The Journal of Chemical Physics

RESEARCH ARTICLE | MARCH 28 2017

Performance of extended Lagrangian schemes for molecular dynamics simulations with classical polarizable force fields and density functional theory

Valerio Vitale; Jacek Dziedzic; Alex Albaugh; Anders M. N. Niklasson ; Teresa Head-Gordon; Chris-Kriton Skylaris

Check for updates

J. Chem. Phys. 146, 124115 (2017)
https://doi.org/10.1063/1.4978684

View Online
Export Citation
CrossMark

This article may be downloaded for personal use only. Any other use requires prior permission of the author and AIP Publishing. This article appeared in (citation of published article) and may be found at https://doi.org/10.1063/1.4978684

26 February 2024 09:53:54

AIP Publishing

THE JOURNAL OF CHEMICAL PHYSICS **146**, 124115 (2017)

Performance of extended Lagrangian schemes for molecular dynamics simulations with classical polarizable force fields and density functional theory

Valerio Vitale,[1] Jacek Dziedzic,[1,2] Alex Albaugh,[3] Anders M. N. Niklasson,[4] Teresa Head-Gordon,[3,5,6,7] and Chris-Kriton Skylaris[1]
[1]*School of Chemistry, University of Southampton, Southampton SO17 1BJ, United Kingdom*
[2]*Faculty of Applied Physics and Mathematics, Gdańsk University of Technology, Gdańsk, Poland*
[3]*Department of Chemical and Biomolecular Engineering, University of California, Berkeley, California 94720, USA*
[4]*Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*
[5]*Department of Chemistry, University of California, Berkeley, California 94720, USA*
[6]*Department of Bioengineering, University of California, Berkeley, California 94720, USA*
[7]*Chemical Sciences Division, Lawrence Berkeley National Laboratory, University of California, Berkeley, California 94720, USA*

(Received 25 January 2017; accepted 2 March 2017; published online 28 March 2017)

Iterative energy minimization with the aim of achieving self-consistency is a common feature of Born-Oppenheimer molecular dynamics (BOMD) and classical molecular dynamics with polarizable force fields. In the former, the electronic degrees of freedom are optimized, while the latter often involves an iterative determination of induced point dipoles. The computational effort of the self-consistency procedure can be reduced by re-using converged solutions from previous time steps. However, this must be done carefully, as not to break time-reversal symmetry, which negatively impacts energy conservation. Self-consistent schemes based on the extended Lagrangian formalism, where the initial guesses for the optimized quantities are treated as auxiliary degrees of freedom, constitute one elegant solution. We report on the performance of two integration schemes with the same underlying extended Lagrangian structure, which we both employ in two radically distinct regimes—in classical molecular dynamics simulations with the AMOEBA polarizable force field and in BOMD simulations with the ONETEP linear-scaling density functional theory (LS-DFT) approach. Both integration schemes are found to offer significant improvements over the standard (unpropagated) molecular dynamics formulation in both the classical and LS-DFT regimes. *Published by AIP Publishing.* [http://dx.doi.org/10.1063/1.4978684]

26 February 2024 09:53:54

MOST WIEDZY Downloaded from mostwiedzy.pl

I. INTRODUCTION AND MOTIVATION

Since the first classical molecular dynamics (MD) calculations in the late 1950s, a plethora of schemes employing a hierarchy of approximations for the description of interatomic interactions have emerged, ranging from classical fixed charge models to fully *ab initio* treatments. While molecular dynamics has proved to be an invaluable tool for elucidating many chemical, biological, and physical processes involving a large number of atomic degrees of freedom,[1–3] challenges remain in finding models and methods that are able to adequately describe complex environments and show a favorable accuracy-efficiency trade-off. In this regard, promising approaches include linear-scaling density functional theory (LS-DFT),[4,5] classical polarizable force fields,[6,7] self-consistent field (SCF) tight-binding,[8] and semi-empirical quantum mechanical (SQM) approaches.[9]

The last two decades have witnessed significant developments in the sophistication of classical force fields,[10,11] motivated by the need to overcome the deficiencies identified in earlier fixed point charge models. In the absence of explicit treatment of polarization, the dynamic response of a molecular system to a given environment, e.g., an organic solvent, or metallic ions in solution becomes difficult to capture.[7] Moreover, for conditions far from the ones used to parametrize the potential and charges, such as at extreme pressures or temperatures, for different physical phases, or at interfaces, fixed point charge models can perform quite poorly.[11] Inclusion of polarization effects is hence crucial for obtaining an adequate description of complex systems and improving the transferability of classical models. As a consequence, a gamut of polarization models has emerged, employing Drude oscillators,[12,13] fluctuating charges,[14,15] and induced point dipoles[16–22] or higher multipoles.[23] In most cases, a physically coherent treatment of polarization involves a relatively costly variational self-consistent determination of the induction response, although alternate schemes have recently been proposed.[24,25] See Refs. 7, 11, and 6 for a review.

Ab initio molecular dynamics (AIMD) techniques approach the transferability problem from the opposite direction, by directly (Born-Oppenheimer MD, BOMD) or indirectly (Car-Parrinello MD, CPMD) modeling the electronic degrees of freedom on a quantum-mechanical (QM)

footing, while treating atomic nuclei classically. The underlying quantum mechanical scheme is selected depending on the desired trade-off between accuracy and computational efficiency—starting from inexpensive, qualitative methods like tight binding (TB), through density-functional-based tight binding (DFTB),[26,27] density functional theory (DFT) to correlated wavefunction methods. The majority of these approaches are self-consistent, requiring an iterative procedure to arrive at the converged energy and associated electronic orbitals, at each MD step.

The existence of a self-consistency loop is a common feature of classical MD calculations with polarizable force fields and BOMD calculations, leading to important consequences for energy conservation in both approaches. Energy conservation is known to be sensitive to how close to convergence the self-consistency loop gets. In polarizable MD with induced dipoles, this is a consequence of a non-Hellmann-Feynman-like force term that is proportional to the energy gradient residual $\frac{\partial U}{\partial \boldsymbol{\mu}}$.[25] This term vanishes in schemes where induced dipoles are not determined self-consistently (e.g., iAMOEBA[24]), and in self-consistent schemes, it can only be neglected when the iterative dipole equations are tightly converged. In DFT BOMD, the error in the forces is first order with respect to the error in the incompletely converged wavefunction, even though the error in the Kohn-Sham energy is second order.[2] This means that it is significantly more difficult to calculate accurate DFT forces for driving the ions than it is to calculate energies. Additionally, the numerical machinery of DFT, which typically involves FFTs on grids with finite spacing and finite extents, numerical integrations on a variety of grid, the use of moment approximations, etc., results in an increase in the inaccuracy of the forces compared to the classical case, where closed analytical expressions are often directly implementable. Furthermore, many DFT approaches have to contend with additional (beyond Hellmann-Feynman) contributions to the forces, known as Pulay forces,[28] that arise as a result of the dependence of the basis set on ionic positions. Although the magnitude of the Pulay forces is, in principle, independent of how close the electronic configuration is to the ground state,[2] errors in *approximate* treatments of Pulay forces can be larger when the system is insufficiently converged.[29] More precisely, this is relevant only for *in situ* optimized localized orbitals, whereas for fixed localized orbitals, the Pulay forces can be computed exactly. Finally, in LS-DFT approaches, additional difficulties arise due to the use of strictly localized orbitals,[30] where further non-Hellmann-Feynman contributions to the forces can be expected to arise as a consequence of localization constraints[31] or of the incomplete convergence of localized orbitals,[29,31] or, to a lesser extent, the density matrix.[29]

For polarizable classical MD and BOMD calculations alike, tight SCF convergence is thus a necessary condition for maintaining adequate energy conservation. The large number of SCF steps mandated by this requirement leads to an undesired increase of calculation wall times. It is tempting to try to speed up convergence by not starting the SCF procedure from the same fixed initial guess at each MD time step, re-using instead the converged solutions obtained in preceding time steps. Although the dangers of such intuitive extrapolation schemes have been known since early 1990s,[32,33] adequate solutions have been proposed only recently.[34–36] Simply put, the self-consistent solution is only independent of the initial guess under exact SCF optimization. In practice, the SCF optimization is always incomplete, leading to memory effects and the breaking of time-reversal symmetry,[36] and, in consequence, to systematic errors in energy gradients that manifest as a drift in microcanonical energy.[35] Such undesired memory effects can be elegantly dealt with through extended Lagrangian BOMD (XL-BOMD, EL/SCF) formulations, where the initial guesses are not extrapolated, but rather *propagated* as extended degrees of freedom, allowing time-reversibility to be recovered.

In this paper, we focus our attention on two integration schemes proposed recently—a *dissipative* formulation by Niklasson *et al.*[36] and an *inertial* formulation by Albaugh *et al.*[37] These improvements to the original scheme by Niklasson *et al.*[38] use different strategies to mitigate the problem of error accumulation that causes two extended potential energy surfaces—that of the real degrees of freedom and that of the initial guesses—to diverge for longer simulation times. In this work, we study the properties and performance of the two schemes in MD simulations of liquid water, which we carry out in two notably different regimes—using the polarizable, classical force field AMOEBA and using the ONETEP LS-DFT formulation. Our comparison employs similar systems, but tests how the schemes operate under distinct conditions (long simulation times with inherently accurate forces and classical potential energy surfaces vs. short simulation times with inherently more noisy forces and quantum-mechanical potential energy surfaces). This allows us to highlight the differences between the two schemes, and their strengths and weaknesses in each of the two regimes.

We begin with a short presentation of the classical and quantum-mechanical approaches used in our study (Section II), followed by a presentation of the original extended Lagrangian scheme (Section III), and the dissipative (Section IV) and inertial (Section V) integration scheme variants, pausing briefly to comment on issues specific to linear-scaling BOMD and its non-orthogonal orbital variants in particular. Sections VI, VII A, and VII B describe our computational setup and the results obtained in the simulations, respectively. We finish with conclusions (Section VIII), where we summarize our observations and lay out directions for promising future research.

II. METHODS

A. Classical polarizable force-field molecular dynamics

For our analysis of classical molecular dynamics, we have adopted the polarizable force field AMOEBA,[20,39] as implemented in the TINKER[18] program. AMOEBA belongs to a newer generation of force fields that go beyond the time-honored model using pairwise-additive interactions of fixed point charges. In AMOEBA, electrostatic interactions are computed from interactions between point multipoles, where each atomic site I is host to a set of permanent multipoles, $\mathbf{M}_I = \{q_I, \boldsymbol{\mu}_I, \mathbf{Q}_I\}$,

26 February 2024 09:53:54

MOST WIEDZY Downloaded from mostwiedzy.pl

representing a charge, dipole, and quadrupole, respectively. Permanent multipoles are parametrized from *ab initio* calculations.[40–42] Alongside permanent multipoles, each atomic site is host to an induced dipole $\boldsymbol{\mu}_I^{\text{ind}}$, allowing polarization effects to be explicitly captured. Scaling of electrostatic interactions based on interatomic connectivity[19] and Thole damping[43] is used to ensure a smooth transition between the electrostatic and bonded (valence) descriptions of interactions and to avoid the polarization catastrophe.

The polarization effect in AMOEBA is modeled by induced dipoles, $\boldsymbol{\mu}_I^{\text{ind}}$, placed on each atomic site, whose magnitude is determined by the site-specific isotropic polarizability α_I and the total external electric field exerted,

$$\boldsymbol{\mu}_I^{\text{ind}} = \alpha_I(\mathbf{E}_I + \mathbf{E}_I'), \tag{1}$$

where $\mathbf{E}_I$ is the electric field owing to the permanent multipoles on other fragments, and $\mathbf{E}_I'$ is the field generated by the induced dipoles on all the other atomic sites,

$$\mathbf{E}_I = \sum_J \mathbf{T}_{IJ}\mathbf{M}_J^{(\text{d})}, \tag{2}$$

$$\mathbf{E}_I' = \sum_{J\neq I} \mathbf{T}_{IJ}'\boldsymbol{\mu}_J^{\text{ind}}, \tag{3}$$

where $\mathbf{T}_{IJ}'$ now refers to appropriate powers of $1/r_{IJ}$ according to the dipole induction and the superscript (d) refers to special scaling factors used for electrostatic interactions in AMOEBA.[18] Since the RHS of Eq. (1) relies on the induced dipoles, a procedure for guaranteeing self-consistency of induced dipoles is required. This is usually achieved through iterative techniques, such as successive over-relaxation (SOR)[18,44] or the more recent use of a precondition conjugate gradient self-consistent field (CG-SCF) approach.[45] With converged $\{\boldsymbol{\mu}_I^{\text{ind}}\}$, the polarization energy is determined by

$$U_{\text{ele}}^{\text{ind}} = -\frac{1}{2}\sum_I \boldsymbol{\mu}_I^{\text{ind}} \cdot \mathbf{E}_I. \tag{4}$$

A detailed discussion of the AMOEBA electrostatics model is beyond the scope of this paper, and the interested reader is referred to Refs. 17 and 46 for a more in-depth discussion, and a description of bonded and van der Waals interactions used in the AMOEBA model.

B. Linear scaling *ab initio* molecular dynamics

In contrast to classical molecular dynamics, in AIMD calculations, the electronic degrees of freedom are not integrated out but treated explicitly by finding approximate solutions to QM equations. Nuclei are treated as classical particles, and the forces acting on them are obtained from electronic structure calculations. In this work, we made use of Kohn-Sham (KS) density functional theory[47] to solve the electronic problem. In particular, all AIMD calculations in this work have been carried out in Onetep.[48,49]

In the Onetep framework, the electronic degrees of freedom are described through the KS reduced (spinless) single-particle density matrix operator $\hat{\rho}$, which in position representation reads

$$\rho(\mathbf{r}, \mathbf{r}') = \phi_\alpha(\mathbf{r})K^{\alpha\beta}\phi_\beta^*(\mathbf{r}'), \tag{5}$$

where $\mathbf{K}$ is the density kernel matrix, a (2,0)-tensor, which is a generalization of the occupation number matrix in a non-orthogonal basis: $K^{\alpha\beta} = \sum_i M^{\dagger\alpha}{}_i f_i M_i{}^\alpha$, where $\phi(\mathbf{r})$ are generalized non-orthogonal Wannier functions, hereafter termed NGWFs, and they are related to the eigenstates of the KS Hamiltonian via a non-unitary matrix $\mathbf{M}$ as

$$\psi_i(\mathbf{r}) = \phi_\alpha(\mathbf{r})M^\alpha{}_i, \tag{6}$$

where we have made use of Einstein's convention for repeated indices. Greek indices are used to label non-orthogonal objects, while Latin indices label orthogonal quantities.

The NGWFs are atom-centered real functions (reflecting the Γ-point approximation) which are strictly localized, that is to say, they are non-zero only within a localization region (LR) centered around the atom they belong to. The use of non-orthogonal functions allows us to impose a tighter LR than for orthogonal functions.[50–52] The disadvantage is having to explicitly take tensorial correctness into account. As a consequence, the (0, 2) metric tensor, also known as the overlap matrix $S_{\alpha\beta} = \langle\phi_\alpha|\phi_\beta\rangle$, is not the identity matrix. The KS electron density function $n(\mathbf{r})$ is simply given by the diagonal part of the density matrix $n(\mathbf{r}) = 2\rho(\mathbf{r}, \mathbf{r})$, where the factor of 2 accounts for spin degeneracy (which we assume throughout this work).

Linear scaling is achieved by exploiting the principle of "nearsightedness" of electronic matter.[53,54] For systems with a non-zero band gap, i.e., insulators and semiconductors, the density matrix (5) decays exponentially as a function of distance between two points in space. In practical calculations, it is therefore possible to truncate the density matrix by imposing a radial cutoff,

$$\rho(\mathbf{r}, \mathbf{r}') = 0 \quad \text{for } |\mathbf{r} - \mathbf{r}'| > R_{\text{cut}}, \tag{7}$$

where R_{cut} is an assumed cutoff distance.

Within Onetep two minimization strategies are available for solving the electronic problem in a self-consistent way:

1. The total energy can be minimized by optimizing both the elements of the density kernel $\mathbf{K}$ and the expansion coefficients of the NGWFs in terms of an underlying periodic cardinal sinc (psinc) function basis[49] (*in situ* optimization).
2. Alternatively, only the elements of the density kernel can be optimized, for a fixed set of NGWFs that have been suitably initialized, e.g., to pseudoatomic orbitals (PAOs), or to orbitals that have been pre-optimized in advance.

When NGWFs are optimized *in situ*, a minimal set of NGWFs is sufficient for obtaining high accuracy and systematic convergence of total energies to those of a plane-wave approach with KS orbitals. When total energy is minimized with respect to the density kernel elements only, a non-minimal basis size is typically necessary to achieve accurate energies and forces, increasing the memory requirements of this strategy, even if it is generally faster.

In both approaches, a modified Li-Nunes-Vanderbilt (LNV) algorithm[50,55] is employed to minimize the energy. The density kernel $\mathbf{K}$ is expressed in terms of an auxiliary matrix $\mathbf{L}$ as $\mathbf{K} = 3\mathbf{LSL} - 2\mathbf{LSLSL}$. Energy is then minimized with

26 February 2024 09:53:54

MOST WIEDZY Downloaded from mostwiedzy.pl

respect to the elements $L^{\alpha\beta}$, which allows the idempotency constraint to be satisfied to first order during the minimization.

In this paper, we have adopted the second strategy, where the NGWFs are not optimized and we used a non-minimal PAOs basis. Convergence is deemed achieved when the RMS of the energy gradient $g^{\alpha\beta}$ falls below a specified threshold,

$$\sqrt{\frac{1}{N}\left|\langle g^{\alpha\beta}|g_{\alpha\beta}\rangle\right|} < \varepsilon_{\mathrm{LNV}}. \tag{8}$$

For a comprehensive description of the energy minimization algorithm and linear-scaling calculations, the reader is referred to Refs. 48, 49, 56, and 57.

III. EXTENDED LAGRANGIAN FORMALISM

Extended Lagrangian methods were originally introduced to perform MD simulations of systems in statistical ensembles other than the microcanonical.[58–60] For example, the action of a thermostat (barostat) can be described through the interaction of the system with a heat bath (piston). By postulating operational expressions for the kinetic and potential energy of the extra dynamical variables, one can write an extended Lagrangian which intrinsically takes into account the new variables. Similarly, CPMD defines the electronic states as extended classical dynamical variables (classical fields), with a fictitious kinetic energy and a fictitious mass, with the idea of avoiding the expensive electronic self-consistency procedure altogether. Recently, several authors have proposed schemes based on the CP Lagrangian, using the density matrix (DM) elements as the extra degrees of freedom,[61,62] where the orthonormality constraints are replaced with the idempotency constraint of the DM.

Starting from the broken time-reversal symmetry problem in BOMD,[34] Niklasson *et al.* introduced a time-reversible extrapolation scheme for the electronic degrees of freedom.[63] It is now recognized that the original scheme can be derived from an extended Lagrangian,[38,64] in which an additional set of degrees of freedom is propagated alongside the nuclei with the purpose of generating good quality time-reversal guesses for the SCF calculations. Since the extra degrees of freedom are only a computational device to reduce the number of iterations in the SCF step, we will refer to them as *auxiliary* degrees of freedom hereafter. In its most recent refinement, it has been shown that it can be formulated to completely avoid the SCF problem.[64]

Quite generally, a system of classical interacting nuclei moving in an external potential field U, which contains a SCF-derived component, has the following Lagrangian:

$$\mathcal{L}^{\mathrm{SCF}}(\{\mathbf{R}_I\},\{\mathbf{V}_I\}) = \frac{1}{2}\sum_I M_I V_I^2 - U(\{\mathbf{R}_I\},\{\chi_{\mathrm{SCF},\mathbf{a}}\}), \tag{9}$$

where $\{\mathbf{R}_I\}$ and $\{\mathbf{V}_I\}$ represent the sets of the nuclear positions and velocities, respectively, M_I is the mass of atom I, and $\{\chi_{\mathrm{SCF},\mathbf{a}}\}$ is the set of converged degrees of freedom that generate the SCF-derived part of the external field. The subscript "**a**" represents a generic collection of indices, which can account for atomic-centered quantities, such as atomic dipoles in classical polarizable force field, as well as global (atom-independent) quantities, such as the density matrix elements in DFT. The detailed form of the potential energy U depends on the model employed, but in general can be cast as a sum of an SCF-independent part, such as permanent electrostatics or dispersion interactions, and a potential energy component obtained through an SCF procedure, for example, the solution of inducible dipoles in the case of a classical polarizable force field.

Following the work of Niklasson,[38] we can introduce generalized auxiliary degrees of freedom $\{\zeta_\mathbf{a}\}$ into the Lagrangian (9), provided we have a definition for their kinetic and potential energies. We want $\{\zeta_\mathbf{a}\}$ to closely follow the dynamics of $\{\chi_{\mathrm{SCF},\mathbf{a}}\}$, so that they can serve as initial guesses for $\{\chi_\mathbf{a}\}$ in the SCF calculations. To keep the energy expression simple, a harmonic potential centered around the converged solution $\{\chi_{\mathrm{SCF},\mathbf{a}}\}$ can be employed, with a single parameter $k = m\omega^2$ to control the steepness of the well, yielding

$$\mathcal{L}^{\mathrm{ext}} = \mathcal{L}^{\mathrm{SCF}} + \frac{1}{2}m\sum_\mathbf{a}\dot{\zeta}_\mathbf{a}^2 - \sum_\mathbf{a}\frac{k}{2}(\chi_{\mathrm{SCF},\mathbf{a}} - \zeta_\mathbf{a})^2, \tag{10}$$

where $\dot{\zeta}_\mathbf{a}$ represent the generalized velocities of the auxiliary degrees of freedom and m represents their mass. The extended Lagrangian (10) can now be used to derive a new set of equations of motion (Euler-Lagrange equations),

$$M_I\ddot{\mathbf{R}}_I = -\frac{\partial U}{\partial \mathbf{R}_I} + k(\chi_{\mathrm{SCF},\mathbf{a}} - \zeta_\mathbf{a})\frac{\partial \zeta_\mathbf{a}}{\partial \mathbf{R}_I}, \tag{11a}$$

$$m\ddot{\zeta}_\mathbf{a} = k(\chi_{\mathrm{SCF},\mathbf{a}} - \zeta_\mathbf{a}). \tag{11b}$$

In the limit $m \to 0$ ($k \to 0, k/m \to \omega^2$), we obtain

$$M_I\ddot{\mathbf{R}}_I = -\frac{\partial U}{\partial \mathbf{R}_I}, \tag{12a}$$

$$\ddot{\zeta}_\mathbf{a} = \omega^2(\chi_{\mathrm{SCF},\mathbf{a}} - \zeta_\mathbf{a}). \tag{12b}$$

It is worth pointing out that (12a) is exactly the same equation we would have obtained where there are no auxiliary variables introduced, i.e., when using (9) instead of (10). This is a consequence of taking the limit $m \to 0$ only after having derived the Euler-Lagrange equations. In so doing, we recover the correct potential energy surface, provided the SCF procedure is converged exactly. Integration of (12b) gives $\{\zeta_\mathbf{a}(t)\}$, which in turn provide the initial guesses for the SCF calculations.

A. Adaptation to classical polarizable force-field methods

In order to extend Niklasson's extended Lagrangian methods to classical polarization, we introduce a set of auxiliary atomic dipoles as extra variables.[37] More specifically, the real self-consistent induced dipoles $\{\boldsymbol{\mu}_{\mathrm{SCF},I}\}$ take the role of $\{\chi_{\mathrm{SCF},\mathbf{a}}\}$, and the auxiliary dipoles $\{\boldsymbol{\mu}_I\}$ replace $\{\zeta_\mathbf{a}\}$ in (10). The nuclear centers $\mathbf{R}_I$ are propagated in the usual way, i.e., according to (13a), while the real self-consistent induced dipoles are solved for using an SCF solver initiated by an initial guess that is propagated through the auxiliary dipoles according to (13b),

$$M_I\ddot{\mathbf{R}}_I = -\left.\frac{\partial U\left(\{\mathbf{R}_I\},\{\boldsymbol{\mu}_{\mathrm{SCF},I}\}\right)}{\partial \mathbf{R}_I}\right|_{\{\boldsymbol{\mu}_I\}}, \tag{13a}$$

$$\ddot{\boldsymbol{\mu}}_I = \omega^2\left(\boldsymbol{\mu}_{\mathrm{SCF},I} - \boldsymbol{\mu}_I\right), \tag{13b}$$

26 February 2024 09:53:54

MOST WIEDZY Downloaded from mostwiedzy.pl

where $U\left(\{\mathbf{R}_I\},\{\boldsymbol{\mu}_{\mathrm{SCF},I}\}\right)$ is the total potential energy from the AMOEBA force field.

B. Adaptation to linear-scaling DFT

In the case of linear-scaling DFT, the density matrix elements $K^{\alpha\beta}$ would take the role of the real degrees of freedom $\chi_{\mathbf{a}}$ in (10). Analogously to the classical case, we would proceed by introducing matrix elements of an auxiliary matrix $X^{\alpha\beta}$ as the auxiliary degrees of freedom $\zeta_{\mathbf{a}}$. However, simply substituting quantities in (12b) leads to equations that are geometrically inconsistent due to the tensorial nature of $\mathbf{K}$ and $\mathbf{X}$ and the fact that the underlying metric also changes with time. It is worth stressing here that the metric tensor $\mathbf{S}$ is not propagated (as the NGWFs are not treated as dynamical variables), but rather it is generated at every MD step from the current NGWFs. Arita *et al.* have proposed an alternative scheme[65] based on propagating the matrix elements $(\mathbf{KS})^{\alpha}{}_{\beta}$ as dynamical variables in (10), with associated $X^{\alpha}{}_{\beta}$ as auxiliary degrees of freedom. This approach has the advantage of propagating a representation of the density matrix which maintains the correct metric. At a given MD step n, the initial guess $\mathbf{K}^{\mathrm{init}}$ for the SCF procedure is computed from the auxiliary matrix $\mathbf{X}$ as

$$\mathbf{K}^{\mathrm{init}} = \mathbf{X}\mathbf{S}^{-1}, \tag{14}$$

where $\mathbf{S}^{-1}$ is the inverse overlap matrix at step n, approximated through an iterative Hotelling algorithm[66] (to maintain a linear scaling behavior).

The disadvantage of the above approach is that it does not preserve the symmetry of the density kernel matrix, i.e., $\mathbf{K}^{\dagger} \neq \mathbf{K}$, since at a given step, $\mathbf{X}$ and $\mathbf{S}^{-1}$ do not commute in general. One possible solution is to instead employ the symmetrized version of (14). In our experience, this quickly leads to instabilities for larger systems, particularly with a velocity-Verlet integrator (to be outlined in Section V).

Here, we propose a different approach, based on a different, orthogonal representation of the density kernel matrix $\mathbf{K}^{\perp} = \mathbf{S}^{\frac{1}{2}}\mathbf{K}\mathbf{S}^{\frac{1}{2}}$, where $\mathbf{S}^{\frac{1}{2}}$ is computed through a modified Newton-Schulz linear-scaling algorithm.[67] This procedure can be performed to a desired level of accuracy and can be made numerically exact, thus not introducing time-irreversibility, or, strictly speaking, it does not introduce any more time-irreversibility than other operations performed with finite-precision floating-point arithmetic, such as matrix multiplications or inversions. The initial guess for the SCF procedure at a given step n is given by

$$\mathbf{K}^{\mathrm{init}} = \mathbf{S}^{-\frac{1}{2}}\mathbf{X}\mathbf{S}^{-\frac{1}{2}}, \tag{15}$$

which ensures that $\mathbf{K}$ is symmetric at all times.

The Euler-Lagrange equations in our framework read

$$\mathrm{M}_I\ddot{\mathbf{R}}_I = -\left.\frac{\partial E[\mathbf{K}_{\mathrm{SCF}},\{\phi_{\mathrm{SCF}}\};\{\mathbf{R}_I\}]}{\partial \mathbf{R}_I}\right|_{\mathbf{K},\{\phi\}}, \tag{16a}$$

$$\ddot{\mathbf{X}} = \omega^2[\mathbf{K}^{\perp}_{\mathrm{SCF}} - \mathbf{X}], \tag{16b}$$

where $E[\mathbf{K}_{\mathrm{SCF}},\{\phi_{\mathrm{SCF}}\};\{\mathbf{R}_I\}]$ is the potential energy in ONETEP,

$$E[\mathbf{K}_{\mathrm{SCF}},\{\phi_{\mathrm{SCF}}\};\{\mathbf{R}_I\}] = E_{\mathrm{elec}}[\mathbf{K}_{\mathrm{SCF}},\{\phi_{\mathrm{SCF}}\};\{\mathbf{R}_I\}] \\ E_{\mathrm{Ewald}}(\{\mathbf{R}_I\}) + E_{\mathrm{disp}}(\{\mathbf{R}_I\}), \tag{17}$$

given by a sum of three terms: (1) the electronic potential energy E_{elec}; (2) the Ewald-Coulombic interaction energy of the atomic cores; (3) an empirical dispersion energy correction for dealing with a well-known deficiency of generalized gradient approximation (GGA) functionals in describing dispersion interactions.[68]

Within KS theory, the potential energy E_{elec} can be cast as a functional of the single-particle density matrix, hence the dependence on $\mathbf{K}$ and $\{\phi_\alpha(\mathbf{r})\}$, as

$$\begin{aligned} E_{\mathrm{elec}}[\mathbf{K}_{\mathrm{SCF}},\{\phi_{\mathrm{SCF}}\};\{\mathbf{R}_I\}] = &-\frac{1}{2}\int \mathrm{d}\mathbf{r}\left[\nabla^2_{\mathbf{r}'}\rho(\mathbf{r},\mathbf{r}')\right]_{\mathbf{r}'=\mathbf{r}} \\ &+\frac{1}{2}\int\int \mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}'\,\frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \\ &+\int \mathrm{d}\mathbf{r}\, v_{\mathrm{ext}}(\{\mathbf{R}_I\})n(\mathbf{r}) \\ &+E_{\mathrm{XC}}[n(\mathbf{r})], \end{aligned} \tag{18}$$

where the terms on the RHS represent, respectively, the kinetic energy of the non-interacting KS states, the classical interaction between charged densities, the potential interaction energy of electrons and clamped nuclei, and the exchange-correlation energy. Equations (16a) and (16b) are the AIMD counterparts to the classical Equations (13a) and (13b).

IV. EXTENDED LAGRANGIAN WITH DISSIPATIVE VERLET INTEGRATOR (dXL)

It is now recognized[37,38] that the above simple formulation suffers from numerical instabilities in the evolution of the auxiliary degrees of freedom. In fact, the velocities of the auxiliary degrees of freedom increase in an unbounded fashion, ultimately resulting in initial guesses that are unacceptably far from the converged values, negating the efficiency gains of the scheme.[37] The origin of this phenomenon lies in the fact that exact convergence is never achieved in practical calculations, which couples (12a) and (12b) through a "memory effect"[36] or

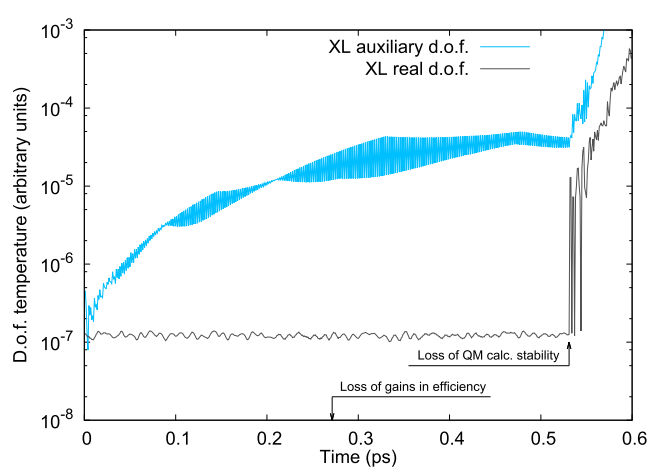


FIG. 1. Illustration of the runaway accumulation of kinetic energy in the auxiliary degrees of freedom in the original XL scheme. See text in Sec. IV.

26 February 2024 09:53:54

MOST WIEDZY Downloaded from mostwiedzy.pl

kinetic resonances.[37] Energy can therefore flow from the real degrees of freedom to the (massless) auxiliary ones, producing a runaway increase in the velocities of the latter. This is demonstrated in Fig. 1, which displays the runaway in the velocities for a 64 H_2O-molecule AIMD simulation, with an LNV convergence threshold of 10^{-5} Ha $a_0^{-3/2}$.The temperature of the auxiliary degrees of freedom was computed as $\mathrm{Tr}[\dot{\mathbf{X}}^2]/N$ and the temperature associated with the real degrees of freedom was calculated as $\mathrm{Tr}[\dot{\mathbf{K}}^2]/N$ (cf. comment following (26)). The arrow at 0.27 ps indicates the point where the initial guesses obtained from propagation become worse than the default initialization, negating the efficiency gains of the scheme. The arrow at 0.53 ps indicates the point where the guesses become so far from the converged values as to make the QM calculation unstable.

Recognizing this issue, Niklasson *et al.* proposed a modified Verlet integrator,[36] which breaks the time-reversal symmetry of the equations of motion of the auxiliary variables to a small degree through the addition of a dissipative-like term in the integration. Since this effect is introduced through the integration, rather than a physical term in the Lagrangian, it does not yield new equations of motion for the auxiliary degrees of freedom. Instead, the approach can be thought as being similar to Langevin-like dynamics for the degrees of freedom $\{\chi_{\mathbf{a}}\}$ with internal numerical error fluctuations and external, approximately energy conserving, dissipative forces f^{diss}.

In order to minimize the breaking of time-reversal symmetry, dissipative forces proportional to $\dot{\zeta}^{\mathrm{diss}}$ are avoided so that the time-reversal symmetry can be maintained to a chosen higher order. Hence, the modified Verlet algorithm to integrate (12b) for the dXL scheme is given by

$$\zeta_{\mathbf{a}}^{n+1} = 2\zeta_{\mathbf{a}}^{n} - \zeta_{\mathbf{a}}^{n-1} + \kappa(\chi_{\mathrm{SCF},\mathbf{a}} - \zeta_{\mathbf{a}})^n + \gamma \sum_{l=0}^{L} c_l \zeta_{\mathbf{a}}^{n-l}, \quad (19)$$

where we δt is the time step chosen to integrate the equations of motion and $\kappa = \delta t^2 \omega^2$. The coefficients γ and c_l are obtained in such a way that for a given L, all the odd-order terms in δt cancel out[36] up to order δt^{2L-3}. Fig. 2 illustrates the behavior of the dXL scheme, for a 64 H_2O-molecule AIMD simulation,

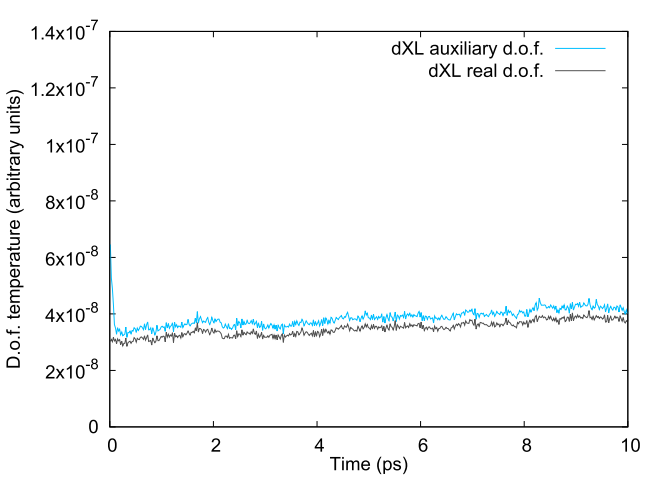


FIG. 2. Illustration of the stability of the dXL ($L = 7$) scheme. In the interest of clarity, averages over 10 fs are shown.

with the same approach for the calculation of temperatures as explained above for Fig. 1. The auxiliary degrees of freedom closely follow their real counterparts and the instability is removed.

Equation (19) is readily adapted to the classical framework with the usual substitution $\zeta_{\mathbf{a}} \to \boldsymbol{\mu}_I$,

$$\boldsymbol{\mu}_I^{n+1} = 2\boldsymbol{\mu}_I^{n} - \boldsymbol{\mu}_I^{n-1} + \kappa(\boldsymbol{\mu}_{\mathrm{SCF},I} - \boldsymbol{\mu}_I)^n + \gamma \sum_{l=0}^{L} c_l \boldsymbol{\mu}_I^{n-l}. \quad (20)$$

Here the dissipative force is given as a linear combination of previous values of the auxiliary dipoles up to some order L, with the optimal expansion coefficients c_l, and the overall scaling parameter γ given in Ref. 36.

The dXL can be analogously introduced for the linear-scaling DFT framework through a linear combination of previous auxiliary degrees of freedom,

$$\mathbf{X}^{n+1} = 2\mathbf{X}^{n} - \mathbf{X}^{n-1} + \kappa(\mathbf{K}_{\mathrm{SCF}}^{\perp} - \mathbf{X})^n + \gamma \sum_{l=0}^{L} c_l \mathbf{X}^{n-l}, \quad (21)$$

where the coefficients γ and c_l are the same as the ones used for the classical approach.[36]

V. EXTENDED LAGRANGIAN WITH THERMOSTAT-CONTROL (iXL)

The success of the dissipative scheme for a number of QM models has been reported in the literature.[36,38] Since the role of the dissipation is to counteract the numerical instabilities generated by the propagation scheme, for short time scales, a *bona fide* microcanonical dynamics can be generated. On the other hand, the main drawback of the scheme lies in the fact that it breaks time-reversibility (though to a high order) and therefore introduces a small, but measurable drift in the total energy. This can be ameliorated by carefully optimizing the coefficients of the expansion, but it cannot be removed completely. For long time scales, the steady drift of total energy is unavoidable, as first demonstrated in Ref. 37, which is consistent with our results obtained with classical polarizable force-field MD (Section VII A).

An alternative approach for overcoming the problem of breaking time-reversal symmetry has been proposed by Albaugh *et al.*[37] The main idea is to apply a simple thermostat to the velocities of the auxiliary variables, resulting in an *inertial* extended Lagrangian SCF formulation (iXL) *in lieu* of dissipation. Here the scheme will be illustrated using a general thermostat, γ, applied to the time-reversible velocity-Verlet integrator (12b),

$$\dot{\zeta}_{\mathbf{a}}^{n+\frac{1}{2}} = \dot{\zeta}_{\mathbf{a}}^{n} + \frac{1}{2}\delta t \omega^2 \ddot{\zeta}_{\mathbf{a}}^{n}, \quad (22)$$

$$\zeta_{\mathbf{a}}^{n+1} = \zeta_{\mathbf{a}}^{n} + \delta t \dot{\zeta}_{\mathbf{a}}^{n+\frac{1}{2}}, \quad (23)$$

$$\{\zeta_{\mathbf{a}}\}^{n+1} \to \mathtt{SCF} \to \{\chi_{\mathrm{SCF},\mathbf{a}}\}^{n+1} \to \mathtt{ENERGY\ AND\ FORCES}$$

$$\tilde{\dot{\zeta}}_{\mathbf{a}}^{n+1} = \dot{\zeta}_{\mathbf{a}}^{n+\frac{1}{2}} + \frac{1}{2}\delta t \omega^2 \ddot{\zeta}_{\mathbf{a}}^{n+1}, \quad (24)$$

$$\dot{\zeta}_{\mathbf{a}}^{n+1} = \gamma^{n+1} \tilde{\dot{\zeta}}_{\mathbf{a}}^{n+1}, \quad (25)$$

26 February 2024 09:53:54

MOST WIEDZY Downloaded from mostwiedzy.pl

where γ^{n+1} is the velocity scaling factor generated by a general thermostatting procedure at full time step update. For example, in the case of Berendsen velocity rescaling, γ^{n+1} is given by

$$\gamma^{n+1} = \sqrt{1 + \frac{\delta t}{\tau}\left(\frac{T^*}{T^{n+1}} - 1\right)}, \tag{26}$$

where T^{n+1} is the temperature of the auxiliary degrees of freedom at MD step $n+1$, T^* is the target (desired) value of this temperature, and τ is the characteristic time of the thermostat. T^* and τ are parameters of the scheme.

Although strictly speaking in the limit $m \to 0$, one can no longer define a kinetic energy, and therefore a temperature, for the auxiliary degrees of freedom $\zeta_{\mathbf{a}}$. Albaugh *et al.*[37] have suggested the ensemble average of the squared auxiliary velocities, i.e., $T = \langle \dot{\zeta}_{\mathbf{a}}^2 \rangle$ as an operational definition for the pseudo-temperature T. In this paper, we follow this convention, referring to the quantity in question simply as "temperature." The characteristic time τ is chosen similarly as in typical applications of thermostats—on the one hand, we want the decay rate of the temperature towards T^* to be much shorter than the length of the simulation t_{sim}: $\tau \ll t_{\text{sim}}$; on the other hand, we want to avoid a strong damping of instantaneous jumps in the temperature, and so $\tau \gg \delta t$. Provided τ satisfies the above constraints, the exact choice is expected to be inconsequential to the dynamics.

The desired auxiliary temperature T^* is chosen to approximately conform to the equipartition of energy consistent with a classical harmonic oscillator.[37] One possible way of obtaining T^* is by approximating the auxiliary velocity with the maximum displacement of the distribution of the real degrees of freedom.[37] One can also run a brief dXL calculation beforehand and subsequently set T^* to the time average of T^n obtained from that run. Another option is to simply compute the temperature of the real degrees of freedom, i.e., $\langle \dot{\chi}_{\text{SCF},\mathbf{a}} \rangle$ over a brief initialization period, and use a value slightly larger than that (since the real degrees of freedom are the minimum around which auxiliary degrees of freedom are meant to harmonically oscillate). A typical behavior of the iXL scheme is illustrated in Fig. 3.

The equations for the iXL approach for a polarizable force field are obtained from (22)–(25), with the substitutions $\boldsymbol{\mu} \to \zeta$ and $\dot{\boldsymbol{\mu}} \to \dot{\zeta}$. The coupling constant γ^n is given by

$$\gamma^n = \sqrt{1 + \frac{\delta t}{\tau}\left(\frac{T^*}{\langle \dot{\boldsymbol{\mu}}_I^2 \rangle^n} - 1\right)}. \tag{27}$$

For the sake of clarity, a simple Berendsen thermostat has been used to illustrate the iXL approach. However, any other (more efficient) thermostat can be used in principle, since the scope of a thermostat in this scheme is only to remove heat (numerical noise) from the auxiliary degrees of freedom. In fact, in all results reported later, we use 4th order Nosé-Hoover chains for thermostatting the auxiliary velocities, which we found to be marginally better than the Berendsen thermostat.

Analogously, for a linear-scaling DFT approach, we apply the following substitutions in (22)–(25): $K^{\perp\alpha\beta} \to \zeta$ and $\dot{K}^{\perp\alpha\beta} \to \dot{\zeta}$. The coupling constant is given by

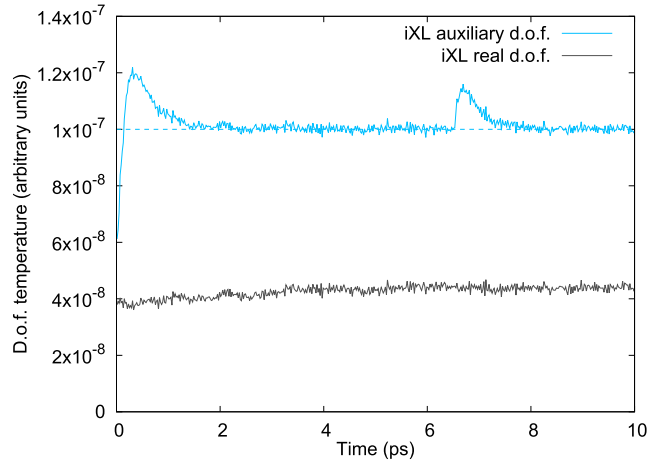


FIG. 3. Illustration of the behavior of the iXL scheme—the temperature of the auxiliary degrees of freedom quickly decays to the preset desired temperature T^* (shown as a dashed line). The two peaks correspond to instants where the number of SCF steps briefly flipped from 7 to 6, causing a temporary increase in the auxiliary temperature. Here $T^* = 10^{-7}$ and $\tau = 30$ fs. Testcase and the approach to calculation of temperatures: as in Fig. 1. In the interest of clarity, averages over 10 fs are shown.

$$\gamma^n = \sqrt{1 + \frac{\delta t}{\tau}\left(\frac{T^*}{\text{Tr}[\dot{\mathbf{X}}^2]^n} - 1\right)}. \tag{28}$$

VI. COMPUTATIONAL DETAILS

We studied the three extended Lagrangian schemes described in Secs. III–V the original extended Lagrangian scheme (extended Lagrangian) and its dissipative (dXL) and inertial (iXL) variants. Calculations not employing any propagation (starting the SCF procedure from scratch at every MD iteration) were used as baseline comparisons. We tested the extended Lagrangian schemes in two regimes—in classical MD calculations with the AMOEBA polarizable force field and in AIMD calculations with linear-scaling DFT. The former calculations were carried out using the TINKER[18] program, and the latter using ONETEP.[48,49]

As test systems we chose pure water-box systems with increasing numbers of water molecules: 16, 32, 64, 128 (and 512 with classical calculations), although all methods described should be generalizable to any molecular system. Liquid water is ubiquitous in biological systems and, as it is well-known, is a prototypical system for hydrogen bonding that influences its anomalous behavior throughout its phase diagram. Standard DFT GGA models struggle to correctly describe the structure and dynamics of water, with the nonlocality of dispersion interactions, deficiencies of local and semi-local exchange, presence of self-interaction error, and the neglect of quantum nuclear effects often cited as culprits.[69–71] At the same time, many successes and failures of both classical and quantum approaches for bulk water systems have been reported.[18,24,69–74] Consequently, bulk water systems provide an appropriate and stringent test for the extended Lagrangian methods where many-body effects are paramount.

For all classical polarizable force-field simulations, we used the water parameters of the AMOEBA14 water model.[74]

26 February 2024 09:53:54

MOST WIEDZY Downloaded from mostwiedzy.pl

The equations of motion for the nuclear degrees of freedom were integrated using the velocity Verlet integrator[75] with a time step of $\delta t = 1.0$ fs. Each system started with the water molecules arranged on a lattice in an equilibrium geometry with a cubic cell, whose volume corresponded to the reported density for the force field $\rho = 1.0003$ g cm^{-3}. The Particle-mesh Ewald method[76,77] with a 9 Å real-space cutoff for long-range electrostatics was employed. Equilibration simulations were carried out in the *NVT* ensemble at 298 K for 0.5 ns, with temperature controlled using a Nosé-Hoover thermostat[78] with a fourth-order chain and a characteristic time of $\tau = 0.1$ ps.

Following equilibration, we ran *NVE* production calculations for 1 ns with each of the above extended Lagrangian schemes, along with a baseline calculation where initial guesses were not propagated using the default conjugate gradient (CG) SCF method in TINKER, with a threshold of 10^{-6} D. For production calculations, we integrated the equation of motion (13b) using the time-reversible velocity Verlet integration for the original extended Lagrangian scheme and iXL, whereas a modified Verlet scheme was used for dXL, as described in Section IV. For the original extended Lagrangian scheme and for the iXL scheme, we set $\omega = \sqrt{2}/\delta t$ according to the criterion in Ref. 38. The target temperature T^* for the auxiliary degrees of freedom in the iXL scheme was estimated by approximating the square of the auxiliary dipole velocity $\langle \dot{\boldsymbol{\mu}}_I^2 \rangle$ with the maximum displacement of the distribution of the real dipoles.[37] For these systems, this gave a value of $T^* \approx 10^5\ e^2\text{Å}^2/\text{ps}^2$, which is the value used in this work. The Nosé-Hoover thermostat[78] with a fourth-order chain and a characteristic time of $\tau = 0.1$ ps was used for the auxiliary degrees of freedom. For all the extended Lagrangian schemes and for all the system sizes N, we ran our simulations with three different thresholds for the SCF optimization: 10^{-1} D (loose), 10^{-4} D (moderate), and 10^{-6} D (tight).

For linear-scaling DFT calculations, we used the pre-equilibrated systems obtained from the classical calculation to avoid lengthy equilibration in AIMD. These were subsequently further equilibrated with ONETEP for 1 ps with conventional BOMD (i.e., in the absence of a propagation scheme) in order for the systems to adjust to the switch from a classical to an *ab initio* Hamiltonian. The *NVT* ensemble was used, with temperature controlled via a Nosé-Hoover chain thermostat. No adjustments were made to the densities of the systems. The LNV convergence threshold was set to an RMS gradient of 10^{-6} Ha $a_0^{-3/2}$, which generally required 19-20 SCF steps to converge.

For both equilibration and production calculations, we used the BLYP exchange-correlation functional[79,80] with Grimme D2 dispersion correction[81] in order to improve the DFT description of water.[70] The kinetic energy cutoff was set to 900 eV, and norm-conserving pseudopotentials were employed. We used 8 a_0 as the localization radii of the NGWFs throughout, except for the 16 H_2O system, where the small size of the periodic box (14.78 a_0) forced us to use a slightly smaller localization radius of 7.35 a_0. Since we used fixed (non-*in situ*-optimized) localized orbitals, we chose a non-minimal double-zeta with polarization (DZ + P) basis set to improve the description. No density kernel truncation was applied, due to the systems' sizes being too small for the truncation to show any benefit. A velocity Verlet scheme was employed, with a time step of $\delta t = 0.5$ fs to integrate the nuclear degrees of freedom.

Production calculations were carried out with all of the extended Lagrangian schemes, and calculations with no propagation as baseline, using a selection of LNV convergence thresholds: 10^{-4} Ha $a_0^{-3/2}$ (loose), 10^{-5} Ha $a_0^{-3/2}$ (moderate), and 10^{-6} Ha $a_0^{-3/2}$ (tight). These sampled the *NVE* ensemble and were carried out for 10 ps. While a longer sampling would certainly be desired, the large computational cost of AIMD simulations precluded that. Analogously to the classical calculations, we employed a velocity Verlet scheme for the auxiliary degrees of freedom both for the original extended Lagrangian scheme and the iXL scheme, with a time step of 0.5 fs. For the dissipative dXL scheme, we used a modified Verlet scheme as explained in Sec. IV. The target temperature T^* for the iXL was set by running a brief dXL calculation, taking the time average of the auxiliary temperature, and using a slightly larger (more conservative) value for iXL, as otherwise the scheme's thermostat struggled to keep the desired temperature, leading to excessive drift. This more heuristic approach has the advantage of avoiding a long simulation to compute the distribution of the displacements of real electronic degrees of freedom, which can be quite computationally demanding for AIMD. The value we settled for was $T^* = 10^{-7}\ e^2/\text{fs}^2$ for all thresholds, except for the 128 H_2O system at the loose threshold, where a larger value of $T^* = 10^{-6}\ e^2/\text{fs}^2$ was necessary to maintain stability. Due to the short time scale of our AIMD simulations, we set $\tau = 0.03$ ps for the thermostat characteristic time, forcing the thermostat to work six times faster than its classical counterpart.

VII. RESULTS

A. Polarizable force fields

We will use the largest (512 H_2O) system as the main test-case for classical calculations, and all results will refer to this system, unless indicated otherwise. We begin by confirming the problems of the original extended Lagrangian scheme due to its quickly deteriorating quality of the propagated guesses. We laid out the origin of this undesired behavior in Sec. III. Fig. 4 shows how the number of SCF iterations increases over time, regardless of the assumed convergence threshold. Given sufficient time, the quality of the propagated guesses becomes worse than in the absence of propagation, negating the efficiency gains of the scheme, even if, in principle, this formulation is expected not to introduce a drift in the energy.

The dissipative scheme addresses the deficiency of the original formulation, at the price of weaker (finite-order) time-reversibility as outlined in Sec. IV. We will now estimate the typical energy drift that manifests as a consequence. Fig. 5 shows the change in energy per atom in an *NVE* simulation with the dXL scheme, for several selected values of the order parameter L (cf. (19)) and for calculations with a loose, moderate, and tight SCF threshold. For all three SCF convergence

26 February 2024 09:53:54

MOST WIEDZY Downloaded from mostwiedzy.pl

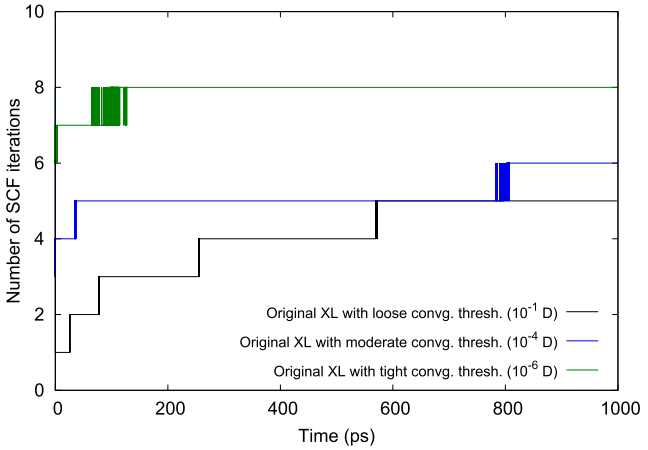


FIG. 4. Decaying quality of the initial guesses propagated through the original extended Lagrangian scheme, evidenced by the number of SCF steps needed to converge the induced dipoles (solid lines).

thresholds, the drift is substantial, with the loose threshold being the most severe case, corresponding to a drift rate of $\sim 2 \times 10^{-1}$ K/ps. While the energy drift rate gets smaller as the SCF tolerance increases, the drift is systematic and will become non-negligible in calculations spanning hundreds of nanoseconds.

The iXL scheme adopts yet another approach to overcome the issues of the original extended Lagrangian scheme. A thermostat is introduced to control the temperature of the auxiliary dipoles in order to avoid the accumulation of noise in the propagation, as described in Sec. V. Fig. 6 shows the change in energy per atom in an *NVE* simulation for this case. Energy conservation is maintained even at loose thresholds, with a drift rate of $\sim 3 \times 10^{-3}$ K/ps and requiring only three SCF iterations at the 10^{-1} D threshold (see inset in Fig. 6), making this scheme very competitive and suitable for performing long (μs-scale) *NVE* simulations. A modest price to pay for the iXL method, apart from the need to choose suitable parameters T^* and τ (cf. Sec. V), is its dependence on a definition of "pseudo-temperature" through the kinetic energy of the system, which may become less valid as the number of degrees of freedom in the system decreases. It is thus prudent to examine the behavior of the method for smaller systems to assess its transferability to other systems. Tables I and II summarize our findings for the smaller systems. In particular, in Table I, we compare the drift in the total energy for the unpropagated (conventional SCF-MD) scheme, which serves as reference, the dXL scheme, and the iXL scheme for four system sizes each with three SCF-CG thresholds. The reasons for large uncertainties in the drifts, even given a long simulation time, are as follows: for the loose threshold, the drift is so severe that over 1000 ps the system heats up (cools down) so much that the drift is no longer linear due to the massive increase (decrease) in kinetic energy. For the moderate and tight thresholds, the large uncertainty reflects the difficulty of accurately estimating extremely small (sub-mK/ps) drifts. The drifts and their uncertainties were calculated by assuming the drifts to be approximately linear. In a simulation with a length of t (t = 1000 ps), we can use a subset of data, viz., the interval $[0, t_0]$ to evaluate the linear coefficient in the drift, $a(t_0)$, over this interval. The final

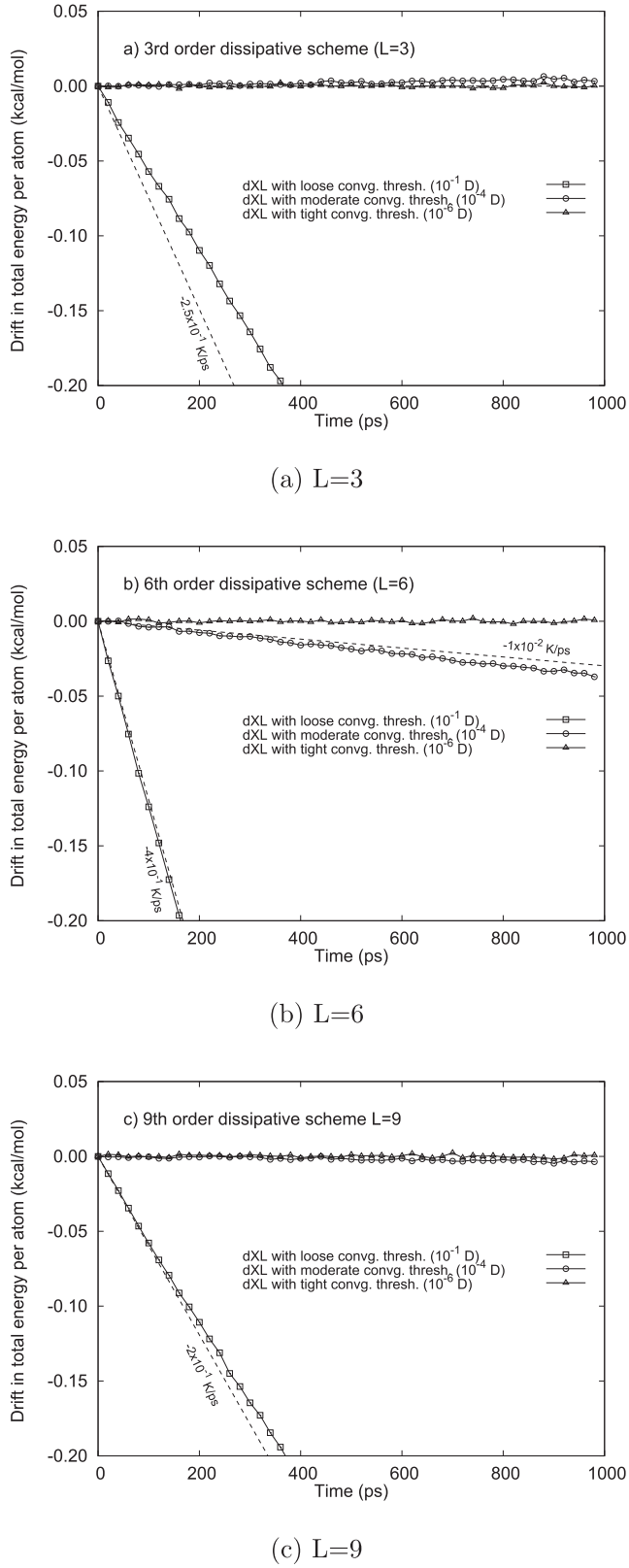


FIG. 5. Total energy conservation of dXL for different dissipation orders: L = 3 (panel (a)), L = 6 (panel (b)), and L = 9 (panel (c)). Different curves within each panel correspond to different SCF thresholds: loose, moderate, and tight. The dashed lines are meant as a guide to the eye and correspond to energy drifts expressed as system cooling/heating rates in K/ps. In the interest of clarity, only points 20 ps apart are shown.

drift estimate is $a(t)$. The uncertainty is taken as the largest difference between $a(t_0)$ and $a(t)$ calculated over $t_0 \geq t/2$.

26 February 2024 09:53:54

MOST WIEDZY Downloaded from mostwiedzy.pl

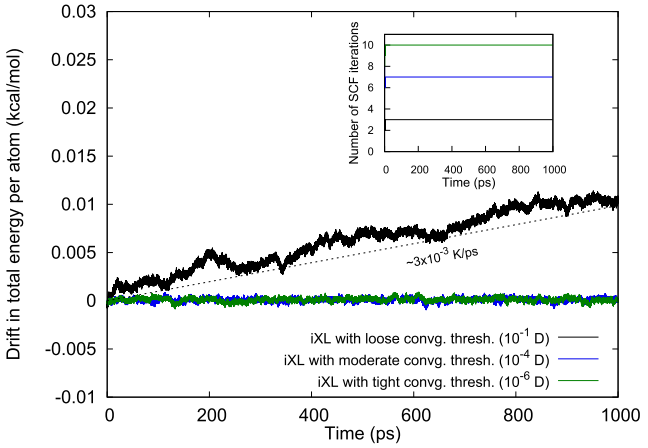


FIG. 6. Total energy conservation of iXL for different convergence thresholds (solid lines): loose, moderate, and tight. The dashed line is meant as a guide to the eye and corresponds to energy drift expressed as system heating rates in K/ps. In the interest of clarity, averages over 10 fs are shown.

TABLE II. A comparison of the number of CG-SCF iterations between calculations with no propagation and calculations using the dXL and iXL schemes—for four different system sizes (16, 32, 64, and 128 H_2O molecules) and three different CG-SCF thresholds.

CG-SCF threshold (D)	Propagation scheme	System size 16	32	64	128
10^{-1} (loose)	No prop.	1.00	1.00	1.00	1.00
	dXL	1.00	1.00	1.00	1.00
	iXL	3.00	3.00	3.00	3.00
10^{-4} (moderate)	No prop.	5.42	5.51	5.65	5.77
	dXL	3.62	3.65	3.83	3.79
	iXL	6.89	6.72	6.84	6.99
10^{-6} (tight)	No prop.	8.38	8.70	8.92	8.99
	dXL	6.68	6.85	6.96	6.99
	iXL	9.02	9.07	9.17	9.16

Table II presents an analogous analysis to Table I, with the number of CG-SCF iterations replacing the energy drift. Table I and Fig. 7 demonstrate that for loose and moderate CG-SCF thresholds, the iXL scheme outperforms all orders of the dXL scheme in terms of energy conservation, even for very small system sizes. In the tight threshold regime, the two schemes are equivalent. However, as shown in Table II, the trend is inverted when considering the average number of CG-SCF iterations. In fact, $\overline{N}_{SCF}$ for iXL is always larger than the corresponding $\overline{N}_{SCF}$ for dXL, regardless of the CG-SCF threshold. Notably, at moderate convergence thresholds, the dXL scheme shows energy drifts of the same order of magnitude (10^{-3} K/ps) as iXL with a loose threshold and also a comparable average number of CG-SCF iterations ($\overline{N}_{SCF} = 3.7$). Therefore, dXL and iXL display similar performance but in different SCF threshold windows: moderate for dXL and loose for iXL.

To assess how these integration schemes affect the actual dynamics, the oxygen-oxygen pair correlation function $g_{OO}(r)$ has been computed, see Fig. 8. Clearly, the red curve and cyan curve in Fig. 8, corresponding to dXL with a moderate threshold and iXL with a loose threshold, respectively, lie on top of each other and they are indistinguishable from the reference curve (solid gray). The latter has been obtained with the conventional (unpropagated) scheme and a tight threshold to provide a robust baseline. No artifacts are hence introduced in the dynamics by these two integration schemes, and they essentially provide the "correct" result as obtained from the reference calculation.

B. Linear-scaling DFT

As the main testcase for linear-scaling DFT calculations, we will use the 64 H_2O system, unless indicated otherwise. As we have done for the classical calculations, we begin by confirming the impracticality of the original extended Lagrangian scheme. Fig. 9 shows that for linear-scaling DFT calculations, the quality of the propagated guesses decays even more rapidly than for the classical calculations. Indeed, for the loose SCF convergence threshold, the original extended Lagrangian scheme becomes less efficient than the unpropagated scheme

TABLE I. Comparison of the energy drift (K/ps) between calculations with no propagation, and calculations using the dXL ($L = 6$) and iXL schemes—for four different system sizes (16, 32, 64, and 128 H_2O molecules) and three different CG-SCF convergence thresholds. $\overline{N}_{SCF}$ is the average number of SCF steps for a combination of method and threshold (detailed breakdown in Table II).

CG-SCF threshold (D)	Propagation scheme	System size 16 H_2O	32 H_2O	64 H_2O	128 H_2O	$\overline{N}_{SCF}$
		Drift (K/ps ($\times 10^{-4}$))				
10^{-1} (loose)	No prop.	48 752.2 ± 5969.3	49 642.4 ± 8490.7	40 048.8 ± 5110.5	38 739.4 ± 3827.8	1.0
	dXL	2395.9 ± 685.2	−3371.2 ± 663.7	−6003.1 ± 3986.1	−5613.2 ± 4632.9	1.0
	iXL	−7.9 ± 154.9	158.4 ± 104.7	84.8 ± 93.8	74.9 ± 18.7	3.0
10^{-4} (moderate)	No prop.	5.0 ± 7.4	−6.7 ± 9.4	−3.7 ± 1.7	1.6 ± 2.1	5.6
	dXL	−20.8 ± 9.2	−23.4 ± 4.6	−20.0 ± 1.7	−22.9 ± 3.2	3.7
	iXL	−0.9 ± 8.0	0.4 ± 2.6	−0.8 ± 1.0	−0.5 ± 29.6	6.9
10^{-6} (tight)	No prop.	5.9 ± 7.9	−0.8 ± 2.7	−0.4 ± 1.5	−0.6 ± 1.7	8.7
	dXL	1.4 ± 5.3	0.3 ± 4.0	0.0 ± 1.8	−0.6 ± 1.4	6.9
	iXL	3.5 ± 6.8	2.3 ± 2.1	0.8 ± 5.0	−0.7 ± 1.5	9.1

26 February 2024 09:53:54

MOST WIEDZY Downloaded from mostwiedzy.pl

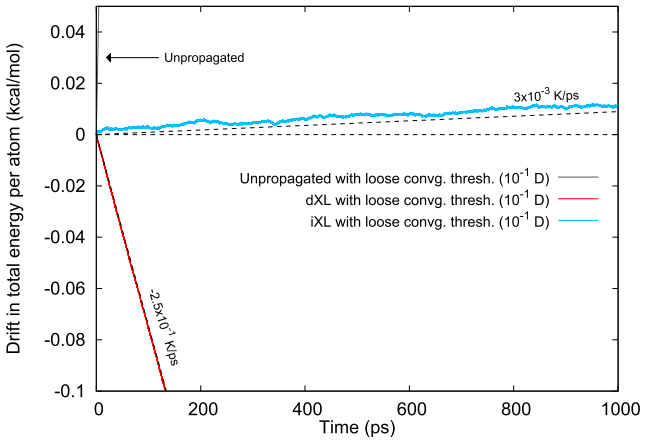


FIG. 7. Comparison of drifts in total energy between calculations with no propagation, dXL and iXL. SCF convergence was set to 10^{-1} D (loose) for all integration schemes. In the interest of clarity, averages over 10 fs are shown.

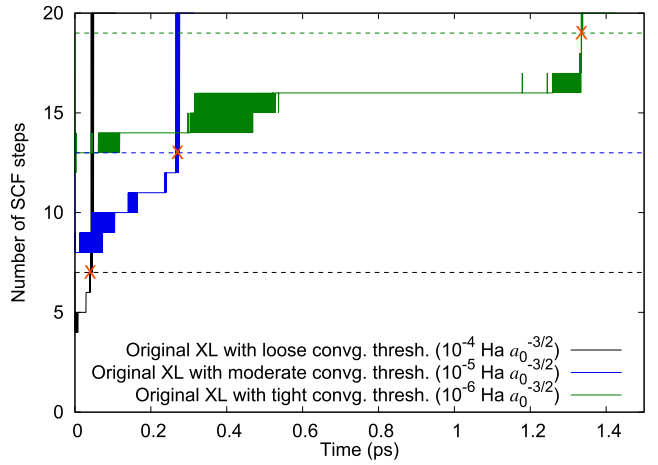


FIG. 9. Decaying quality of the initial guesses propagated through the original extended Lagrangian scheme, evidenced by the number of SCF steps needed to converge the induced dipoles (solid lines). Dashed lines denote the corresponding number of SCF steps in the absence of propagation. Crosses at 0.04 ps, 0.27 ps, and 1.33 ps denote points where the efficiency gain of extended Lagrangian over the unpropagated scheme is lost.

as early as after 0.04 ps (80 MD steps), probably reflecting the good quality of ONETEP's (unpropagated) initial guesses for the simpler scenario of fixed NGWFs.

A significant difference between classical MD calculations and their linear-scaling DFT counterparts is the much higher "noise floor" of *ab initio*-derived forces. This is a consequence of more intricate numerical machinery involved in DFT calculations (commonplace use of grid-based operations, such as Fast Fourier Transforms (FFTs), numerical integration of quantities on a variety of grids with up- and downsampling between grids, inexact translational and rotational invariance ("egg-box effect"), and use of polynomial interpolations in the handling of pseudopotentials), and the fact that the error in Kohn-Sham DFT forces is first order with respect to the error in the incompletely converged wavefunction, even though the error in the energy is second order.[2] While tighter SCF convergence can be imposed for dynamical calculations, this can quickly become impractical, as to reach those stricter convergence threshold grids must be made finer, and other approximations need to be well-controlled. In practice, the resultant noise in DFT forces, even though perfectly acceptable, e.g., for geometry optimization, leads to energy drifts in the order of 10^{-1} K/ps, while in classical MD, even when iterative schemes are involved, drifts are not expected to exceed 10^{-3} K/ps (cf. Table I).

Linear-scaling formulations of DFT necessarily add further approximations to conventional DFT, even if in robust approaches these approximations are controllable. For example, the use of finite-box FFTs leads to a slight delocalization of gradients beyond the localization regions of local orbitals, exchange-correlation energy is typically evaluated on a finite

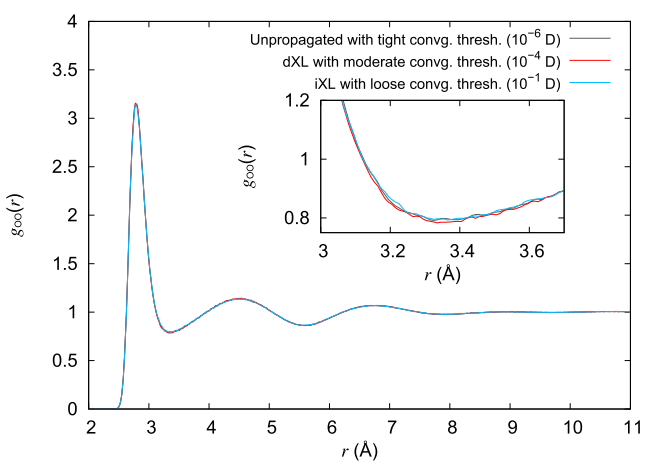


FIG. 8. Oxygen-oxygen pair correlation function $g_{OO}(r)$ obtained in the absence of a propagation scheme, and with the dXL and iXL schemes with different thresholds. Testcase: 512 H_2O molecule system. The reference unpropagated calculation uses a tight RMS threshold of 10^{-6} D. The SCF convergence threshold was set to RMS 10^{-1} D (loose) for the iXL scheme (solid cyan). The SCF convergence threshold was set to RMS 10^{-4} D (moderate) for the dXL scheme (solid red). The differences between the predictions obtained with dXL and iXL and in the absence of propagation are seen to be minor, as also highlighted in the inset.

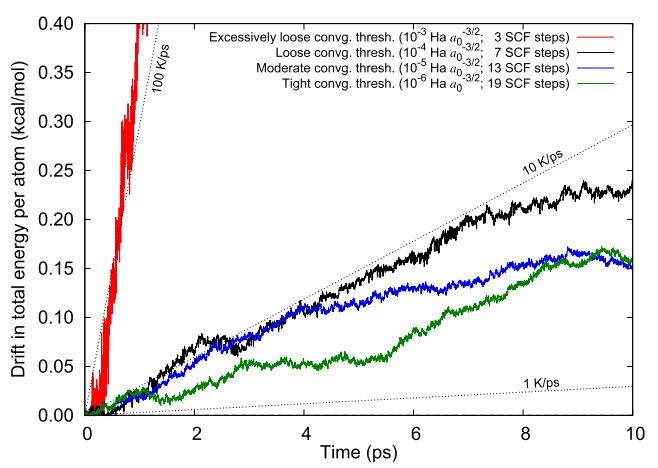


FIG. 10. Drift in the total energy for linear-scaling DFT for unpropagated BOMD (testcase: 64 H_2O molecules), for a selection of SCF convergence thresholds. Excessively loose thresholds lead to extreme drift, but inherent drift at ~5 K/ps is apparent regardless of how tight SCF convergence is. The dotted lines are meant as a guide to the eye and correspond to energy drifts expressed as system cooling/heating rates in K/ps.

26 February 2024 09:53:54

MOST WIEDZY Downloaded from mostwiedzy.pl

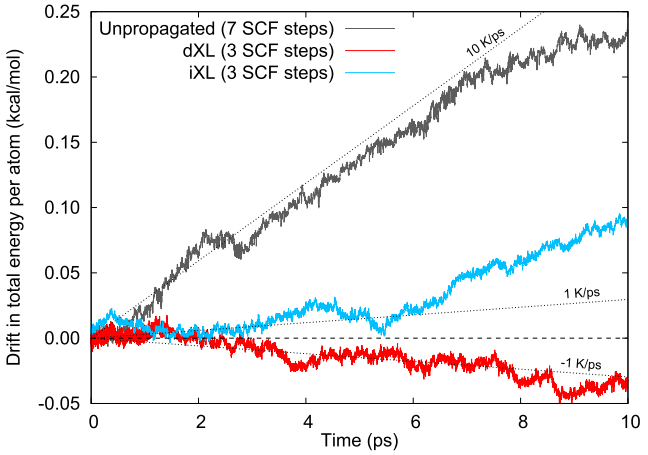


FIG. 11. Comparison of drifts in total energy between calculations with no propagation and dXL and iXL (testcase: 64 H_2O molecules). SCF convergence threshold was set to 10^{-4} (loose), as this is what would be used in practical simulations. The dotted lines are meant as a guide to the eye and correspond to energy drifts expressed as system cooling/heating rates in K/ps.

Cartesian grid, and Pulay-like corrections to forces are often not numerically exact for reasons of efficiency. Some of these approximations are common to all LS-DFT formulations, whereas others are specific to the different implementations. A consequence of these approximations is a further increase in the inaccuracy of forces, with residual errors in ONETEP typically in the order of 0.1%—this can be estimated from the magnitude of the net force, which, in the absence of noise, should be zero by Newton's 3rd law of motion. While approaches for correcting some of these approximations have recently been proposed,[29,31] the current state of the art necessitates using mild thermostatting to control for energy drifts in LS-DFT, as these typically result in temperature increase/decrease rates of several K/ps.

To wit, in Fig. 10, we show drifts obtained for the 64 H_2O system in the absence of any propagation scheme. For consistency, we adopt the same criterion to initialise the NGWFs as outlined in Sec. VI. While the drift in excess of 100 K/ps obtained for the LNV convergence threshold of 10^{-3} Ha $a_0^{-3/2}$ is clearly due to an excessively loose threshold, there is little improvement when the threshold is tightened from 10^{-4} Ha $a_0^{-3/2}$ to 10^{-5} Ha $a_0^{-3/2}$, and no further gain whatsoever with 10^{-6} Ha $a_0^{-3/2}$, which indicates the presence of an inherent drift that cannot be mitigated by improving the degree of SCF convergence. The magnitude of this drift is in the order of 5 K/ps, which leads us to expect that the drift due to time-reversal symmetry breaking in dXL (estimated between 10^{-3} and 10^{-4} K/ps in the classical case, cf. Table I, for moderate and tight thresholds) will be entirely obscured by the intrinsic drift due to forces.

Fig. 11 compares the energy drift of the unpropagated calculation and the two extended Lagrangian approaches for a loose SCF threshold (10^{-4} Ha $a_0^{-3/2}$). The drift in the absence of propagation is 8.5 ± 0.9 K/ps, with dXL and iXL performing better: at -1.5 ± 0.3 K/ps and 2.8 ± 2.1 K/ps, respectively. The fact that the drifts are comparable between dXL and iXL is in line with our expectations—the drift that makes dXL less desirable for long classical MD simulations does not play an appreciable role in LS-DFT MD, since it is dwarfed by inherent drift due to the noise in LS-DFT forces.

To add more weight to this argument, we examine the drifts for the remaining systems and for tighter SCF thresholds. The results are summarized in Table III. For all system sizes, regardless of how and if the initial guesses are propagated, we observe drifts of several K/ps. No correlation is apparent between the magnitude of the drift and the SCF convergence threshold, which indicates that the threshold is not excessively loose, and that no accuracy

TABLE III. Comparison of the energy drift (K/ps) between calculations with no propagation, and calculations using the dXL and iXL schemes—for four different system sizes (16, 32, 64, and 128 H_2O molecules) and three different SCF(LNV) convergence thresholds. Large uncertainties of computed drifts reflect necessarily short simulation times.[a] $\overline{N}_{SCF}$ is the average number of SCF steps for a combination of method and threshold (detailed breakdown in Table IV).

		System size				
		16 H_2O	32 H_2O	64 H_2O	128 H_2O	
SCF threshold (Ha $a_0^{-3/2}$)	Propagation scheme	Drift (K/ps [$\times 10^0$])				$\overline{N}_{SCF}$
10^{-4} (loose)	No prop.	4.8 ± 0.4	4.2 ± 1.9	8.5 ± 0.9	3.3 ± 1.1	7.8
	dXL	3.5 ± 1.8	6.6 ± 0.5	−1.5 ± 0.3	0.2 ± 0.2	3.3
	iXL	1.5 ± 0.8	−3.9 ± 2.1	2.8 ± 2.1	−2.4 ± 0.4	4.0
10^{-5} (moderate)	No prop.	3.8 ± 5.3	2.3 ± 2.1	5.6 ± 3.2	3.0 ± 0.5	13.5
	dXL	2.9 ± 0.8	0.2 ± 3.0	1.9 ± 2.9	3.4 ± 0.2	5.7
	iXL	4.8 ± 0.7	3.8 ± 1.5	0.9 ± 0.8	2.6 ± 0.3	7.7
10^{-6} (tight)	No prop.	3.2 ± 2.7	5.7 ± 1.6	5.8 ± 2.2	2.5 ± 0.9	19.3
	dXL	2.2 ± 1.4	4.6 ± 1.9	1.5 ± 0.4	3.1 ± 0.7	10.2
	iXL	2.8 ± 2.5	7.1 ± 1.5	5.4 ± 0.7	2.8 ± 0.4	13.5

[a]The drifts and their uncertainties were calculated as follows. We assume the drift to be approximately linear. In a simulation with a length of t ($t = 10$ ps), we can use a subset of data, viz., the interval $[0, t_0]$ to evaluate the linear coefficient in the drift, $a(t_0)$, over this interval. The final drift estimate is $a(t)$. The uncertainty is taken as the largest difference between $a(t_0)$ and $a(t)$ calculated over $t_0 \geq t/2$.

26 February 2024 09:53:54

MOST WIEDZY Downloaded from mostwiedzy.pl

TABLE IV. A comparison of the number of SCF iterations between calculations with no propagation, and calculations using the dXL and iXL schemes—for four different system sizes (16, 32, 64, and 128 H_2O molecules) and three different SCF thresholds.

SCF threshold ($\mathrm{Ha}\,a_0^{-3/2}$)	Propagation scheme	System size 16	32	64	128
10^{-4} (loose)	No prop.	9.0	8.0	7.1	7.0
	dXL	3.8	3.4	3.1	3.0
	iXL	4.0	4.0	3.0	5.0
10^{-5} (moderate)	No prop.	14.0	14.0	13.0	13.0
	dXL	6.3	6.0	5.5	5.0
	iXL	9.0	8.0	7.0	7.0
10^{-6} (tight)	No prop.	20.0	20.0	19.1	18.0
	dXL	11.0	10.8	10.0	9.0
	iXL	14.4	13.9	13.7	12.0

gains can be achieved by using tighter thresholds. Neither of the extended Lagrangian methods is seen to discernibly outperform the other, and the drifts are comparable to the case with no propagation. This confirms that the observed drifts are a result of inherent noise in LS-DFT forces and are hardly affected by the properties of an extended Lagrangian scheme.

Of course, the use of an extended Lagrangian approach is expected to yield a performance improvement by reducing the number of SCF steps needed to converge to a given threshold. We report that number in Table IV for all systems and methods under study. Both dXL and iXL offer a performance gain of 30%–60%, depending on the convergence thresholds, with looser thresholds offering larger gains. Except for an isolated case (64 H_2O molecules, loose convergence threshold), the use of the dissipative scheme leads to slightly faster convergence compared to the inertial scheme. This is a consequence of manually setting the auxiliary temperature for iXL

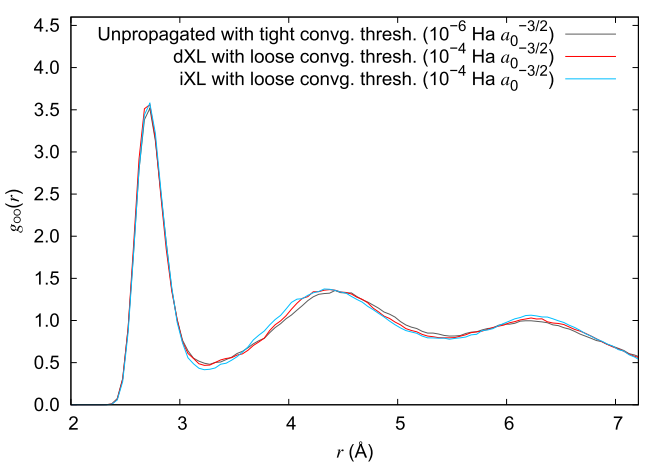


FIG. 12. Oxygen-oxygen pair correlation function $g_{OO}(r)$ obtained in the absence of a propagation scheme and with the dXL and iXL schemes. Test-case: 64 H_2O molecule system. For calculations with propagation, the SCF convergence threshold was set to 10^{-4} Ha $a_0^{-3/2}$ (loose), as this is what would be used in practical simulations. The reference (unpropagated) calculation uses a tight threshold (10^{-6} Ha $a_0^{-3/2}$). The differences between the predictions obtained with dXL and iXL and in the absence of propagation are seen to be minor.

to a slightly more conservative value than the exact average auxiliary temperature of dXL (cf. Sec. VI). With more aggressive settings, we would expect the performance of iXL to be indistinguishable from that of dXL, but with the caveat that iXL immediately starts to drift appreciably when T^* is set to an excessively low value, necessitating care in the choice of this parameter. The drift in this case is a consequence of the thermostat's persistent scaling of velocities, with the time average of γ^n (Eq. (28)) being excessively lower than 1.

Apart from not impairing energy conservation and improving performance, a natural requirement for a propagation scheme is for it not to visibly affect the calculated properties of the system. In this study, we examine the basic structural characteristic of water, the oxygen-oxygen pair correlation function $g_{OO}(r)$, shown in Fig. 12. The reference (unpropagated) calculation was run with a tight threshold, to provide a robust baseline that we assume is the correct result. The calculations with dXL and iXL were performed with a loose threshold, as this is what would be used in practice. The differences between the predictions of the two integration schemes are minuscule, and the differences between the propagated calculations and the reference are also very modest—the first peak is described practically identically, and for larger distances, the introduction of propagation (in particular with the iXL scheme) appears to introduce a very slight over-structuring, which we find perfectly acceptable given the time scale of the simulations. We do not include experimental results in the comparison, knowing full well that the DFT model of water significantly over-structures for reasons that we do not expect the propagation schemes to compensate for. The same over-structuring exacerbates the problem (the system takes longer to fully equilibrate). We thus attribute the small differences to the noise stemming from the limited time scale.

VIII. CONCLUSIONS

In this work, we assessed the performance of two integrators for the extended Lagrangian introduced by Niklasson *et al.*[36]—a dissipative formulation (dXL)[36] and an inertial formulation (iXL),[37] in two distinct regimes, by employing them in classical and in LS-DFT *NVE* MD calculations on condensed water systems. We confirmed the previously reported[37,38,65] necessity of counteracting the unbounded increase in the kinetic energy of the auxiliary degrees of freedom through some form of energy leaching, to which the two schemes take different approaches.

In the classical polarizable force field regime, we reproduced the observations of Albaugh *et al.*,[37] showing that over long (~ns) time scales, for maximally loose SCF thresholds, where an unpropagated scheme drifts catastrophically, iXL offers better energy conservation compared to dXL. This advantage is a consequence of a different approach to "energy leaching" taken by the iXL method, which strictly preserves time reversibility. However, at this loose SCF threshold, dXL has the advantage of needing substantially fewer SCF steps (1 vs. 3) to converge and thus offers an efficiency gain over iXL. As the SCF threshold is made progressively tighter, the performance of the two schemes begins to

26 February 2024 09:53:54

MOST WIEDZY Downloaded from mostwiedzy.pl

converge, with the drift of dXL improving considerably, and the almost constant absolute efficiency advantage of 2-3 SCF steps that dXL maintains over iXL becoming relatively less important.

In the LS-DFT MD regime, the picture is substantially different due to the high intrinsic noise of LS-DFT forces, which leads to energy drifts that are much larger than the additional drift due to energy leaching from the extended Lagrangian approach. As such, the dXL and iXL propagation schemes can be used with the loosest of LNV-SCF threshold investigated (10^{-4} Ha $a_0^{-3/2}$) beyond which an additional drift of an extended Lagrangian scheme starts to become apparent. Although there is no discernible difference in practice between the dXL and iXL schemes at the loose convergence threshold, the dXL would currently be the mildly preferred choice for LS-DFT MD calculations due to its parameter-free nature.

As it is now, we believe that in practical and sufficiently long (many-ps) LS-DFT MD calculations, a gentle thermostatting would currently be necessary to counteract the excessive increase/decrease in temperature that is a consequence of intrinsic drift, and which is as large as several K/ps. Further work is desirable in the area of linear-scaling DFT to improve the accuracy of forces in the presence of artifacts that result from localization constraints and the approximate treatment of Pulay-like terms in the forces. We feel obliged to point out that the energy drifts in LS-DFT would likely change noticeably if the localized were to be optimized (not studied in this work). This is because on the one hand the *in situ* optimization significantly reduces the magnitude of Pulay-like terms, while on the other hand it introduces a further approximation into their calculation. It is possible that the change that the optimization of localized orbitals would introduce to the "noise background" of LS-DFT forces, and the ensuing change in intrinsic energy drift, would uncover differences in the behavior of the two extended Lagrangian integration schemes that are currently obscured in *ab initio* MD. We intend to study this behavior in the future.

We highlighted the non-triviality of correctly evolving the auxiliary degrees of freedom over a curved manifold in *ab initio* calculations, a fact that is not always appreciated in the literature. We presented and tested a viable scheme for propagating the density kernel in this scenario, using a fixed, non-orthogonal generalized Wannier function basis. Further work is necessary to develop a scheme where the localized orbitals could be similarly propagated.

Finally, we showed that both dXL and iXL consistently enable significant, and usually similar, increases in performance over calculations not employing propagation, both for classical MD and LS-DFT MD. Thus, both of these schemes constitute important algorithmic improvements that markedly extend the time scales accessible to classical and LS-DFT MD simulations alike.

ACKNOWLEDGMENTS

V.V. acknowledges the support of EPSRC the Doctoral Training Centre of the Institute for Complex System Simulations (Grant No. EP/G03690X/1). J.D. and C.K.S. acknowledge the support of the Engineering and Physical Sciences Research Council (EPSRC Grant Nos. EP/J015059/1 and EP/K039156/1) and of the UKCP consortium (EPSRC Grant No. EP/K013556/1) for access to the ARCHER national supercomputer and the IRIDIS High Performance Computing Facility of the University of Southampton. A.A. and T.H.G. acknowledge the support by Grant No. CHE-1363320 from the U.S. National Science Foundation. A.M.N.N. acknowledges support from the Department of Energy Offices of Basic Energy Sciences (Grant No. LANL2014E8AN). Most of the calculations in this work were carried out at the TASK Computer Centre in Gdańsk, Poland (supported in part by the PL-Grid Infrastructure, Grant No. POIG.02.03.00-00-096/10).

APPENDIX: STABILITY OF VELOCITY VERLET WITH WEAK COUPLING

In Section V, we have derived the equations of motion for the auxiliary degrees of freedom $\{\zeta\}$, starting from the extended Lagrangian (10) and coupling $\{\zeta\}$ to an external bath,

$$\dot{\zeta}^{n+\frac{1}{2}} = \dot{\zeta}^{n} + \frac{\Delta t}{2}\ddot{\zeta}^{n} = \dot{\zeta}^{n} + \frac{\omega^2 \Delta t}{2}[\chi^{n}_{\mathrm{SCF}} - \zeta^{n}], \tag{A1}$$

$$\zeta^{n+1} = \zeta^{n} + \dot{\zeta}^{n+\frac{1}{2}}\Delta t, \tag{A2}$$

$$\mathtt{SCF} \rightarrow \mathtt{ENERGY\ AND\ FORCES}$$

$$\dot{\tilde{\zeta}}^{n+1}_{\mathbf{a}} = \dot{\zeta}^{n+\frac{1}{2}} + \frac{\Delta t}{2}\ddot{\zeta}^{n+1} = \dot{\zeta}^{n+\frac{1}{2}} + \frac{\omega^2 \Delta t}{2}[\chi^{n+1}_{\mathrm{SCF}} - \zeta^{n+1}], \tag{A3}$$

$$\dot{\zeta}^{n+1} = \alpha^{n+1}\dot{\tilde{\zeta}}^{n+1}, \tag{A4}$$

where $\zeta, \dot{\zeta}$, and $\ddot{\zeta}$ represent the auxiliary generalized coordinates, velocities, and accelerations, respectively, whereas χ_{SCF} is the converged SCF solution (real degrees of freedom). The coupling strength with the external bath is given by α. In deriving (A1)–(A4), we have adopted the velocity Verlet integrator with a time step Δt. A possible choice for α is given by the Berendsen thermostat,

$$\alpha^{n} = \sqrt{1.0 + \frac{\Delta t}{\tau}\left(\frac{T^{*}}{T^{n}} - 1.0\right)}, \tag{A5}$$

where τ represents the characteristic time scale of the thermostat. T^{*} is the target temperature. The instantaneous temperature T^{n} is generally a quadratic function of the auxiliary generalized velocities,

$$T \propto f(\dot{\zeta}^2). \tag{A6}$$

The discretized equations of motion (A1)–(A4) can be cast in matrix form as

$$\begin{bmatrix} \zeta^{n+1} \\ \dot{\zeta}^{n+1} \end{bmatrix} = \mathbf{T}(\alpha, \chi_{\mathrm{SCF}}, \Delta t) \begin{bmatrix} \zeta^{n} \\ \dot{\zeta}^{n} \end{bmatrix}. \tag{A7}$$

Following a change of variable $\dot{\zeta}\Delta t \rightarrow \dot{\zeta}$, the propagation matrix $\mathbf{T}$ takes the form

26 February 2024 09:53:54

MOST WIEDZY Downloaded from mostwiedzy.pl

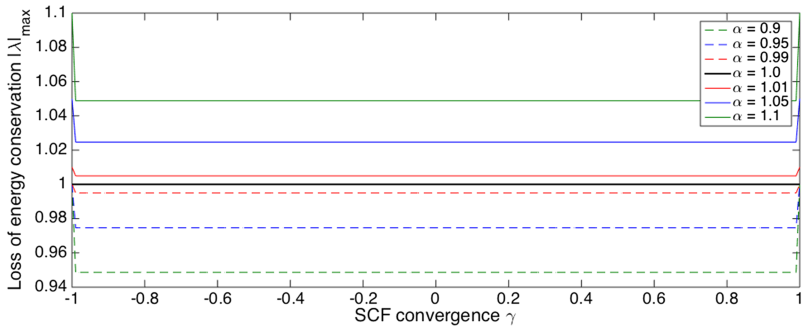
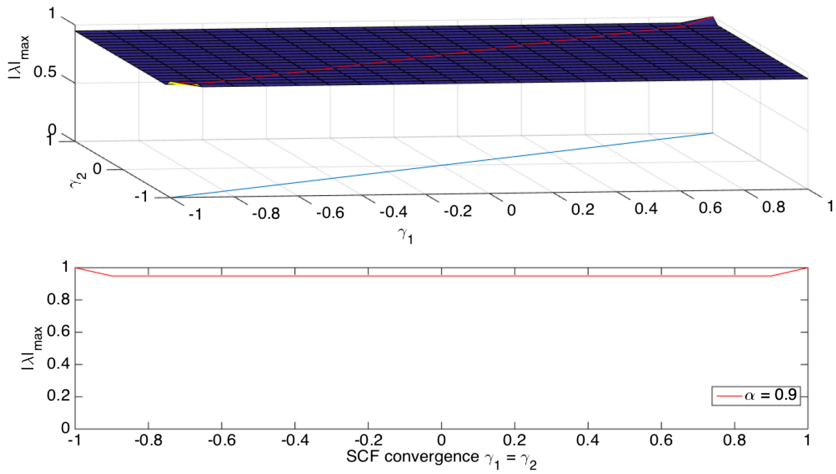
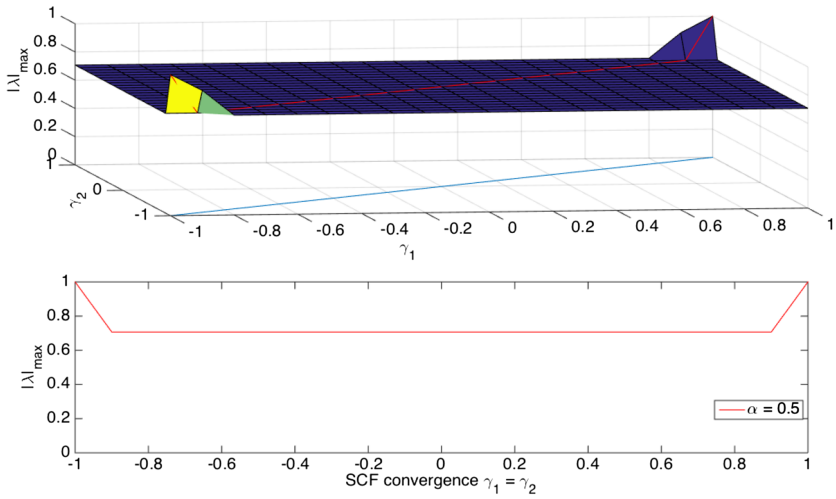


FIG. 13. Stability measured by the largest eigenvalue $|\lambda|_{\max}$ of (A10) as a function of the degree of SCF convergence measured by $\gamma \in [-1, 1]$ for selected values of α for the velocity Verlet scheme.



(a) $\alpha = 0.9$



(b) $\alpha = 0.5$

FIG. 14. Stability measured by the largest eigenvalue $|\lambda|_{\max}$ of (A11) as a function of the degree of SCF convergence when $\gamma^{n+1} \neq \gamma^n$ and they are independently allowed to vary in the range $[-1, 1]$ for $\alpha = 0.9$ panel (a), and $\alpha = 0.5$ panel (b) for the velocity Verlet scheme.

$$
\begin{aligned}
\mathbf{T} &= \begin{bmatrix} 1 & 0 \\ \frac{\alpha\omega^2\Delta t^2}{2}[\chi_{\mathrm{SCF}}^{n+1}\left(\zeta^{-1}\right)^{n+1} - 1] & \alpha \end{bmatrix} \times \begin{bmatrix} 1 + \frac{\omega^2\Delta t^2}{2}[\chi_{\mathrm{SCF}}^{n}\left(\zeta^{-1}\right)^{n} - 1] & 1 \\ 0 & 1 \end{bmatrix} \\
&\quad + \begin{bmatrix} 0 & 0 \\ \frac{\alpha\omega^2\Delta t^2}{2}[\chi_{\mathrm{SCF}}^{n}\left(\zeta^{-1}\right)^{n} - 1] & 0 \end{bmatrix} \\
&= \begin{bmatrix} 1 + \frac{\omega^2\Delta t^2}{2}Q^n & 1 \\ \frac{\alpha\omega^2\Delta t^2}{2}Q^{n+1} + \frac{\alpha\omega^4\Delta t^4}{4}Q^{n+1}Q^n + \frac{\alpha\omega^2\Delta t^2}{2}Q^n & \frac{\alpha\omega^2\Delta t^2}{2}Q^{n+1} + \alpha \end{bmatrix} \\
&= \begin{bmatrix} 1 + \kappa Q^n & 1 \\ \alpha\kappa\{Q^{n+1} + Q^n + \kappa Q^{n+1}Q^n\} & \alpha(\kappa Q^{n+1} + 1) \end{bmatrix},
\end{aligned}
\tag{A8}
$$

26 February 2024 09:53:54

MOST WIEDZY Downloaded from mostwiedzy.pl

where in the last two steps we have set $Q = \chi_{\mathrm{SCF}}\zeta^{-1} - 1$ and we have grouped together all scalar parameters, but α, into $\kappa = \omega^2\Delta t^2/2$. Following the analysis in Ref. 82, we can study the stability of the propagation by linearizing the SCF optimization around an exact ground state χ_*,

$$\chi_{\mathrm{SCF}} = \mathrm{SCF}[\zeta] \approx \chi_* + \Gamma(\zeta - \chi_*), \tag{A9}$$

where the linearized SCF optimization kernel is represented by Γ. Now we look at the homogeneous equation, i.e., $\chi_* \equiv 0$, and replace Γ with its largest eigenvalue γ, and insert this equation into the propagation matrix $\mathbf{T}$, which reads

$$\mathbf{T} = \begin{bmatrix} 1 + \kappa(\gamma^n - 1) & 1 \\ \alpha\kappa\{(\gamma^{n+1} - 1) + (\gamma^n - 1) + \kappa(\gamma^{n+1} - 1)(\gamma^n - 1)\} & \alpha[\kappa(\gamma^{n+1} - 1) + 1] \end{bmatrix}, \tag{A10}$$

where we want to ensure the propagation to be stable for the entire range of convergence, $\gamma \in [-1, 1]$.

The stability of the integrator is determined by the largest eigenvalue $|\lambda_{\max}|$ of the propagation matrix in (A10), which is a measure of the loss of energy conservation. To simplify the discussion, we can look at

$$\gamma^{n+1} \approx \gamma^n = \widetilde{\gamma},$$

which is achieved when two subsequent SCF steps converge with the same accuracy. In this case, the propagation matrix $\mathbf{T}$ simplifies to

$$\mathbf{T} = \begin{bmatrix} 1 + \kappa(\widetilde{\gamma} - 1) & 1 \\ \alpha\kappa[2(\widetilde{\gamma} - 1) + \kappa(\widetilde{\gamma} - 1)^2] & \alpha[\kappa(\widetilde{\gamma} - 1) + 1] \end{bmatrix}. \tag{A11}$$

This simplification captures most of the results for the stability as demonstrated by comparing Fig. 13 with Figs. 14(a) and 14(b), where we let both γ^n and γ^{n+1} to independently vary in the range $[-1, 1]$, with no significant impact on the stability. This implies that the stability of the integrator is mainly determined by the coupling parameter α. In fact, the propagation matrix in (A11) shows the important symplectic property,

$$\mathbf{T}^{\mathrm{T}}\mathbf{J}\mathbf{T} = \alpha\mathbf{J}, \tag{A12}$$

where $\mathbf{J}$ is the symplectic structure matrix

$$\mathbf{J} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{A13}$$

This is simply a re-statement of the fact that when $\alpha = 1$, we recover the canonical velocity Verlet scheme, which is of course symplectic. Our analysis is in line, at least qualitatively, with the results obtained by Albaugh *et al.* for the simpler Verlet scheme.[37] We also find that any $\kappa \neq 1$ results in a suboptimal regime, cf. Table V, which means that the optimal choice for ω ought to be $\sqrt{2}/\Delta t$.

TABLE V. Different values for the α parameter and the corresponding optimal values for κ and maximal dissipation.

Method/coefficients	α	κ_{opt}	$\lvert\lambda\rvert_{\max}$
	1.0	1.0	1.0
	0.99	1.0	0.995
	0.95	1.0	0.975
Velocity verlet	0.9	1.0	0.949
	1.01	1.0	1.005
	1.05	1.0	1.025
	1.1	1.0	1.049

[1] M. Karplus and J. A. McCammon, Nat. Struct. Biol. **9**, 646 (2002).
[2] M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos, Rev. Mod. Phys. **64**, 1045 (1992).
[3] N. Bernstein, J. R. Kermode, and G. Csányi, Rep. Prog. Phys. **72**, 026501 (2009).
[4] S. Goedecker, Rev. Mod. Phys. **71**, 1085 (1999).
[5] D. R. Bowler, T. Miyazaki, and M. J. Gillan, J. Phys.: Condens. Matter **14**, 2781 (2002).
[6] S. W. Rick and S. J. Stuart, "Potentials and algorithms for incorporating polarizability in computer simulations," in *Reviews in Computational Chemistry* (John Wiley and Sons, Inc., 2003), pp. 89–146.
[7] P. Cieplak, F.-Y. Dupradeau, Y. Duan, and J. Wang, J. Phys.: Condens. Matter **21**, 333102 (2009).
[8] M. Elstner and G. Seifert, Philos. Trans. R. Soc., A **372**, 20120483 (2014).
[9] N. D. Yilmazer and M. Korth, Comput. Struct. Biotechnol. J. **13**, 169 (2015).
[10] W. L. Jorgensen, J. Chem. Theory Comput. **3**, 1877 (2007).
[11] O. Demerdash, E.-H. Yap, and T. Head-Gordon, Annu. Rev. Phys. Chem. **65**, 149 (2014).
[12] G. Lamoureux, A. D. MacKerell, and B. Roux, J. Chem. Phys. **119**, 5185 (2003).
[13] D. P. Geerke and W. F. van Gunsteren, J. Phys. Chem. B **111**, 6425 (2007).
[14] A. C. T. van Duin, S. Dasgupta, F. Lorant, and W. A. Goddard, J. Phys. Chem. A **105**, 9396 (2001).
[15] D.-X. Zhao, C. Liu, F.-F. Wang, C.-Y. Yu, L.-D. Gong, S.-B. Liu, and Z.-Z. Yang, J. Chem. Theory Comput. **6**, 795 (2010).
[16] G. A. Kaminski, H. A. Stern, B. J. Berne, R. A. Friesner, Y. X. Cao, R. B. Murphy, R. Zhou, and T. A. Halgren, J. Comput. Chem. **23**, 1515 (2002).
[17] P. Ren and J. W. Ponder, J. Comput. Chem. **23**, 1497 (2002).
[18] P. Ren and J. W. Ponder, J. Phys. Chem. B **107**, 5933 (2003).
[19] P. Ren, C. Wu, and J. W. Ponder, J. Chem. Theory Comput. **7**, 3143 (2011).
[20] J. W. Ponder, C. Wu, P. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio, M. Head-Gordon, G. N. I. Clark, M. E. Johnson, and T. Head-Gordon, J. Phys. Chem. B **114**, 2549 (2010).
[21] P. Cieplak, J. Caldwell, and P. Kollman, J. Comput. Chem. **22**, 1048 (2001).
[22] P. N. Day, J. H. Jensen, M. S. Gordon, S. P. Webb, W. J. Stevens, M. Krauss, D. Garmer, H. Basch, and D. Cohen, J. Chem. Phys. **105**, 1968 (1996).
[23] A. Holt, J. Boström, G. Karlström, and R. Lindh, J. Comput. Chem. **31**, 1583 (2010).
[24] L.-P. Wang, T. Head-Gordon, J. W. Ponder, P. Ren, J. D. Chodera, P. K. Eastman, T. J. Martinez, and V. S. Pande, J. Phys. Chem. B **117**, 9956 (2013).
[25] A. C. Simmonett, F. C. Pickard, Y. Shao, T. E. Cheatham, and B. R. Brooks, J. Chem. Phys. **143**, 074115 (2015).
[26] D. Porezag, T. Frauenheim, T. Köhler, G. Seifert, and R. Kaschner, Phys. Rev. B **51**, 12947 (1995).
[27] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, and G. Seifert, Phys. Rev. B **58**, 7260 (1998).
[28] P. Pulay, Mol. Phys. **17**, 197 (1969).
[29] A. Ruiz-Serrano, N. D. M. Hine, and C.-K. Skylaris, J. Chem. Phys. **136**, 234101 (2012).
[30] M. V. Fernández-Serra and E. Artacho, J. Chem. Phys. **121**, 11136 (2004).
[31] S. Mohr, L. E. Ratcliff, P. Boulanger, L. Genovese, D. Caliste, T. Deutsch, and S. Goedecker, J. Chem. Phys. **140**, 204110 (2014).
[32] T. A. Arias, M. C. Payne, and J. D. Joannopoulos, Phys. Rev. Lett. **69**, 1077 (1992).
[33] D. K. Remler and P. A. Madden, Mol. Phys. **70**, 921 (1990).
[34] P. Pulay and G. Fogarasi, Chem. Phys. Lett. **386**, 272 (2004).

26 February 2024 09:53:54

MOST WIEDZY Downloaded from mostwiedzy.pl

[35]J. M. Herbert and M. Head-Gordon, Phys. Chem. Chem. Phys. **7**, 3269 (2005).
[36]A. M. N. Niklasson, P. Steneteg, A. Odell, N. Bock, M. Challacombe, C. J. Tymczak, E. Holmstrm, G. Zheng, and V. Weber, J. Chem. Phys. **130**, 214109 (2009).
[37]A. Albaugh, O. Demerdash, and T. Head-Gordon, J. Chem. Phys. **143**, 174104 (2015).
[38]A. M. N. Niklasson, Phys. Rev. Lett. **100**, 123004 (2008).
[39]Y. Shi, Z. Xia, J. Zhang, R. Best, C. Wu, J. W. Ponder, and P. Ren, J. Chem. Theory Comput. **9**, 4046 (2013).
[40]R. T. Bradshaw and J. W. Essex, J. Chem. Theory Comput. **12**, 3871 (2016).
[41]L.-P. Wang, J. Chen, and T. Van Voorhis, J. Chem. Theory Comput. **9**, 452 (2013).
[42]L.-P. Wang, T. J. Martinez, and V. S. Pande, J. Phys. Chem. Lett. **5**, 1885 (2014).
[43]P. T. van Duijnen and M. Swart, J. Phys. Chem. A **102**, 2399 (1998).
[44]*Iterative Solution of Large Linear Systems*, edited by D. M. Young (Academic Press, 1971).
[45]W. Wang and R. D. Skeel, J. Chem. Phys. **123**, 164107 (2005).
[46]J. Dziedzic, Y. Mao, Y. Shao, J. Ponder, T. Head-Gordon, M. Head-Gordon, and C.-K. Skylaris, J. Chem. Phys. **145**, 124106 (2016).
[47]W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[48]P. D. Haynes, A. A. Mostof, C.-K. Skylaris, and M. C. Payne, J. Phys.: Conf. Ser. **26**, 143 (2006).
[49]C.-K. Skylaris, P. D. Haynes, A. A. Mostofi, and M. C. Payne, J. Chem. Phys. **122**, 084119 (2005).
[50]R. W. Nunes and D. Vanderbilt, Phys. Rev. B **50**, 17611 (1994).
[51]E. Hernández and M. J. Gillan, Phys. Rev. B **51**, 10157 (1995).
[52]F. Mauri and G. Galli, Phys. Rev. B **50**, 4316 (1994).
[53]W. Kohn, Phys. Rev. Lett. **76**, 3168 (1996).
[54]E. Prodan and W. Kohn, Proc. Natl. Acad. Sci. U. S. A. **102**, 11635 (2005).
[55]J. M. Millam and G. E. Scuseria, J. Chem. Phys. **106**, 5569 (1997).
[56]P. D. Haynes, C.-K. Skylaris, A. A. Mostofi, and M. C. Payne, J. Phys.: Condens. Matter **20**, 294207 (2008).
[57]N. D. M. Hine, M. Robinson, P. D. Haynes, C.-K. Skylaris, M. C. Payne, and A. A. Mostofi, Phys. Rev. B **83**, 195102 (2011).
[58]H. C. Andersen, J. Chem. Phys. **72**, 2384 (1980).
[59]S. Nosé, Mol. Phys. **100**, 191 (2002).
[60]M. Parrinello and A. Rahman, J. Appl. Phys. **52**, 7182 (1981).
[61]H. B. Schlegel, J. M. Millam, S. S. Iyengar, G. A. Voth, A. D. Daniels, G. E. Scuseria, and M. J. Frisch, J. Chem. Phys. **114**, 9758 (2001).
[62]J. M. Herbert and M. Head-Gordon, J. Chem. Phys. **121**, 11542 (2004).
[63]A. M. N. Niklasson, C. J. Tymczak, and M. Challacombe, Phys. Rev. Lett. **97**, 123001 (2006).
[64]A. M. N. Niklasson and M. J. Cawkwell, J. Chem. Phys. **141**, 164123 (2014).
[65]M. Arita, D. R. Bowler, and T. Miyazaki, J. Chem. Theory Comput. **10**, 5419 (2014).
[66]T. Ozaki, Phys. Rev. B **64**, 195110 (2001).
[67]B. Jansík, S. Høst, P. Jørgensen, J. Olsen, and T. Helgaker, J. Chem. Phys. **126**, 124104 (2007).
[68]Q. Hill and C.-K. Skylaris, Proc. R. Soc. A **465**, 669 (2009).
[69]M. J. Gillan, D. Alfé, and A. Michaelides, J. Chem. Phys. **144**, 130901 (2016).
[70]I.-C. Lin, A. P. Seitsonen, I. Tavernelli, and U. Rothlisberger, J. Chem. Theory Comput. **8**, 3902 (2012).
[71]R. A. DiStasio, B. Santra, Z. Li, X. Wu, and R. Car, J. Chem. Phys. **141**, 084502 (2014).
[72]C. J. Burnham and S. S. Xantheas, J. Chem. Phys. **116**, 5115 (2002).
[73]B. Santra, R. A. DiStasio Jr., F. Martelli, and R. Car, Mol. Phys. **113**, 2829 (2015).
[74]M. L. Laury, L.-P. Wang, V. S. Pande, T. Head-Gordon, and J. W. Ponder, J. Phys. Chem. B **119**, 9423 (2015).
[75]G. J. Martyna, M. E. Tuckerman, D. J. Tobias, and M. L. Klein, Mol. Phys. **87**, 1117 (1996).
[76]J. Kolafa and J. W. Perram, Mol. Simul. **9**, 351 (1992).
[77]T. Darden, L. Perera, L. Li, and L. Pedersen, Structure **7**, R55 (1999).
[78]W. G. Hoover, Phys. Rev. A **31**, 1695 (1985).
[79]A. D. Becke, Phys. Rev. A **38**, 3098 (1988).
[80]C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
[81]S. Grimme, J. Comput. Chem. **27**, 1787 (2006).
[82]A. Odell, A. Delin, B. Johansson, M. J. Cawkwell, and A. M. N. Niklasson, J. Chem. Phys. **135**, 224105 (2011).

26 February 2024 09:53:54
MOST WIEDZY Downloaded from mostwiedzy.pl