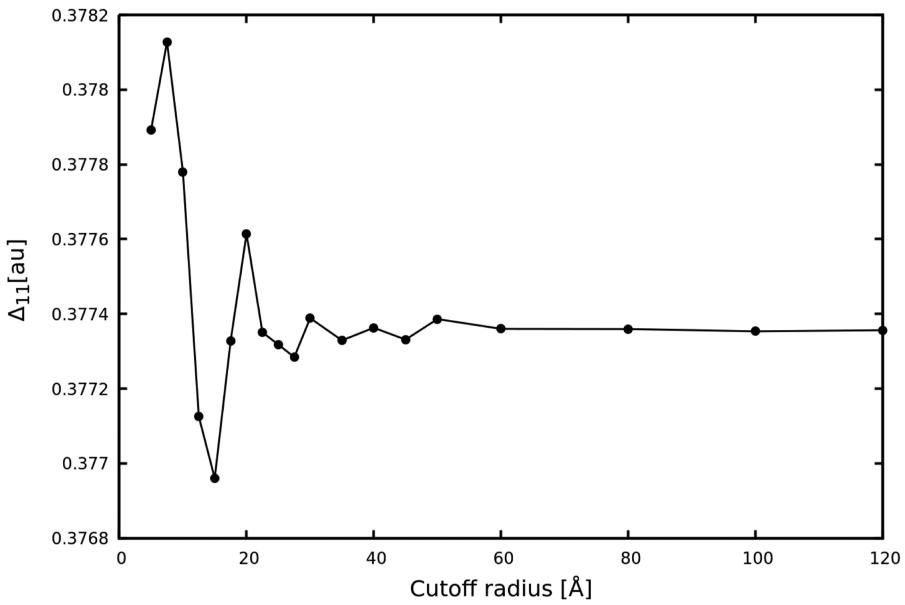


FIG. 5. First diagonal element of Δ as a function of the cutofff radius, computed via the replica method for a cubic box of diamond (14.2 Å).

Published under an exclusive license by AIP Publishing

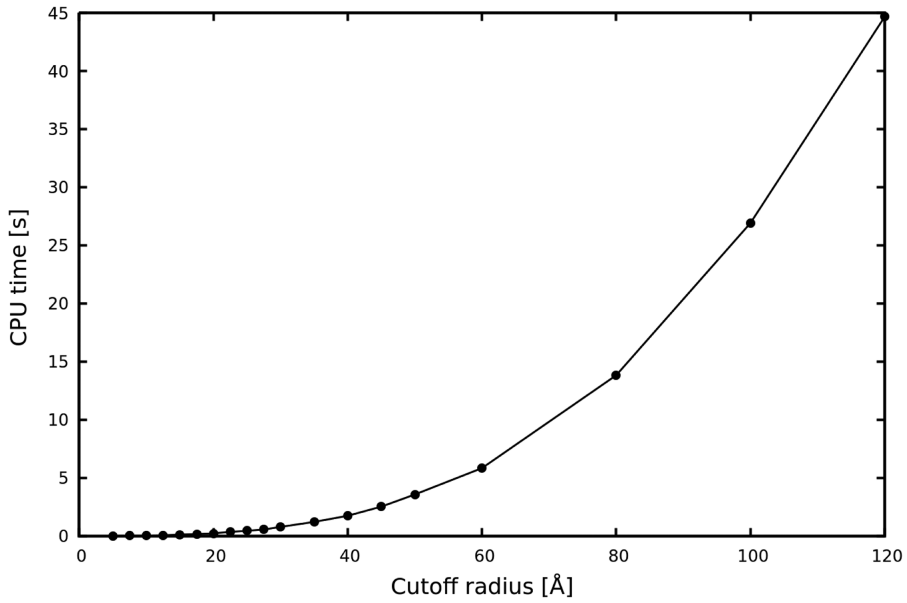


FIG. 6. CPU time as a function of the cutoff radius for the replica method for diamond in a cubic box of side 14.2 Å.

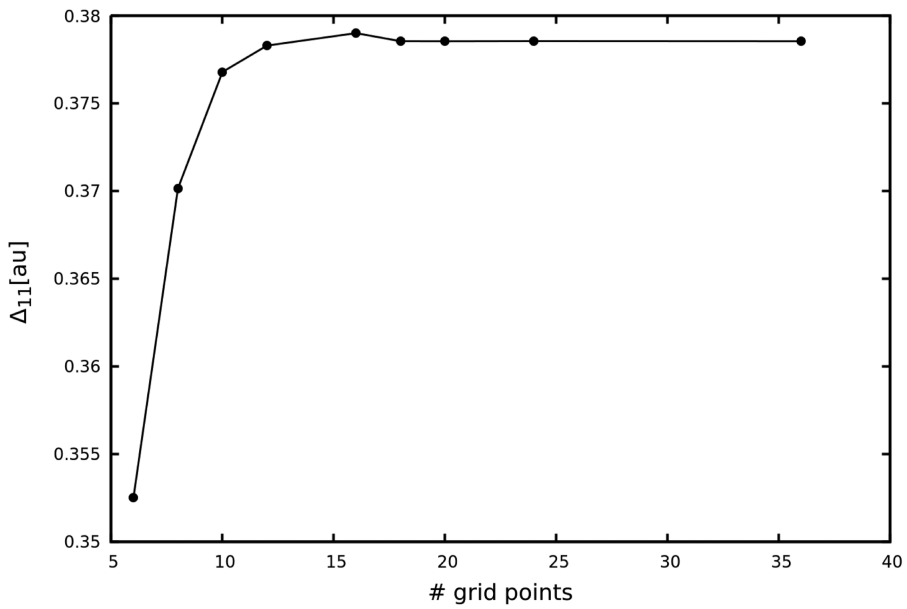


FIG. 7. First diagonal element of Δ computed via the SPME-Lanczos method as a function of the grid points (only K_1 is reported as the box is cubic) for diamond in a box of side 14.2 Å.

subspace $\mathscr{K}_{M+1}$, can be set to 15, regardless of the system size. This implies that the SPME-Lanczos algorithm does not suffer from the numerical instability (loss of orthogonality among basis vectors) of standard Lanczos algorithm,[35] typically encountered in applications where very large Krylov subspaces and thus basis vectors are required.[36]

Being the construction of the tridiagonal matrix Δ the bottleneck step of the overall algorithm, it is of interest to probe its scaling as a function of the system size as shown in Fig. 8 for an increasingly large box of liquid water.

The plot shows that the SPME-SL algorithm deviates from linearity for larger system sizes and this is explained by virtue of the

11 November 2025 13:46:34

Published under an exclusive license by AIP Publishing

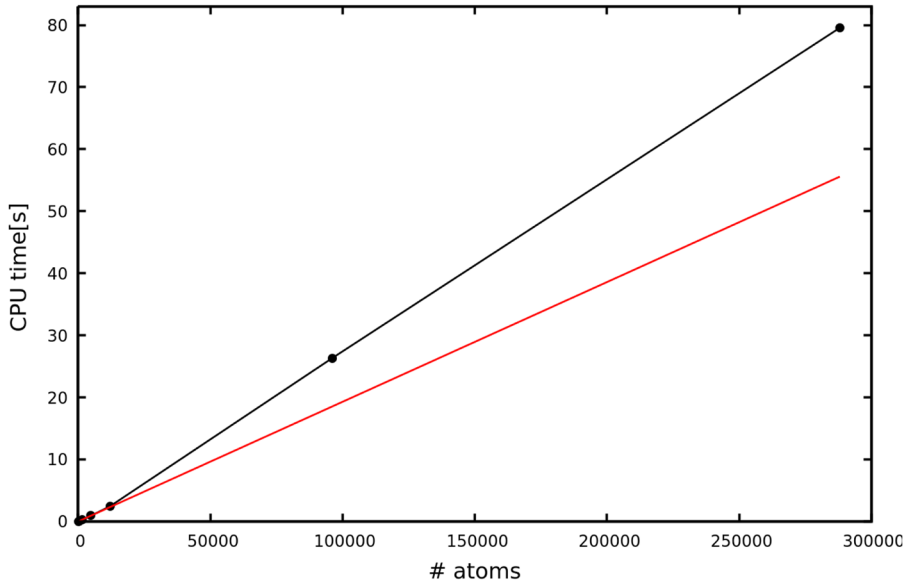


FIG. 8. CPU time in seconds as a function of the number of atoms of increasingly larger water boxes (black line). The time refers to the ten iterations required to build Δ for a given random sample within a standard $\mathscr{K}_{10}$ Krylov subspace with the implemented SPME-SL algorithm. The red line represents the ideal linear scaling.

$N\log(N)$ scaling of the SPME method employed to compute the generalized field vectors which are key contributions in the construction of the $\mathbf{\Delta}$ matrix. The deviation from linearity is, however, rather contained even for the largest system considered composed of ~100 000 water molecules that is completely out of reach for the standard replica method discussed earlier.

For one single core, the overall time necessary to compute the final energy is equivalent to the time required to build $\mathbf{\Delta}$ (Fig. 8) multiplied by the number of random samples R involved in Hutchinson's trace estimator. For large systems in the order of 10^4 atoms or above, R can be taken to be around 300 with a resulting low relative standard deviation (0.5%) as analyzed in depth in Ref. 11.

However, the SPME-SL algorithm's strength is found in its embarrassingly parallel nature since the random samples can be divided among the available processes while a simple reduction is required before the final trace evaluation [Eq. (9)]. Since the parallelization scheme is essentially the same as the one discussed in the original SL-MBD algorithm, we refer to a previous work[11] for an in depth analysis of the scalability with respect to the number of processes as well as a detailed discussion of the parallelization strategy.

V. CONCLUSIONS

We have derived, implemented an discussed the SPME-SL algorithm where the stochastic Lanczos trace estimation scheme is coupled to the state of the art Smooth Particle Mesh Ewald method. This was made possible by introducing the generalized field term contribution from the Lanczos iterative equations. Our combined approach allows for an embarrassingly parallel computation of many-body dispersion energies with the full inclusion of long-range interactions arising from all periodic images of the central simulations cell.

The proposed algorithm undoubtedly outperforms truncation-based approaches such as the replica method that is affected by slow and conditionally convergence as well as by the employed quadratic-scaling double loops making the computation highly inefficient for large systems.

The parallelism features of the SPME-SL algorithm together with the $N\log(N)$ scaling with the system size allows for a fast Many-Body dispersion treatment of very large periodic systems composed of hundreds of thousands atoms and more. This work represents the natural extension to long-range PBC of our recent stochastic Lanczos MBD algorithm[11] and it focuses uniquely on the energy evaluation. Our focus will now be dedicated to the extension of these achievements to the energy nuclear gradients towards large scale condensed phase molecular dynamics simulations including many-body dispersion effects. Furthermore, alternative handling of PBC within the recently introduced ANKH linear scaling framework will pursued.[37]

ACKNOWLEDGMENTS

This work has been funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant No. 810367), project EMC2 (J.P.P.). Computations have been performed at GENCI (IDRIS, Orsay, France and TGCC, Bruyères le Châtel) on Grant No. A0070707671.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

11 November 2025 13:46:34

Published under an exclusive license by AIP Publishing

Author Contributions

Pier Paolo Poier: Conceptualization (lead); Methodology (lead); Software (lead); Writing – original draft (lead); Writing – review & editing (equal). **Louis Lagardère**: Conceptualization (equal); Methodology (equal); Software (equal); Supervision (equal); Writing – original draft (equal); Writing – review & editing (equal). **Jean-Philip Piquemal**: Conceptualization (equal); Funding acquisition (lead); Methodology (equal); Project administration (lead); Resources (lead); Supervision (lead); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1]T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular Electronic Structure Theory* (John Wiley & Sons, Ltd., Chichester, 2000).

[2]F. Jensen, *Introduction to Computational Chemistry*, 3rd ed. (Wiley, Chichester, UK, 2017).

[3]S. Grimme, A. Hansen, J. G. Brandenburg, and C. Bannwarth, "Dispersion-corrected mean-field electronic structure methods," Chem. Rev. **116**, 5105–5154 (2016).

[4]A. G. Donchev, "Many-body effects of dispersion interaction," J. Chem. Phys. **125**, 074713 (2006).

[5]A. Ambrosetti, A. M. Reilly, R. A. DiStasio, and A. Tkatchenko, "Long-range correlation energy calculated from coupled atomic response functions," J. Chem. Phys. **140**, 18A508 (2014).

[6]T. Gould, S. Lebègue, J. G. Ángyán, and T. Bučko, "A fractionally ionic approach to polarizability and van der Waals many-body dispersion calculations," J. Chem. Theory Comput. **12**, 5920–5930 (2016).

[7]K. Carter-Fenk, K. U. Lao, K.-Y. Liu, and J. M. Herbert, "Accurate and efficient *ab initio* calculations for supramolecular complexes: Symmetry-adapted perturbation theory with many-body dispersion," J. Phys. Chem. Lett. **10**, 2706–2714 (2019).

[8]P. P. Poier, T. Jaffrelot Inizan, O. Adjoua, L. Lagardère, and J.-P. Piquemal, "Accurate deep learning-aided density-free strategy for many-body dispersion-corrected density functional theory," J. Phys. Chem. Lett. **13**, 4381–4388 (2022).

[9]P. P. Poier, O. Adjoua, L. Lagardère, and J.-P. Piquemal, "Generalized many-body dispersion correction through random-phase approximation for chemically accurate density functional theory," J. Phys. Chem. Lett. **14**, 1609–1617 (2023).

[10]S. Ubaru, J. Chen, and Y. Saad, "Fast estimation of $tr(f(A))$ via stochastic Lanczos quadrature," SIAM J. Matrix Anal. Appl. **38**, 1075–1099 (2017).

[11]P. P. Poier, L. Lagardère, and J.-P. Piquemal, "$O(N)$ stochastic evaluation of many-body van der Waals energies in large complex systems," J. Chem. Theory Comput. **18**, 1633–1645 (2022).

[12]M. Stöhr and A. Tkatchenko, "Quantum mechanics of proteins in explicit water: The role of plasmon-like solute–solvent interactions," Sci. Adv. **5**, eaax0024 (2019).

[13]T. Bučko, S. Lebègue, T. Gould, and J. G. Ángyán, "Many-body dispersion corrections for periodic systems: An efficient reciprocal space implementation," J. Phys.: Condens. Matter **28**, 045201 (2016).

[14]T. Darden, D. York, and L. Pedersen, "Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems," J. Chem. Phys. **98**, 10089–10092 (1993).

[15]U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, "A smooth particle mesh Ewald method," J. Chem. Phys. **103**, 8577–8593 (1995).

[16]T. J. Giese, M. T. Panteva, H. Chen, and D. M. York, "Multipolar Ewald methods, 1: Theory, accuracy, and performance," J. Chem. Theory Comput. **11**, 436–450 (2015).

[17]T. Brinck, J. S. Murray, and P. Politzer, "Polarizability and volume," J. Chem. Phys. **98**, 4305–4306 (1993).

[18]A. D. Becke and E. R. Johnson, "Exchange-hole dipole moment and the dispersion interaction: High-order dispersion coefficients," J. Chem. Phys. **124**, 014104 (2006).

[19]A. Tkatchenko and M. Scheffler, "Accurate molecular van der Waals interactions from ground-state electron density and free-atom reference data," Phys. Rev. Lett. **102**, 073005 (2009).

[20]K. R. Bryenton and E. R. Johnson, "Many-body dispersion in model systems and the sensitivity of self-consistent screening," J. Chem. Phys. **158**, 204110 (2023).

[21]D. Massa, A. Ambrosetti, and P. L. Silvestrelli, "Many-body van der Waals interactions beyond the dipole approximation," J. Chem. Phys. **154**, 224115 (2021).

[22]M. Hutchinson, "A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines," Commun. Stat. Simul. Comput. **19**, 433–450 (1990).

[23]G. H. Golub and G. Meurant, *Matrices, Moments and Quadrature with Applications* (Princeton University Press, 2009).

[24]M. A. Blood-Forsythe, T. Markovich, R. A. DiStasio, R. Car, and A. Aspuru-Guzik, "Analytical nuclear gradients for the range-separated many-body dispersion model of noncovalent interactions," Chem. Sci. **7**, 1712–1728 (2016).

[25]S. W. de Leeuw, J. W. Ewald, and E. R. Smith, "Simulation of electrostatic systems in periodic boundary conditions. I. Lattice sums and dielectric constants," Proc. R. Soc. London, Ser. A **373**, 27–56 (1980).

[26]P. P. Poier and J. Frank, "Including implicit solvation in the bond capacity polarization model," J. Chem. Phys. **151**, 114118 (2019).

[27]P. P. Poier and F. Jensen, "Polarizable charges in a generalized Born reaction potential," J. Chem. Phys. **153**, 024111 (2020).

[28]B. Stamm, L. Lagardère, E. Polack, Y. Maday, and J.-P. Piquemal, "A coherent derivation of the Ewald summation for arbitrary orders of multipoles: The self-terms," J. Chem. Phys. **149**, 124103 (2018).

[29]A. Redlack and J. Grindlay, "Coulombic potential lattice sums," J. Phys. Chem. Solids **36**, 73–82 (1975).

[30]T. R. C. Kenny, B. Lipkowitz, and D. B. Boyd, *Reviews in Computational Chemistry* (Wiley, 2007), Vol. 24.

[31]M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon Press, New York, 1989).

[32]C. Sagui, L. G. Pedersen, and T. A. Darden, "Towards an accurate representation of electrostatics in classical force fields: Efficient implementation of multipolar interactions in biomolecular simulations," J. Chem. Phys. **120**, 73–87 (2004).

[33]L. Lagardère, L.-H. Jolly, F. Lipparini, F. Aviat, B. Stamm, Z. F. Jing, M. Harger, H. Torabifard, G. A. Cisneros, M. J. Schnieders, N. Gresh, Y. Maday, P. Y. Ren, J. W. Ponder, and J.-P. Piquemal, "Tinker-HP: A massively parallel molecular dynamics package for multiscale simulations of large complex systems with advanced point dipole polarizable force fields," Chem. Sci. **9**, 956–972 (2018).

[34]O. Adjoua, L. Lagardère, L.-H. Jolly, A. Durocher, T. Very, I. Dupays, Z. Wang, T. J. Inizan, F. Célerse, P. Ren, J. W. Ponder, and J. P. Piquemal, "Tinker-HP: Accelerating molecular dynamics simulations of large complex systems with advanced point dipole polarizable force fields using GPUs and multi-GPU systems," J. Chem. Theory Comput. **17**, 2034–2053 (2021).

[35]C. Lanczos, "An iteration method for the solution of the eigenvalue problem of linear differential and integral operators," J. Res. Natl. Bur. Stand. **45**, 255–282 (1950).

[36]C. Brezinski, M. Zaglia, and H. Sadok, "A review of formal orthogonality in Lanczos-based methods," J. Comput. Appl. Math. **140**, 81–98 (2002), Int. Congress on Computational and Applied Mathematics 2000.

[37]I. Chollet, L. Lagardère, and J.-P. Piquemal, "ANKH: A generalized $O(N)$ interpolated Ewald strategy for molecular dynamics simulations," J. Chem. Theory Comput. **19**, 2887–2905 (2023).

11 November 2025 13:46:34

Published under an exclusive license by AIP Publishing