

UC Irvine

UC Irvine Previously Published Works

Title

Optimal Scheme to Achieve Energy Conservation in Induced Dipole Models

Permalink

<https://escholarship.org/uc/item/6pw5c644>

Journal

Journal of Chemical Theory and Computation, 19(15)

ISSN

1549-9618

Authors

Huang, Zhen
Zhao, Shiji
Cieplak, Piotr
[et al.](#)

Publication Date

2023-08-08

DOI

10.1021/acs.jctc.3c00226

Peer reviewed

An Optimal Scheme to Achieve Energy Conservation in Induced Dipole Models

Zhen Huang,^{#,†} Shiji Zhao,^{†,§} Piotr Cieplak,[‡] Yong Duan,^{*,§} Ray Luo,^{*,#,†} and Haixin Wei^{*,†,%}

[#]Chemical and Materials Physics Graduate Program, University of California, Irvine. Irvine, California 92697, United States

[†]Departments of Molecular Biology and Biochemistry, Chemical and Biomolecular Engineering, Materials Science and Engineering, and Biomedical Engineering, University of California, Irvine. Irvine, California 92697, United States

[§]Nurix Therapeutics, Inc., 1700 Owens St. Suite 205, San Francisco, CA 94158, United States

[‡]SBP Medical Discovery Institute, 10901 North Torrey Pines Road, La Jolla, California 92037, United States

[§]UC Davis Genome Center and Department of Biomedical Engineering, University of California, Davis, One Shields Avenue, Davis, California 95616, United States

[%]Department of Chemistry & Biochemistry, University of California, San Diego. 9500 Gilman Drive, La Jolla, California 92093, United States

*Corresponding author:

Yong Duan: duan@ucdavis.edu

Ray Luo: rluo@uci.edu

Haixin Wei: haixinw1@uci.edu

Abstract

Induced dipole models have proven to be effective tools for simulating electronic polarization effects in biochemical processes, yet their potential has been constrained by issues of energy conservation, particularly when historical data is utilized for dipole prediction. This study identifies error outliers as the primary factor causing this failure of energy conservation and proposes a comprehensive scheme to overcome this limitation. Leveraging maximum relative errors as a convergence metric, our data demonstrates that energy conservation can be upheld even when using historical information for dipole predictions. Our study introduces the Multi-Order Extrapolation (MOE) method to quicken induction iteration and optimize the use of historical data, while also developing the Preconditioned Conjugate Gradient with Local Iterations (LIPCG) to refine the iteration process and effectively remove error outliers. This scheme further incorporates a "peek" step via Jacobi Under-Relaxation (JUR) for optimal performance. Simulation evidence

suggests that our proposed scheme can achieve energy convergence akin to that of point-charge models within a limited number of iterations, thus promising significant improvements in efficiency and accuracy.

Introduction

Computer simulations are increasingly vital for elucidating molecular mechanisms at the microscopic level. To achieve dependable molecular dynamics (MD) simulations, force fields of high precision and consistent atomistic interaction descriptions are paramount. Polarizable force fields, encapsulating interactions between charges and induced higher moments like dipoles and quadrupoles, have shown superiority in depicting atomistic interactions over previous fixed-charge additive force field models, per comparisons with high-level quantum mechanical calculations. Strategies for polarizable force field development are varied, involving Drude oscillators^{1, 2}, fluctuating charges³, induced dipoles⁴⁻⁶, and continuum dielectrics^{7, 8}. Recently, the polarizable Gaussian Multipole (pGM) model⁹⁻¹⁴ has emerged as a promising alternative approach, employing Gaussian-shaped multipoles and dipoles to consistently treat intra- and intermolecular electrostatic interactions and circumvent "polarization catastrophe."^{15, 16}

The pGM model has been made available to the molecular modeling community through a series of recent papers. Specifically, a set of isotropic atomic polarizabilities and radii for the pGM model has been optimized at the B3LYP/aug-cc-pVTZ level of theory.⁹ A local reference frame with unit vectors defined along covalent or virtual bonds has been suggested, enabling closed-form analytical expressions to be obtained for atomic forces.¹⁰ For molecular simulations under the periodic boundary condition, the pGM model's electrostatic terms have been interfaced with the particle mesh Ewald (PME) approach.^{10, 17-20} The pGM internal stress tensor expression for constant pressure MD simulations of both flexible and rigid body molecular systems has also been derived.¹¹ Additionally, a python-based PyRESP program has been implemented, which enables parameterizations for various induced dipole polarizable models by reproducing the quantum mechanical (QM) electrostatic potential (ESP) around molecules.¹² Finally, the pGM model has been shown to perform with high accuracy in predicting many-body interactions in peptide oligomers and to exhibit good transferability in predicting ESPs of molecules and oligomers.^{13, 14}

In comparison to traditional point-charge models, MD simulations with polarizable induced-dipole models bring considerably higher computational costs due to the need for computing induced dipoles and electric fields. This study explores strategies for efficient induction calculations within the pGM model framework, with findings applicable to other induced dipole polarizable force fields.

A key step in the polarizable induced-dipole simulations is to compute the induced dipoles to describe the electrostatic induction. Like in other induced dipole models, induced dipoles in pGM can be obtained by solving the following equations as

$$p_i = \alpha_i \left(E_i - \sum_{j \neq i}^N T_{ij} \cdot p_j \right) \quad (1)$$

Iterative methods in principle converge towards the true solution by reducing the residue, which can be viewed as minimization of a residual function $f(P) = \frac{1}{2} P^T \hat{T} P - P^T E$.²¹

When residue $E - \hat{T} P = 0$, the residue function reaches minimum. One of the primary approaches to minimize the function value is the gradient descent algorithm. At each iteration, the optimal direction and the amount of move are found. We denote the search direction to be $e^{(i)}$ for the i^{th} iteration and $a_{(i)}$ to be the amount of move along the search direction. The new iterative solution is then a linear combination of all previous search directions. $P_{i+1} = a_{(0)} e^{(0)} + a_{(1)} e^{(1)} + a_{(2)} e^{(2)} + \dots$.^{21, 22}

The conjugate gradient (CG) method is instrumental in reducing search iterations by utilizing search directions of the initial gradient and vectors conjugate to that gradient.^{21, 22} In most cases, a preconditioner matrix that mirrors the original matrix yet is easy to invert can expedite the iteration. This method is widely employed in biomolecular applications of Poisson-Boltzmann solvent models²⁴⁻²⁷ and is also used in induction iteration, as reviewed by Wang and Skeel, Lipparini et al, and Aviat et al.^{23, 28, 29} In this work, we introduce a local iteration-based preconditioner instead of the preconditioner presented by Wang and Skeel. This preconditioner more effectively captures the optimal search direction for updating induced dipoles as it is a better approximation to the inversed matrix.

Moreover, another efficiency-improving technique can be applied in the CG method: an extra step—known as the "peek" step—in the form of a JUR iteration can be used to accelerate convergence.²⁸ This strategy involves an additional attempted movement along the current residual direction (available during convergence evaluation). If this movement converges with the existing iteration solution, the peek step saves computation time by eliminating the need to calculate a new CG search direction. If it does not converge, the process proceeds to the next CG iteration.²⁸ In this work, we explored an alternate strategy of applying a peek step after convergence, which, as demonstrated by Aviat et al., can effectively enhance the quality of induced dipoles.²³

Aside from efficient iterative algorithms, the quality of the initial guess substantially impacts the convergence steps required. Toukmaji et al proposed using the external electric field due to permanent moments and atomic polarizability (αE_{perm}) instead of zero as the initial guess.³⁰ In the context of MD simulations, which represent continuous system time evolution, induced dipoles from prior steps present viable initial guesses. A more contemporary proposal by Wang and Skeel involves an extrapolation scheme utilizing induced dipoles from numerous previous steps, thereby enhancing convergence efficiency. However, it has been observed that NVE simulations using prior steps' induced dipoles as initial guesses were found to lead to subpar energy conservation compared to simulations utilizing αE_{perm} as initial guesses, even when employing the same convergence criterion. This phenomenon, while well-known, remains unexplained.²⁸ Given that energy conservation is a crucial validation step for any force field development for MD simulations, addressing this issue is essential. In this context, Aviat et al. put forth a truncated conjugate gradient method involving a fixed number of iterations to achieve an approximated solution, established at a user-defined accuracy level.²³ This approximated solution is treated as the "correct" solution, ensuring consistency in energy and forces with the approximation, thereby ostensibly achieving

energy conservation in NVE simulations.²³ In this study, we delve deeper into understanding this energy drift and propose an enhanced multi-order extrapolation method.

One last point worth mentioning is the convergence check in iterative solutions of linear systems. As solutions of all linear systems in the form of $Ax=b$ would lead to residues ($r=b-Ax$) approaching zero as close as possible, a standard practice is to claim convergence once the relative norm is less than a preset error tolerance (τ), i.e. $\|r\|/\|b\| \leq \tau$.^{21, 22} Another way is to check whether the difference between two consecutive solutions is smaller than the tolerance. Either average unsigned difference or root mean squared difference can be used.²⁸ A risk is that the iterative procedure may be stuck somewhere leading to very little change in the solution, resulting in a false convergence claim. This occurs often in solutions of the Poisson-Boltzmann equation.²⁴⁻²⁷

In this study, we strive to enhance the induction calculation efficiency of the pGM model, balancing optimal efficiency and energy conservation in molecular dynamics. We delve into the reasons behind poor energy conservation behavior when using history in initial guesses and propose the multi-order extrapolation (MOE) scheme to maximize the use of previous steps' induced dipoles in initial guesses. We also developed the preconditioned conjugate gradient method with a local-iteration-based preconditioner (LIPCG), enhancing convergence rates. Finally, we optimize the "peek" step in the form of Jacobi under-relaxation (JUR) to better the convergence quality and efficiency.

Method

1. Residue Norm is a Poor Measure in Convergence Check

In the development of polarizable models, it is common to use induced dipoles from previous steps to initialize the induced dipoles in the current step, as this can speed up convergence. However, it has been observed that using history in the initial guesses can make it more difficult to achieve energy conservation, compared to simulations without history unless a more stringent convergence criterion is used.^{23, 28} In the end the time saving would be reduced in induction calculation. This issue also exists in pGM simulations, as we show in the Results and Discussion. We believe that this problem arises because the system does not truly reach the intended convergence quality whether history is used or not, when convergence is determined by residue norm.

To understand this issue, let us first consider the case when no history is used for initial guesses. In this case, all induced dipoles start in a similar situation with respect to convergence. As the induction iteration progresses, all induced dipoles converge in a similar manner, with individual dipole errors being reduced similarly. Thus, the residue norm can be used to check convergence reasonably well.

However, when history is used, the induced dipoles start in different situations. On one hand, many initial guesses are so good that they converge quickly with few iterations. On the other hand, there are dipoles that are far from convergence and require many iterations to converge. Due to the larger range of initial errors, the overall residue norm is not an appropriate measure for checking convergence because the ratio, named residue error: $\varepsilon = \|r\|/\|b\|$ would mask the large errors, which accumulate and cause significant energy drifts.

Therefore, it is better to use the maximum relative error in convergence checking whether history is used or not. (In the following, errors are all relative by default, so we drop “relative” for presentation clarity). This approach makes it possible to identify the slowest-converging induced dipole and ensure that convergence is achieved for all induced dipoles in the system. By using this approach, energy conservation can finally be achieved as we shown in Results and Discussion.

2. Multi-Order Extrapolation (MOE) Scheme for Initial Guesses

Given the maximum error as the measure for checking convergence, the energy drift problem would not be an issue as shown in Results and Discussion. Thus, we can safely use history in initial guesses to accelerate the induction calculation.

Since in molecular simulations atoms do not move much at each step, we can assume that induced dipoles change in a linear fashion for a few simulation steps. In general, the following linear extrapolation can be used to approximate the induced dipoles,²⁸

$$P_n \approx \sum_{i=1}^k P_{n-i} \cdot \lambda_i \quad (9)$$

where P_n is a $3N$ -dimensional vector representing all induced dipoles of the system at the current step n . λ_i is the extrapolation coefficient, and k is the number of previous steps that are used. Obviously, the above equations are over-determined and can be solved by a least square fitting approach,

$$P_{n-j}^T \cdot P_n = \sum_{i=1}^k P_{n-j}^T \cdot P_{n-i} \cdot \lambda_i \quad (10)$$

where j takes values from 1 through k .

However, since P_n is not known until the induction calculation is finished, the above extrapolation scheme is not feasible. One way to address this problem, as described in Ref²⁸, is to solve the above equation for P_{n-1} and reuse those obtained λ values to predict P_n .

Here, we introduce an alternative method, which is to solve the least square fitting equation with E_n , the permanent electric field (due to permanent multipoles) instead of P_n at the current step n .

$$E_{n-j}^T \cdot E_n = \sum_{i=1}^k E_{n-j}^T \cdot E_{n-i} \cdot \lambda_i \quad (11)$$

Next the λ 's are used to extrapolate P_n . The advantage of the scheme is that we are using the most up-to-date information from the current step to obtain the coefficients.

Our extrapolation scheme is based on the assumption that $\hat{T}^{-1} = (\alpha^{-1} + T)^{-1}$ varies very little during simulations as noted before.²⁸ Thus, we have

$$P_n = (\alpha^{-1} + T_n)^{-1} E_n \approx (\alpha^{-1} + T)^{-1} E_n \approx (\alpha^{-1} + T)^{-1} \sum_{i=1}^k E_{n-i} \cdot \lambda_i \approx \sum_{i=1}^k (\alpha^{-1} + T_{n-i})^{-1} E_{n-i} \cdot \lambda_i \approx \sum_{i=1}^k P_{n-i} \cdot \lambda_i \quad (12)$$

In addition, the scheme can be extended to the second-order extrapolation since the difference between the true and extrapolated P_n values also satisfy the same relation,

$$P_n^{real} - P_n^{extrapolated} \approx \sum_{i=1}^k (P_{n-1}^{real} - P_{n-1}^{extrapolated}) \cdot \lambda_i \quad (13)$$

Similarly, the third-order extrapolation can also be applied to extrapolate the "difference of the difference".

Thus, this scheme is termed the multi-order extrapolation (MOE) scheme.

3. Local Iterations for Accelerating Maximum Error Convergence

In order to accelerate the convergence of induced dipoles with poor initial guesses, which significantly impact the overall convergence rate, additional methods are required even when using the maximum error as the convergence measure and an aggressive MOE scheme to take advantage of history as much as possible.

Our approach is to use "local" iterations after each "global" iteration, which can be demonstrated most clearly with a stationary point method. Eq. (3) shows that a stationary point method can be used to solve the linear system. However, if this method is used to conduct iterations to solve the system, dipoles with very poor initial guesses will converge much more slowly than most other dipoles. If the maximum error is used to verify convergence, the majority of the iterations will be wasted attempting to improve

the dipoles with the poorest initial guesses. To address the limitation of the general iterations, we rewrite Eq. (3) as

$$P = a[E - (T + T_{local} - T_{local}) \cdot P] \quad (14)$$

Here

$$(T_{ij})_{local} = \partial_i^a \partial_j^b \frac{\text{erf}(\beta_{ij} R_{ij})}{R_{ij}}, \text{ if } R_{ij} \leq d; \wedge (T_{ij})_{local} = 0, \text{ if } R_{ij} > d. \quad (15)$$

Cutoff distance “ d ” can be chosen for optimal efficiency of the overall convergence, usually 3~4 \AA . Eq. (15) above can be rewritten as:

$$(a^{-1} + T_{local})P = E - (T - T_{local}) \cdot P \quad (16)$$

Thus, we have

$$P = (a^{-1} + T_{local})^{-1} [E - (T - T_{local}) \cdot P] \quad (17)$$

Since the local iterations only consider very short-range dipole-dipole interactions, it is a highly efficient way to account for the fast-changing local interactions that cause the poor initial guesses.

Eq. (17) leads to a different stationary point method, which requires two different matrix-vector operations at each iteration to solve the system, i.e.

$$P_{n+1} = (a^{-1} + T_{local})^{-1} [E - (T - T_{local}) \cdot P_n] \quad (18)$$

The first one is $(T - T_{local}) \cdot P_n$, similar to any stationary point method. Because this iteration requires the calculation of $T \cdot P_n$, which represents interactions between all pairs of induced dipoles in the system, this step can be called the global iteration. The second one involves the inversion of $a^{-1} + T_{local}$. This can be done iteratively by solving the following equation,

$$(a^{-1} + T_{local}) \cdot P_{n+1} = b \quad (19)$$

where $b = E - (T - T_{local}) \cdot P_n$. Because this iteration does not require the calculation of $T \cdot P_{n+1}$ but only a local truncated $T_{local} \cdot P_{n+1}$, this step is called the local iteration. Using Eq. (17) instead of Eq. (3) adds a local iteration on top of a global iteration, which further relaxes the system in a more efficient manner.

It is worth noting that solution of Eq. (19) can be achieved by a preconditioned CG algorithm, where the diagonal term of the polarization tensor (a) can be used to speed up convergence of the local iteration.

$$r_0 = b - (a^{-1} + T_{local}) \cdot x_0$$

$$z_0 = ar_0$$

$$p_0 = z_0$$

do

$$\alpha_k = \frac{r_k^T \cdot z_k}{p_k^T \cdot (a^{-1} + T_{local}) \cdot p_k}$$

$$x_{k+1} = x_k + \alpha_k p_k$$

$$r_{k+1} = r_k - \alpha_k (a^{-1} + T_{local}) \cdot p_k$$

if $\|r_{k+1}\| / \|b\| \leq \tau$, **break**

$$z_{k+1} = ar_{k+1}$$

$$\beta_k = \frac{r_{k+1}^T \cdot z_{k+1}}{r_k^T \cdot z_k}$$

$$p_{k+1} = r_{k+1} + \beta_k p_k$$

end do

4. Implementation of Local Iterations as Preconditioner for Conjugate Gradient (LIPCG)

As mentioned in Introduction, conjugate gradient (CG) methods fall into another category of linear system solvers, e.g., the Krylov method.²¹ It is often believed and practically true that CG methods are much faster than stationary point methods for induction iterations.^{23, 28} In this section, local iterations will be incorporated into the CG algorithm within the preconditioning framework.

As reviewed in the introduction, the preconditioned conjugate gradient (PCG) method leverages a preconditioner to accelerate convergence. For simpler cases, the preconditioner is often selected as the diagonal elements of matrix A . However, for larger and more complex systems, the preconditioner can be an approximation of A . One such example of a preconditioner was proposed by Wang and Skeel, who used a first-order approximation: $\hat{T}^{-1} \approx (\alpha^{-1} + T_{local})^{-1} \approx \alpha - \alpha^T T_{local} \alpha$.²⁸ The implementation of this preconditioned conjugate gradient method provides a robust and swift path to convergence. The standard operational scheme of a PCG method with a preconditioning matrix (M) is detailed below.

$$r_0 = b - A \cdot x_0$$

$$z_0 = M^{-1} r_0$$

$$p_0 = z_0$$

do

$$\alpha_k = \frac{r_k^T \cdot z_k}{p_k^T \cdot A \cdot p_k}$$

$$x_{k+1} = x_k + \alpha_k p_k$$

$$r_{k+1} = r_k - \alpha_k A \cdot p_k$$

if $\|r_{k+1}\| / \|b\| \leq \tau$, **break**

$$z_{k+1} = M^{-1} r_{k+1}$$

$$\beta_k = \frac{r_{k+1}^T \cdot z_{k+1}}{r_k^T \cdot z_k}$$

$$p_{k+1} = r_{k+1} + \beta_k p_k$$

end do

In this work, we extended the potential utility of preconditioners. We not only investigated the first-order approximation of \hat{T}^{-1} by Wang and Skeel,²⁸ but also utilized the local iterations previously mentioned to enhance the approximation of \hat{T}^{-1} . To understand how to incorporate this in CG, we introduce a quantity, pseudo-solution x'_k . For each "real" solution, x_k ,

$$x'_k = (\alpha^{-1} + T_{local})^{-1} [b - (T - T_{local}) \cdot x_k] \quad (20)$$

From the Eq. (17) of the last section, we know that this is, in fact, the result of a local iteration.

Thus, instead of generating r_k as $b - (\alpha^{-1} + T) \cdot x_k$, we use $x'_k - x_k$ as the new residue, r'_k ,

$$\begin{aligned}
r'_k &= x'_k - x_k = (\alpha^{-1} + T_{local})^{-1} [b - (T - T_{local}) \cdot x_k] - x_k \\
\hat{r}_k &= (\alpha^{-1} + T_{local})^{-1} [b - (T - T_{local}) \cdot x_k] - (\alpha^{-1} + T_{local})^{-1} (\alpha^{-1} + T_{local}) \cdot x_k \\
&= (\alpha^{-1} + T_{local})^{-1} [b - (\alpha^{-1} + T) \cdot x_k] = (\alpha^{-1} + T_{local})^{-1} r_k
\end{aligned} \tag{21}$$

Therefore, the new residue is the standard CG residue with a constant transformation $(\alpha^{-1} + T_{local})^{-1}$. In another word, this extra step of local iteration effectively plays the role of a preconditioner, M , i.e.

$$M^{-1} = (\alpha^{-1} + T_{local})^{-1} \tag{22}$$

In summary, we intend to compute $z_0 = M^{-1} r_0$ by calling the local-iteration CG as shown at the end of last section. Of course, this process can be expensive if we insist for high accuracy. However, this is not necessary as we do not rely on the precise search direction derived from the preconditioning step. In our testing, a few steps of iterations (2 to 3) were found to be sufficient for the local-iteration PCG method to improve convergence and reduce the overall CPU time notably.

$$r_0 = b - A \cdot x_0$$

$$\begin{aligned}
z_0 &= local_{iter}(r_0) & p_0 &= z_0 \\
\mathbf{do} & & & \\
\alpha_k &= \frac{r_k^T \cdot z_k}{p_k^T \cdot A \cdot p_k} & & \\
& & x_{k+1} &= x_k + \alpha_k p_k \\
& & r_{k+1} &= r_k - \alpha_k A \cdot p_k \\
\mathbf{if} \quad \|r_{k+1}\| / \|b\| &\leq \tau, \mathbf{break} & & \\
& & z_{k+1} &= local_{iter}(r_{k+1}) \\
& & \beta_k &= \frac{r_{k+1}^T \cdot z_{k+1}}{r_k^T \cdot z_k} \\
& & p_{k+1} &= r_{k+1} + \beta_k p_k \\
\mathbf{end do} & & &
\end{aligned}$$

5. Molecular dynamics simulation for performance tests

In all conducted tests, we employed a cubic box containing 512 water molecules, identical to the one used in our prior research.¹⁰ The box dimensions were $(33 \text{ \AA})^3$, and all simulations were performed under the NVE conditions. To meet the high precision necessary for energy computations, we utilized the following parameters for the PME setup: a $\beta_0 = 0.4 \text{ \AA}^{-1}$ setting for the Ewald coefficient, a B-spline interpolation order of 8, a Fast Fourier Transform (FFT) grid spacing size of 0.66 \AA , and a direct space cutoff distance of 9.0 \AA . These parameters yield a PME accuracy level of 10^{-6} , which ensures that PME is not the primary contributor to energy drift. The timestep of 1 femtosecond was used to guarantee a good quality of energy conservation. The Jacobi-under relaxation scheme is applied for iterative approach as mentioned in the appendix. The maximum number of PCG iterations was limited to 50, and the under-relaxation parameter was set to 0.65. For local iterations, a cutoff distance of 4.0 \AA and three

iterations were established. The other parameters followed the default values provided in the Amber simulation package.³¹

Results and Discussion

1. Detailed Error Analysis of Induction Iteration

A vital aspect of implementing a new force field in a molecular dynamics program is validation of energy conservation. As depicted in Table 1, energy drifts are generally more substantial when using the previous step's induced dipoles as the initial guess (μ_{n-1}). In all examined instances, energy drifts were less than half when history was disregarded (αE_{perm}) as compared to when history was considered (μ_{n-1}). This discrepancy between the two methods suggests a difference in convergence quality, with the scheme disregarding history achieving better convergence despite both methods employing the same tolerance. This observation aligns with Wang and Skeels' experiments, where employing history information increased the likelihood of energy conservation failure.²⁸ This detrimental effect is likely to be more significant in pure water due to its high flexibility. However, since water molecules predominantly mimic the biomolecular environment in biomolecular simulations, addressing this issue is essential for molecular dynamics applications. This observation led us to hypothesize that the use of history information might result in more unconverged induced dipoles, as we elaborated in the Methods section. As delineated in Table S1, the average iteration count and CPU time for a simulation comprising 1000 steps are documented. It is discernible from the table that, when maintaining identical convergence criteria, the employment of historical data facilitates a reduction of approximately two iterations before convergence is achieved, thereby yielding substantial time savings. If we consider the time component, an equal number of iterations correlates with a similar time duration. This implies that utilizing historical data at certain convergence criteria will result in a simulation time comparable to a scenario where historical data is not utilized, but with looser convergence criteria. Furthermore, this observation underscores that the utilization of the standard norm as a convergence check is not unassailable.

In this section, we first discuss why history information forms a valuable component of the initial guess. Figure 1 illustrates the error distribution of induced dipoles on 512 water molecules before iteration, with and without history information. As depicted in the figure, employing the permanent field as an initial guess delivers relatively larger errors. Using the previous step as an initial guess, however, can expedite convergence by providing a more accurate starting point. Interestingly, despite smaller initial errors when using history, the quality of convergence tends to be worse compared to when history is not used. This conclusion is supported by Figure 2, where the error distribution post-convergence is depicted. Tests incorporating history information yielded larger errors than those disregarding history, signifying that the use of history leads to relatively poor convergence quality. As mentioned in the Methods section, when history information is employed, some large errors converge more slowly compared to quickly converging errors, potentially contributing to energy drift in induced dipole models.

We also aim to shed light on the impact of error outliers on the issue of energy conservation failure. A detailed analysis of the error outliers' distribution is presented in

Figure 3. Here, we define the outliers when the error exceeds 1000 times the current tolerance τ . As indicated in the figure, the use of history information consistently resulted in a higher number of error outliers compared to those with αE_{perm} . Moreover, a stricter convergence tolerance diminishes the difference between error distributions, suggesting that energy drift should be similar for μ_{n-1} and αE_{perm} given a sufficiently stringent convergence tolerance. To achieve convergence with high tolerance, we propose the maximum error as a potential candidate for convergence checking.

Table 1: Comparative Analysis of Energy Drift (expressed in kcal/(mol·ps)) and Convergence Tolerance (τ) Utilizing Residue Norm as a Convergence Measure. The number of degrees of freedom in the system is 3069. The table exhibits that energy drifts in simulations where the initial values stem from the last step (μ_{n-1}) consistently exceed those where the initial values are drawn from the permanent field (αE_{perm}). Refer to Figures S1, S2, and S3 for more detailed energy plots pertaining to these simulations. A comprehensive timing analysis for these simulations can be found in Table S1.

| τ | 10^{-4} | 10^{-5} | 10^{-6} | 10^{-7} |
|--------------------|------------------------|------------------------|------------------------|-----------------------|
| μ_{n-1} | -8.25×10^{-2} | -1.76×10^{-3} | -1.11×10^{-4} | 6.59×10^{-4} |
| αE_{perm} | 1.92×10^{-3} | -5.33×10^{-4} | -3.47×10^{-5} | 1.98×10^{-5} |
| TIP3P energy drift | -5.56×10^{-4} | | | |

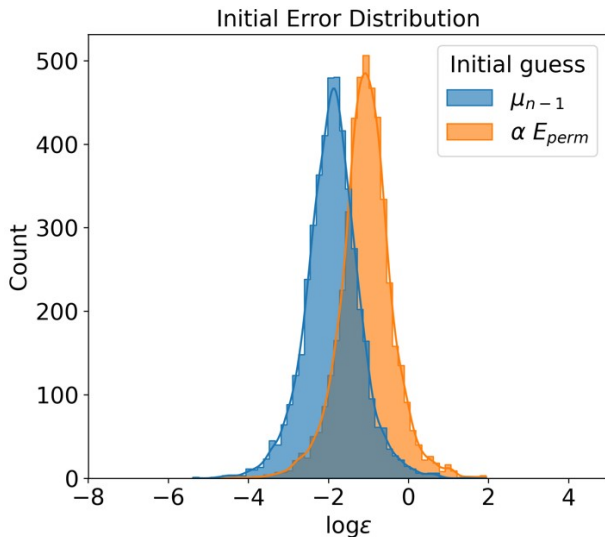


Figure 1: Initial Error Distribution Comparison for μ_{n-1} and αE_{perm} as Initial Guesses. This figure showcases the difference in initial error distribution when using either the last step (μ_{n-1}) or the permanent field (αE_{perm}) as the initial guess. It clearly indicates that using historical data (μ_{n-1}) as an initial guess proves to be a significantly better option than not utilizing historical data. The convergence tolerance for this comparison is set at 10^{-4} .

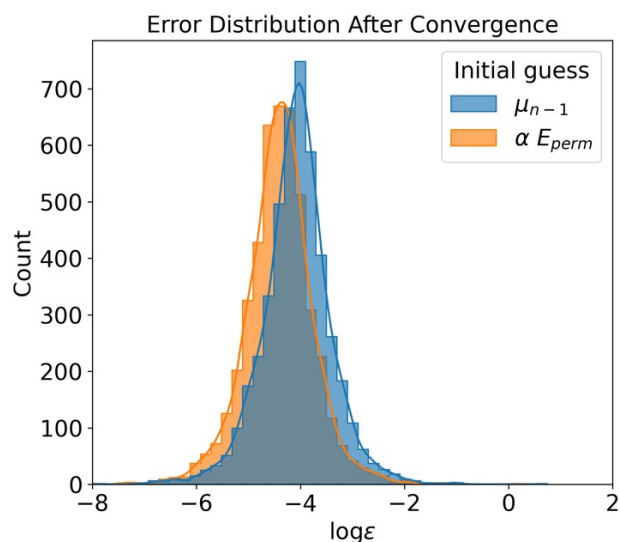


Figure 2: Post-convergence Error Distribution Comparison with and without the Use of Historical Data. This figure presents the error distribution after achieving convergence when employing historical data compared to not using it. It illustrates that the use of historical data results in larger errors compared to using the permanent field (αE_{perm}) as the initial guess. The convergence tolerance for this analysis is set at 10^{-4} .

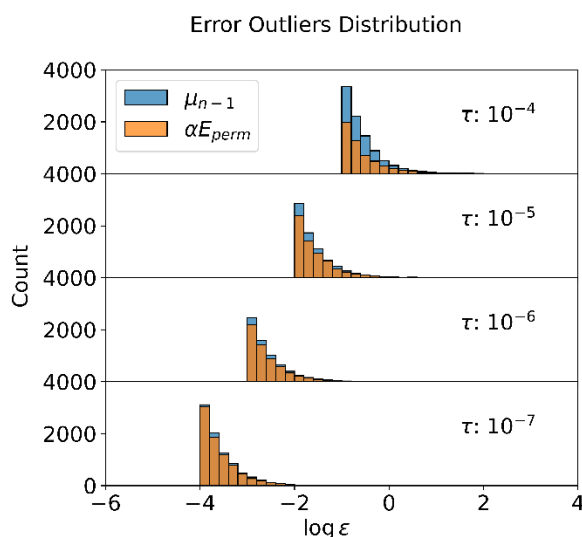


Figure 3: Distribution of Error Outliers Versus Convergence Tolerance (τ) in 100ps NVE Simulations Employing Residue Norm as a Convergence Measure. An error outlier is defined when the error is 1000 times larger than the convergence tolerance τ , with individual relative errors analyzed. The error axes are presented on a logarithmic scale. Each plot within the figure illustrates error distributions from simulations utilizing two different initial guesses for induced dipoles (μ_{n-1} and αE_{perm}). It is evident that simulations adopting μ_{n-1} as the initial guess introduce more errors compared to those employing αE_{perm} . However, these differences diminish as tighter convergence criteria are applied.

2. Maximum Error Is a Better Choice for Convergence Check

In our investigation, we strived to rectify the limitations of employing residue norm in convergence evaluation, thereby considering the maximum error as a possible candidate. The outcomes of our experiments demonstrated that the energy drifts in the majority of simulations, both those using and not using the history for the initial dipole assignment, were largely akin to those observed in TIP3P water (Table 2). This corresponded to an energy drift of $-5.56 \times 10^{-4} \text{ kcal / (mol} \cdot \text{ps)}$ (Refer to Table 1). The only deviations from this pattern were observed with the most lenient convergence tolerance of 10^{-1} . Consequently, it was noted that an implementation of a maximum error of 10^{-2} led to sufficiently low energy drift for both initial dipole assignments. Interestingly, tightening the convergence tolerance further (from 10^{-2} to 10^{-4}) did not yield any significant improvements in energy drifts. This observation suggests that induction error is no longer the primary determinant of the overall system energy drift when the maximum error in induction iteration is set to less than 10^{-2} .

It is essential to underscore that when choosing the maximum error as the convergence criterion, the quality of convergence must be high, devoid of extreme errors. This aspect becomes particularly critical when employing historical data as an initial estimate since improved convergence quality paves the way for a more accurate initial estimate.

To extend our analysis on the energy conservation properties of the tested systems, we also evaluated total energy fluctuations, as illustrated in Table 2. The table demonstrates that energy fluctuations at 10^{-2} , 10^{-3} and 10^{-4} convergence criteria were all in the order of 10^{-1} kcal/mol , both for tests conducted with and without history. This data corroborates the energy drift analysis presented earlier. Detailed energy plots are available in Figure S4 and S5. Lastly, Table 2 also encapsulates the average number of iteration steps, indicating that the utilization of history could still lessen the iteration by 2 to 3 steps. This observation reinforces the adoption of history for a more efficient induction iteration, provided the maximum error is employed as the convergence measure.

Table 2: Energy Drift (expressed in kcal/(mol·ps)), Energy RMSD (in kcal/mol), and Average Iteration Count for NVE Simulations Utilizing Maximum Error for Convergence Checking. This table outlines the outcomes of NVE simulations, detailing metrics such as energy drift, energy Root Mean Square Deviation (RMSD), and the average number of iterations, with the simulations specifically employing the maximum error as the criterion for convergence checking.

| τ | | 10^{-1} | 10^{-2} | 10^{-3} | 10^{-4} |
|------------------------------|-------------------|------------------------|------------------------|------------------------|------------------------|
| TIP3P energy drift | | -5.56×10^{-4} | | | |
| Energy drift | μ_{n-1} | -5.78×10^{-2} | 2.55×10^{-5} | 2.39×10^{-4} | 8.00×10^{-5} |
| | αE_{perm} | 2.11×10^{-3} | -7.33×10^{-4} | -2.61×10^{-4} | -5.63×10^{-4} |
| Energy RMSD | μ_{n-1} | 16.69 | 1.23×10^{-1} | 1.40×10^{-1} | 1.68×10^{-1} |
| | αE_{perm} | 6.64×10^{-1} | 2.43×10^{-1} | 1.45×10^{-1} | 2.03×10^{-1} |
| Average number of iterations | μ_{n-1} | 7.29 | 9.62 | 11.82 | 14.11 |
| | αE_{perm} | 10.06 | 12.31 | 14.60 | 16.84 |

3. Optimization of Initial Guess using MOE Scheme

To enhance the initial estimate for induced dipoles, we employed Multiple Order Extrapolation (MOE) schemes, capitalizing on historical information in an effective manner. Our investigative data implies that while the MOE scheme theoretically possesses the capacity for unlimited expansion, its potential for productivity improvement becomes marginal beyond the third order.

A rigorous investigation was conducted to assess the performance timing across diverse extrapolation orders and varying counts of preceding steps included for extrapolation. Specifically, we evaluated extrapolations of the first, second, and third orders in coordination with two to six preceding steps. A comprehensive encapsulation of the mean iteration number and cumulative CPU time per 1000 molecular dynamics (MD) steps under all test conditions is found in Table 3. Our inquiry identified several settings – five preceding steps in conjunction with first-order extrapolation, four preceding steps paired with second-order extrapolation and two preceding steps paired with third-order extrapolation – as consistently delivering the minimal iteration count and the shortest aggregate CPU duration. Moreover, in scenarios with a looser convergence tolerance, specifically around 10^{-2} , the second-order extrapolation proves to be a more advisable choice. Under the conditions of a tighter tolerance, in the realm of 10^{-3} and 10^{-4} , the third-order extrapolation is also a reasonable choice. It is pertinent to mention that the overall performance of our MOE could be further enhanced through the integration of a more efficient solver (as will be mentioned in later section), thereby generating richer and more valuable historical data for the MOE's optimal application. Not surprisingly, the ideal number of historical steps for the first-order MOE converges with the results of Wang and Skeel's study,²⁸ which is also a first-order extrapolation scheme. Consequently, aggressive utilization of historical data culminates in expected reductions in convergence duration and iteration count.

The distribution of initial errors utilizing different initial guesses is exhibited in Figure 4. It is apparent from the figure that a solitary historical step only mildly decreases error, while Wang and Skeel's extrapolation (WSE) effectuates a more substantial reduction. Our MOE, with optimal parameters, performs marginally better than WSE and subsequently reduces simulation time, though the performance improvements are higher for the tighter tolerance conditions.

Lastly, Table S2 portrays the energy drift values associated with varying extrapolation combinations across 1 ns NVE simulations. Contrary to observations by Wang and Skeel,²⁸ the energy drift values demonstrate remarkable consistency and similarly align with those documented in Table 2 (where $\tau \leq 10^{-2}$). This offers further substantiation for the reliability of utilizing maximum error in the convergence check as a means of ensuring energy conservation.

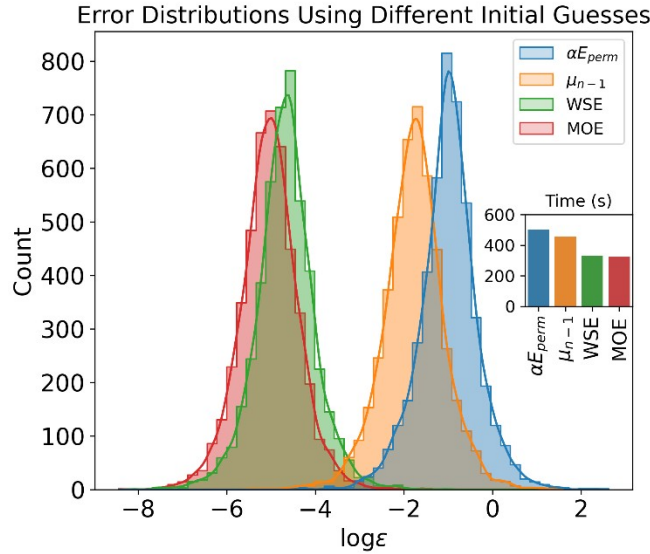


Figure 4. Distribution of Initial Errors with Varying Initial Guesses and their Associated CPU Time for 1000 MD Steps. The figure illustrates the initial error distribution for different initial guesses, with a convergence tolerance set at a maximum error of 10^{-2} . The simulation is performed using a standard Conjugate Gradient (CG). The initial guesses include: αE_{perm} (Permanent Field); μ_{n-1} (last step); WSE (Wang and Skeel’s extrapolation, which employs 5 preceding steps); and MOE (Multi-Order Extrapolation, which employs a third-order extrapolation using 2 previous steps). The inset demonstrates the average CPU time consumed for 1000 MD steps. The illustration suggests that the application of MOE results in the smallest error distribution and quickest overall performance.

4. Use of Local Iterations to Speed up Convergence

In the light of the maximum error employed for induction convergence checking and the optimized initial induced dipole guess, we ventured into further diminishing induction costs via the Preconditioned Conjugate Gradient (PCG) method. The concept of deploying local iterations to refine induced dipoles possessing suboptimal initial guesses was detailed in the Methods section. Initially, we aim to demonstrate the potency of the local iteration method. Figure 5 reveals the error distribution following one iteration of computation. It is evident from the figure that while traditional CG minimally reduces error, the application of a preconditioner (as per Wang and Skeel)²⁸ significantly curtails the error in just one iteration. Further, substituting our local iteration for the preconditioner leads to even greater error reduction.

Table 3: Performance Evaluation of the MOE (Multi-order Extrapolation) Scheme and Number of Preceding Steps Employed in the Least Square Method under Different Convergence Criteria. The table presents a detailed analysis of the performance of the MOE scheme in relation to the number of previous steps incorporated in the least square method, subject to various convergence criteria. The left column enumerates the average iteration count per MD step necessary to achieve convergence, while the right column depicts the average CPU time required to complete 1000 MD steps. Data pertaining to the WSE (Wang and Skeel’s extrapolation) scheme is included as a reference point.

| Convergence Criterion: 10^{-2} | | | |
|----------------------------------|---------|---------|---------|
| Number of | Order 1 | Order 2 | Order 3 |
| | | | |

| previous steps | | | |
|----------------------------------|----------------|----------------|----------------|
| Last 2 | 429.43 (9.98) | 324.83 (4.94) | 344.88 (6.13) |
| Last 3 | 368.03 (6.89) | 345.73 (6.17) | 405.18 (9.24) |
| Last 4 | 384.98 (7.59) | 325.62 (5.09) | 364.00 (7.07) |
| Last 5 | 330.77 (5.14) | 354.69 (6.59) | 412.23 (9.44) |
| Last 6 | 344.73 (5.95) | 339.86 (5.97) | 398.04 (8.47) |
| WSE | 331.02 (5.18) | | |
| Convergence Criterion: 10^{-3} | | | |
| | Order 1 | Order 2 | Order 3 |
| Last 2 | 493.56 (13.26) | 398.00 (8.04) | 352.96 (6.62) |
| Last 3 | 428.71 (9.96) | 355.74 (6.64) | 423.45 (9.64) |
| Last 4 | 445.86 (10.89) | 343.77 (6.17) | 373.76 (7.64) |
| Last 5 | 395.01 (8.39) | 370.13 (7.11) | 426.52 (9.96) |
| Last 6 | 408.13 (9.06) | 365.97 (7.00) | 414.91 (9.54) |
| WSE | 396.31 (8.43) | | |
| Convergence Criterion: 10^{-4} | | | |
| | Order 1 | Order 2 | Order 3 |
| Last 2 | 552.78 (16.48) | 455.05 (11.23) | 419.93 (8.86) |
| Last 3 | 496.20 (13.25) | 414.70 (8.91) | 481.73 (11.61) |
| Last 4 | 505.79 (14.01) | 427.31 (9.12) | 430.84 (9.74) |
| Last 5 | 457.68 (11.69) | 427.64 (9.14) | 493.01 (11.98) |
| Last 6 | 470.49 (12.35) | 431.08 (9.69) | 490.83 (11.81) |
| WSE | 458.19 (11.83) | | |

The efficacy of local iterations is inherently tied to the cut-off distance employed. An extended cut-off invariably enhances the convergence rate, albeit at a higher computational cost. Building upon the optimal initial guess identified in the previous subsection, we scrutinized different cut-off distances, the results of which are presented in Table 4. The data indicate that larger cut-off distances correlate with a decrease in iterations, which aligns with expectations, but these also increase the CPU time per MD step. Thus, it is vital to evaluate the total CPU time consumed, as portrayed in Table 4. It demonstrates that a cut-off distance of 4 Å demands the least amount of total CPU time.

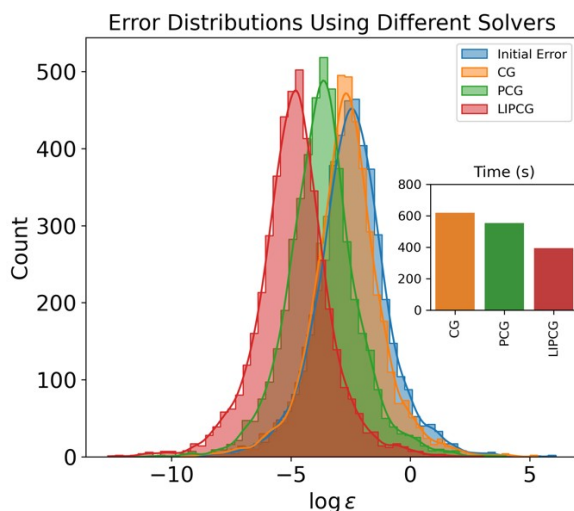


Figure 5: Error Distribution Following a Single Step of Iterative Method Implementation. The convergence tolerance for this analysis is set at a maximum error of 10^{-2} . The inset figure displays the CPU timing necessary for executing 1000 steps of Molecular Dynamics (MD). As demonstrated in the figure, the LIPCG method results in the smallest overall error distribution and is also associated with the least CPU time, thus indicating its superior efficiency in this context.

5. Use of Peek Step to Further Improve Performance

Striving to reduce the computational iterations and enhance overall system performance, we incorporated an additional 'peek' step — constructed as a single Jacobi under-relaxation iteration — into the LIPCG protocol. The optimization strategy for this procedural element is elaborated upon in the appended section. As represented in Table 5, this novel peek step hastens the convergence process when the convergence criteria are set to either 10^{-2} or 10^{-3} , culminating in an average reduction of approximately half an iteration step. Figure 6 visually illustrates the effects of the peek step, showing that the norm of residuals is halved, and maximum errors are significantly diminished post the peek step implementation.

Table 4: Optimal Cutoff Distance Determination for the Local Preconditioner using Best Parameters Derived from Table 3. This table employs the optimal parameters ascertained from Table 3 to ascertain the most suitable cutoff distance for the local preconditioner. The data reveals that the optimal cutoff is 4 Å. The left column represents the average iteration count per MD step necessary to achieve convergence, while the right column indicates the average CPU time required to execute 1000 MD steps.

| Convergence Criterion: 10^{-2} | | | |
|----------------------------------|---------------|---------------|---------------|
| Cut-off distance | Order1 | Order2 | Order3 |
| 3 Å | 267.17 (2.60) | 255.91 (2.31) | 271.04 (2.83) |
| 4 Å | 260.40 (2.34) | 255.23 (2.17) | 260.93 (2.60) |
| 5 Å | 264.16 (2.11) | 258.90 (2.09) | 272.84 (2.47) |
| 6 Å | 264.23 (1.96) | 265.62 (1.95) | 280.19 (2.36) |
| Convergence Criterion: 10^{-3} | | | |
| | Order1 | Order2 | Order3 |
| 3 Å | 305.44 (4.02) | 272.87 (3.03) | 280.23 (3.09) |
| 4 Å | 294.54 (3.57) | 271.01 (2.76) | 268.99 (2.81) |
| 5 Å | 307.48 (3.38) | 273.53 (2.58) | 284.64 (2.67) |
| 6 Å | 314.29 (3.17) | 279.57 (2.46) | 295.20 (2.58) |
| Convergence Criterion: 10^{-4} | | | |
| | Order1 | Order2 | Order3 |
| 3 Å | 332.88 (5.52) | 298.14 (4.20) | 295.26 (4.15) |
| 4 Å | 330.32 (5.05) | 296.37 (3.80) | 292.37 (3.74) |
| 5 Å | 334.50 (4.71) | 302.83 (3.56) | 301.60 (3.48) |
| 6 Å | 341.90 (4.43) | 314.29 (3.32) | 312.98 (3.25) |

Table 5: Comparative Analysis of Average Iteration Counts for Tests Conducted Without and With Peek Steps. This table provides a contrastive evaluation of the average number of iterations required in tests conducted both without and with the inclusion of peek steps. The left column represents the average iteration count per MD step needed to achieve convergence, while the right column denotes the average time necessary to complete 1000 MD steps.

| Convergence Criterion: 10^{-2} | | | |
|----------------------------------|---------------|---------------|---------------|
| | Order1 | Order2 | Order3 |
| Without Peek | 260.40 (2.34) | 255.23 (2.17) | 260.93 (2.60) |
| With Peek | 258.26 (2.23) | 241.33 (1.64) | 245.99 (2.10) |
| Convergence Criterion: 10^{-3} | | | |
| | Order1 | Order2 | Order3 |
| Without Peek | 294.54 (3.57) | 271.01 (2.76) | 268.99 (2.81) |
| With Peek | 292.74 (3.54) | 261.23 (2.43) | 257.71 (2.53) |
| Convergence Criterion: 10^{-4} | | | |
| | Order1 | Order2 | Order3 |
| Without Peek | 330.32 (5.05) | 296.37 (3.80) | 292.37 (3.74) |
| With Peek | 328.98 (5.03) | 290.94 (3.73) | 285.44 (3.69) |

6. Final Performance Using Optimal Strategy

With the implementation of our optimal strategy which includes local-iteration preconditioning, multi-order extrapolation, and a more proper selection of the convergence criterion, the overall performance of the induced dipole model has been significantly optimized. Our comparison of Wang and Skeel's approach with our optimal strategy, as illustrated in Table 6, highlights this performance improvement. As the table shows, our optimal strategy managed to achieve the same convergence quality with less time. Moreover, the degree of improvement was particularly pronounced under conditions requiring tighter convergence criteria. To further emphasize this point, let us consider a comparison with the CG method which utilizes only the last step's history information (as per Table S1). If we insist on a similar energy conservation quality which can be achieved at the convergence tolerance of 10^{-6} , we have managed to decrease the CPU time taken from 533.14s to 245.99s; this translates to an efficiency/speed gain over 210%. The results clearly illustrate the effectiveness of our proposed enhancements to the induced dipole model.

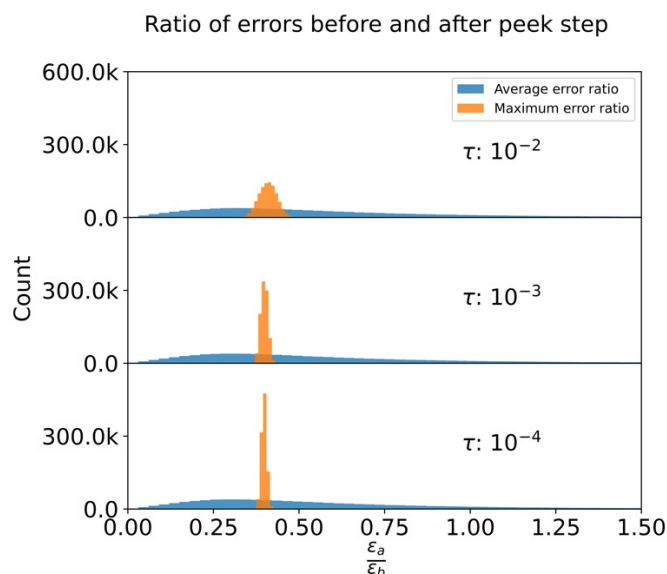


Figure 6: Distributions of Ratios Between Maximum/Average Errors Post-Peek Step and Pre-Peek Step. This figure illustrates the distributions of the ratios derived from the maximum and average errors after the peek step in relation to those prior to the peek step. A smaller ratio is desirable as it indicates a reduction in error.

Table 6: Performance Comparison Between the Optimal Strategy Developed in This Study and the Approach Proposed by Wang and Skeel. This table presents a comparative evaluation of the performance between the optimal strategy derived in this research and the methodology proposed by Wang and Skeel. In all simulations, the maximum error was employed as the convergence check to ensure energy conservation.

| Method | Relative Norm/CG/Last Step | | Maximum Relative Error/LIP-CG/MOE/Peek | | |
|-------------------------------|----------------------------|------------------|--|----------------|----------------|
| | 10^{-6} | 10^{-7} | 10^{-2} | 10^{-3} | 10^{-4} |
| Convergence Criterion | | | | | |
| Time per 1000 steps(Avg Iter) | 533.14 s (12.00) | 588.60 s (14.03) | 241.33s (1.64) | 257.71s (2.53) | 285.44s (3.69) |

Conclusion

Polarizable force fields, which account for interactions between charges and higher moments like dipoles and quadrupoles, have proven more accurate in describing atomistic interactions compared to high-level quantum mechanical calculations. Despite this, polarizable models such as the pGM model have grappled with energy conservation issues in NVE simulations unless a stringent convergence criterion is employed. Large energy drifts are indicative of inadequate convergence of induced dipoles. Tests performed using the pGM model have revealed that energy drifts are more significant when historical data is leveraged, irrespective of the tolerance used, aligning with observations made with other polarizable force fields. In all test scenarios, energy drifts when history is not incorporated are less than half of those where history is used. This discrepancy implies differences in the quality of induction convergence, even when using the same tolerance level.

Our detailed error analysis shows that the convergence quality and outlier errors are two reasons that prevents energy conservation. In general, using history will lead to a poorer

convergence quality with more outlier errors compared to not using history at the same convergence criterion. As a consequence, using the residue norm in convergence check fails to safeguard the convergence of the poorly converged dipoles, leading to unacceptable energy conservation properties unless an unusually tight convergence criterion is used. We therefore explored using the maximum error in the convergence check. Our tests show that energy drifts are qualitatively the same in NVE simulations with or without history when the maximum error is set to be less than 10^{-2} . Additionally, history can still improve the convergence rate in these tests; the iteration steps are reduced by 2 to 3 for the tested systems.

Based on our observations, we explored a scheme to improve induction convergence by utilizing more prudent initial guesses based on previous steps' induced dipoles. The MOE scheme was developed to produce better initial guesses. Our tests show that the third-order interpolation with two previous steps is consistently the most effective, reducing the iteration steps by another 2 to 5 for all tested conditions without impacting the energy conservation of the simulations.

Furthermore, we studied the effect of local iterations as a preconditioner to improve the convergence of error outliers when maximum errors are used in convergence check. The error analysis for the tests with and without history shows that the LIPCG scheme leads to smaller and more uniform errors than the standard CG and PCG. These analyses indicate that local iterations can be used to improve the convergence rate of the PCG method. We explored the optimal cutoff distance to be 4 \AA for the local iteration and under such condition, iteration steps are reduced by another 3 to 5, and overall simulation time is also reduced, despite the extra cost of local iterations.

Finally, we studied the effect of the "peek step" on the convergence quality. Overall, the residue norms are reduced to half, and the maximum errors are mostly reduced after the peek step. Interestingly, the peek step further improves the convergence rate when the criterion is set as 10^{-2} or 10^{-3} , with average iteration steps reduced by as much as a half step, likely due to better initial guesses when the convergence quality is better.

In summary, the development documented here shows that it is possible to achieve reasonable energy convergence with just a few iterations when both MOE and LIPCG schemes are used along with the peek step, if the maximum error is used to check the convergence quality. Apparently, the development and optimization documented here will need to be further tested in heterogeneous systems such as ions, proteins, nucleic acids in water. We will certainly revisit the issue in our future development when pGM force fields for ions, proteins, and nucleic acids become available. In addition, our current focus has been on the quality of energy conservation, so we have focused on NVE simulations with the 1fs time step. Clearly larger time steps are often used for NVT and NPT simulations and will lead to worse initial guesses of induced dipoles. Nevertheless, the timing with 2fs is still more efficient than that with 1fs due to the use of fewer PME calls as Wang and Skeel have pointed out. Nevertheless, this indicates further optimization is still feasible when the pGM model is applied to NVT and NPT simulations.

Supporting Information

Performance analysis of CG method, energy drift values of various MOE schemes with different convergence tolerances, energy plots under different convergence criterion with history and without history, parameter scanning for Jacobi under-relaxation.

This information is available free of charge at the website: <https://pubs.acs.org/>.

Acknowledgements

The authors gratefully acknowledge the research support from NIH (GM79383 to Y.D. and GM130367 to R.L.).

Appendix

Jacobi under-relaxation (JUR) is utilized as the peek step in our overall optimized induction iteration. The use of JUR instead of the standard Jacobi is to minimize the effect of overshooting by introducing a relaxation factor (ω) to the Jacobi method as reviewed in the Introduction. To fully take advantage its benefit as the peek step, we have optimized ω for the pGM model as shown in Figure S6, which is similar to that identified for the amoeba induced dipole model.²³

References

- (1) Lamoureux, G.; MacKerell Jr, A. D.; Roux, B. A simple polarizable model of water based on classical Drude oscillators. *The Journal of chemical physics* **2003**, *119* (10), 5185-5197.
- (2) Yu, H.; Hansson, T.; van Gunsteren, W. F. Development of a simple, self-consistent polarizable model for liquid water. *The Journal of chemical physics* **2003**, *118* (1), 221-234.
- (3) Rick, S. W.; Stuart, S. J.; Berne, B. J. Dynamical fluctuating charge force fields: Application to liquid water. *The Journal of chemical physics* **1994**, *101* (7), 6141-6156.
- (4) Ren, P.; Ponder, J. W. Polarizable atomic multipole water model for molecular mechanics simulation. *The Journal of Physical Chemistry B* **2003**, *107* (24), 5933-5947.
- (5) Caldwell, J.; Dang, L. X.; Kollman, P. A. Implementation of nonadditive intermolecular potentials by use of molecular dynamics: development of a water-water potential and water-ion cluster interactions. *Journal of the American Chemical Society* **1990**, *112* (25), 9144-9147.
- (6) Burnham, C. J.; Li, J.; Xantheas, S. S.; Leslie, M. The parametrization of a Thole-type all-atom polarizable water model from first principles and its application to the study of water clusters (n= 2-21) and the phonon spectrum of ice Ih. *The Journal of chemical physics* **1999**, *110* (9), 4566-4581.
- (7) Tan, Y.-H.; Luo, R. Continuum treatment of electronic polarization effect. *Journal of Chemical Physics* **2007**, *126* (9), 094103. DOI: 10.1063/1.2436871.
- (8) Tan, Y.-H.; Tan, C.; Wang, J.; Luo, R. Continuum polarizable force field within the Poisson-Boltzmann framework. *Journal of Physical Chemistry B* **2008**, *112* (25), 7675-7688. DOI: 10.1021/jp7110988.
- (9) Wang, J.; Cieplak, P.; Luo, R.; Duan, Y. Development of Polarizable Gaussian Model for Molecular Mechanical Calculations I: Atomic Polarizability Parameterization To Reproduce

- ab Initio Anisotropy. *Journal of Chemical Theory and Computation* **2019**, *15* (2), 1146-1158. DOI: 10.1021/acs.jctc.8b00603.
- (10) Wei, H.; Qi, R.; Wang, J.; Cieplak, P.; Duan, Y.; Luo, R. Efficient formulation of polarizable Gaussian multipole electrostatics for biomolecular simulations. *The Journal of Chemical Physics* **2020**, *153* (11), 114116.
- (11) Wei, H.; Cieplak, P.; Duan, Y.; Luo, R. Stress tensor and constant pressure simulation for polarizable Gaussian multipole model. *The Journal of chemical physics* **2022**, *156* (11), 114114.
- (12) Zhao, S.; Wei, H.; Cieplak, P.; Duan, Y.; Luo, R. PyRESP: A Program for Electrostatic Parameterizations of Additive and Induced Dipole Polarizable Force Fields. *Journal of Chemical Theory and Computation* **2022**, *18* (6), 3654-3670. DOI: 10.1021/acs.jctc.2c00230.
- (13) Zhao, S.; Wei, H.; Cieplak, P.; Duan, Y.; Luo, R. Accurate Reproduction of Quantum Mechanical Many-Body Interactions in Peptide Main-Chain Hydrogen-Bonding Oligomers by the Polarizable Gaussian Multipole Model. *Journal of Chemical Theory and Computation* **2022**, *18* (10), 6172-6188. DOI: 10.1021/acs.jctc.2c00710.
- (14) Zhao, S.; Cieplak, P.; Duan, Y.; Luo, R. Transferability of the Electrostatic Parameters of the Polarizable Gaussian Multipole Model. *Journal of Chemical Theory and Computation* **2023**, *19*(3), 924-941. DOI: 10.1021/acs.jctc.2c01048.
- (15) Applequist, J.; Carl, J. R.; Fung, K.-K. Atom dipole interaction model for molecular polarizability. Application to polyatomic molecules and determination of atom polarizabilities. *Journal of the American Chemical Society* **1972**, *94* (9), 2952-2960.
- (16) Cieplak, P.; Dupradeau, F.-Y.; Duan, Y.; Wang, J. Polarization effects in molecular mechanical force fields. *Journal of Physics: Condensed Matter* **2009**, *21* (33), 333102.
- (17) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *The Journal of chemical physics* **1993**, *98* (12), 10089-10092.
- (18) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *The Journal of chemical physics* **1995**, *103* (19), 8577-8593.
- (19) Crowley, M.; Darden, T.; Cheatham, T.; Deerfield, D. Adventures in improving the scaling and accuracy of a parallel molecular dynamics program. *The Journal of Supercomputing* **1997**, *11* (3), 255-278.
- (20) Duke, R. E.; Cisneros, G. A. Ewald-based methods for Gaussian integral evaluation: application to a new parameterization of GEM*. *Journal of molecular modeling* **2019**, *25* (10), 1-9.
- (21) Bulirsch, R.; Stoer, J. Iterative methods for the solution of large systems of linear equations. *Introduction to numerical analysis. 3rd edn.* Springer, New York, 2002, pp. 619-729.
- (22) Saad, Y. Preconditioned Iterations; Preconditioning Techniques. *Iterative methods for sparse linear systems.* 2nd edn. SIAM, Philadelphia, 2003, pp. 261-351.
- (23) Aviat, F.; Levitt, A.; Stamm, B.; Maday, Y.; Ren, P.; Ponder, J. W.; Lagardère, L.; Piquemal, J.-P. Truncated conjugate gradient: an optimal strategy for the analytical evaluation of the many-body polarization energy and forces in molecular simulations. *Journal of chemical theory and computation* **2017**, *13* (1), 180-190.
- (24) Davis, M. E.; McCammon, J. A. Solving the Finite-Difference Linearized Poisson-Boltzmann Equation - a Comparison of Relaxation and Conjugate-Gradient Methods. *Journal of Computational Chemistry* **1989**, *10* (3), 386-391.

- (25) Luo, R.; David, L.; Gilson, M. K. Accelerated Poisson-Boltzmann calculations for static and dynamic systems. *Journal of Computational Chemistry* **2002**, *23* (13), 1244-1253. DOI: 10.1002/jcc.10120|ISSN 0192-8651.
- (26) Wang, J.; Luo, R. Assessment of Linear Finite-Difference Poisson-Boltzmann Solvers. *Journal of Computational Chemistry* **2010**, *31* (8), 1689-1698. DOI: 10.1002/jcc.21456.
- (27) Cai, Q.; Hsieh, M.-J.; Wang, J.; Luo, R. Performance of Nonlinear Finite-Difference Poisson-Boltzmann Solvers. *Journal of Chemical Theory and Computation* **2010**, *6* (1), 203-211. DOI: 10.1021/ct900381r.
- (28) Wang, W.; Skeel, R. D. Fast evaluation of polarizable forces. *The Journal of chemical physics* **2005**, *123* (16), 164107.
- (29) Lipparini, F.; Lagardère, L.; Stamm, B.; Cancès, E.; Schnieders, M.; Ren, P.; Maday, Y.; Piquemal, J.-P. Scalable Evaluation of Polarization Energy and Associated Forces in Polarizable Molecular Dynamics: I. Toward Massively Parallel Direct Space Computations. *Journal of Chemical Theory and Computation* **2014**, *10* (4), 1638-1651. DOI: 10.1021/ct401096t.
- (30) Toukmaji, A.; Sagui, C.; Board, J.; Darden, T. Efficient particle-mesh Ewald based approach to fixed and induced dipolar interactions. *The Journal of chemical physics* **2000**, *113* (24), 10913-10927.
- (31) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *Journal of Computational Chemistry* **2005**, *26* (16), 1668-1688. DOI: 10.1002/jcc.20290.

TOC Graphic:

