# Multiple-timestep *ab initio* molecular dynamics based on two-electron integral screening

Shervin Fatehi[1] and Ryan P. Steele[1, a)]

*Department of Chemistry and Henry Eyring Center for Theoretical Chemistry, University of Utah, 315 South 1400 East, Salt Lake City, Utah 84112-0850, USA*

A multiple-timestep *ab initio* molecular dynamics scheme based on varying the two-electron integral screening method used in Hartree–Fock or density functional theory calculations is presented. Although screening is motivated by numerical considerations, it is also related to separations in the length- and timescales characterizing forces in a molecular system: Loose thresholds are sufficient to describe fast motions over short distances, while tight thresholds may be employed for larger length scales and longer times, leading to a practical acceleration of *ab initio* molecular dynamics simulations. Standard screening approaches can lead, however, to significant discontinuities in (and inconsistencies between) the energy and gradient when the screening threshold is loose, making them inappropriate for use in dynamics. To remedy this problem, a consistent window-screening method that smoothes these discontinuities is devised. Further algorithmic improvements reuse electronic-structure information within the dynamics step and enhance efficiency relative to a naïve multiple-timestepping protocol. The resulting scheme is shown to realize meaningful reductions in the cost of Hartree–Fock and B3LYP simulations of a moderately large system, the protonated-sarcosine/glycine dipeptide embedded in a 19-water cluster.

[a)]Electronic mail: ryan.steele@utah.edu

## I. INTRODUCTION

*Ab initio* molecular dynamics (AIMD) techniques combine the locality and efficiency of classical nuclear dynamics with on-the-fly generation of forces from electronic structure theory to generate real-time chemical information at a microscopic level.[1–4] Electronic structure theory can provide significantly more accurate forces than classical force fields, particularly for systems exhibiting strong polarization, charge transfer, and bond rearrangements. The cost of electronic-structure forces is many orders of magnitude larger than that of force fields, however, effectively prohibiting the use of AIMD for many systems and timescales of chemical interest.

Classical, non-polarizable force fields partition the potential energy of chemical systems into several bonded and non-bonded contributions,[1]

$$E = \underbrace{E_{\text{bonds}} + E_{\text{angles}} + E_{\text{dihedrals}}}_{E_{\text{bonded}}} + \underbrace{E_{\text{van der Waals}} + E_{\text{electrostatics}}}_{E_{\text{non-bonded}}}, \tag{1}$$

each of which is represented by analytic functions which can be differentiated to obtain forces. The bonded terms are typically parameterized to favor the equilibrium bond lengths, bond angles, and dihedral angles of the discrete chemical species in the system; the non-bonded terms primarily describe intermolecular interactions. While the Lennard-Jones potentials typically used to represent the van der Waals dispersion forces are basically short-range (decaying as $r^{-6}$), the $(1/r)$ Coulomb potential mediating the electrostatics is inherently long-range. As a result, evaluation of the electrostatic energy and forces is the most time-consuming part of force-field treatments of large systems.[5]

This problem is exacerbated in MD simulations by the fact that the fastest motions in the system — typically vibrations of hydrogen-containing bonds — set the maximum viable timestep; even though long-range electrostatic forces may change negligibly on the timescale of a bond vibration, a standard MD integrator would require that they be calculated at each timestep. A common strategy for eliminating this redundancy is to treat long-range electrostatics as an infrequent correction to a reference system containing the bonded, van der Waals, and short-range electrostatic interactions, such that the reference and correction are effectively integrated with different ("multiple") timesteps.[5–8]

Multiple-timestep (MTS) molecular dynamics can be derived most naturally from a Liouvillian formulation of classical mechanics,[9] although the MTS idea predates[10–15] the resulting

reversible reference-system propagator algorithm (r-RESPA). The basic algorithm is presented in ref 9 and will not be reviewed here. For the purposes of this work, multiple-timestep MD constitutes a sequence of $n$ velocity-Verlet "inner timesteps" using a computationally expedient force, followed by a momentum correction (or "outer timestep") associated with the remaining contributions to the total force.

In AIMD, analytic potential and force functions are replaced by electronic-structure energies and gradients, which are determined from the full-system Hamiltonian $H(\mathbf{R})$ and geometry $\mathbf{R}$ via a nonlinear self-consistent-field (SCF) procedure. This black-box character comes at the cost of separability; in particular, an explicit partitioning of forces is manifestly impossible. It would seem, then, that *ab initio* molecular dynamics simulations must be limited in scope to short timescales.

Multiple-timestep AIMD schemes *can* be formulated, nevertheless. The full-system force from a low-level theory $\mathbf{F}_{\text{low}}$ is taken as a reference and subsequently corrected with the force *difference* from a higher-level theory, $\Delta\mathbf{F} \equiv \mathbf{F}_{\text{high}} - \mathbf{F}_{\text{low}}$. Over the last decade, MTS has been applied to AIMD in this way. For example, an affordable density functional theory (such as PBE) can be corrected with forces from higher-level DFT (such as B3LYP).[16,17] These schemes were treated as *ad hoc* approaches for accelerating dynamics calculations, and general physical justifications for their use remains an open question. More formally justified MTS methods have also been used to propagate electronic information in the context of extended-Lagrangian MD approaches.[18,19]

Recently, an *ab initio* multiple-timestep scheme was presented[20] in which Hartree–Fock theory (HF) was corrected with a correlated wavefunction method, second-order Møller–Plesset perturbation theory (MP2).[21] In the limit of large system size $N$, HF/MP2 multiple-timestepping realizes a reduction in the computational cost of an MD simulation which is linear in the number of inner timesteps, as expected from the cost dominance of MP2 over HF [$\mathcal{O}(N^5)$ versus $\mathcal{O}(N^4)$ formal scaling].[22,23] Moreover, this scheme has a solid physical foundation — the observed separation between the timescales of variation in the HF forces and in the additional correlation forces obtained from MP2. To the extent that similar separations can be identified between HF, MP2, and more powerful correlated methods (such as coupled-cluster theory), highly accurate dynamics simulations of small systems may soon become feasible.

Of course, correlated wavefunction theories are cost-prohibitive for studying truly large

systems, and even Hartree–Fock or DFT may be unappealingly costly. If these latter methods were considered monolithic, then the usefulness of multiple-timestepping as a AIMD acceleration tool would already be exhausted. These basic methods can, however, vary widely in accuracy and computational cost, depending on a range of choices made in advance of a calculation. Some of these choices reflect purely numerical considerations, such as the convergence criterion for the self-consistent-field procedure; others have physical content, such as the basis set or density functional employed. The extent to which variations in each of these (generalized) parameters would support a multiple-timestepping scheme is a subject that remains largely unexplored. The present work is focused on the effect of changes in two-electron integral screening, keeping other parameters fixed.

### A. Motivation for screening-based multiple-timestepping

Screening-based multiple-timestepping is motivated by the connection between the magnitude of each two-electron integral and the length scale of the interactions it describes. Consider the integral

$$(\mu\nu|\lambda\sigma) = \int \mathrm{d}\mathbf{r}_1 \int \mathrm{d}\mathbf{r}_2 \; \phi_\mu^*(\mathbf{r}_1;\mathbf{R}_A)\phi_\nu(\mathbf{r}_1;\mathbf{R}_B)\frac{1}{r_{12}}\phi_\lambda^*(\mathbf{r}_2;\mathbf{R}_C)\phi_\sigma(\mathbf{r}_2;\mathbf{R}_D), \tag{2}$$

where $\phi_\mu(\mathbf{r}_1;\mathbf{R}_A)$ is a contracted-Gaussian atomic orbital centered on the nucleus at position $\mathbf{R}_A$ and occupied by an electron with coordinate $\mathbf{r}_1$. Orbital pair $\mu\nu$ forms a charge distribution centered at $\mathbf{R}_{\mu\nu}$ which decays exponentially as $S_{\mu\nu} \sim \exp\left(-\alpha_{\mu\nu}R_{AB}^2\right)$, with $R_{AB} \equiv |\mathbf{R}_A - \mathbf{R}_B|$ the internuclear distance and $\alpha_{\mu\nu}$ a composite of primitive-Gaussian widths. The asymptotic behavior of the integral is therefore given by[24]

$$(\mu\nu|\lambda\sigma) \sim \frac{S_{\mu\nu}S_{\lambda\sigma}}{R_{\mu\nu\lambda\sigma}}, \tag{3}$$

where $R_{\mu\nu\lambda\sigma} \equiv |\mathbf{R}_{\mu\nu} - \mathbf{R}_{\lambda\sigma}|$ is the distance between the charge centers.

If either charge distribution is highly attenuated — its constituent orbitals are centered on distant nuclei — or if the charge distributions themselves are well-separated, the integral will be small. It may be possible to discard such an integral altogether without measurably affecting the accuracy of the calculation. "Screening" is the umbrella term for a constellation of techniques which exclude integrals on this basis, thereby reducing the formal $\mathcal{O}(N^4)$ cost of constructing these four-index quantities to a more modest quadratic (or even linear) scaling.[24]

Eq 3 clearly states that the size of a two-electron integral is inversely proportional to the length scale of the electronic interaction it describes. Thus, the screening threshold is equivalent, in some sense, to a range-separation parameter for the interelectronic repulsion. Given that interactions over small length scales tend to be associated with motions of high frequency, the forces associated with integrals of disparate size should also vary on disparate timescales, which is a scenario tailor-made for multiple-timestepping. This intuition is borne out in the analysis that follows, and it is the foundation of a practical approach for accelerating AIMD simulations.

To the extent that the intuitive correspondence between the screening threshold and a range-separation parameter is valid, the multiple-timestep approach developed here is similar in spirit to the MTS-CASE method of Luehr and coworkers.[25] MTS-CASE is based on an explicit partitioning of the bare Coulomb potential felt by both electrons *and* nuclei,

$$\frac{1}{r} = \frac{\mathrm{erf}(\omega r)}{r} + \frac{\mathrm{erfc}(\omega r)}{r}, \tag{4}$$

where $\mathrm{erf}(\omega r)$ is the (long-range) error function; $\mathrm{erfc}(\omega r)$ is the (short-range) complementary error function; and the parameter $\omega$ sets the length scale of the range separation. (The name "MTS-CASE" stems from the fact that neglecting the long-range Coulomb tail in the inner timesteps is equivalent to adopting the Coulomb-attenuated Schrödinger equation, or CASE model.[26]) Luehr *et al.* noted[25] that MTS-CASE inner timesteps can be made linearly scaling, owing to efficient screening of the CASE two-electron integrals,[27] although this connection has not yet been pursued. Evidence of a rigorous timescale separation between CASE and standard Hartree–Fock forces also remains to be presented, but such a separation is to be expected. In the analysis that follows, the range separation implicit in the screening threshold is explicitly shown to be associated with a separation in timescale, as required to justify MTS. Direct comparisons between our method and MTS-CASE are reserved for future work.

### B. Outline of the present work

This manuscript is organized as follows. In Section II, a detailed description of the development of the screening-based multiple-timestep scheme is provided, including a review of the simplest complete sequence of screening methods used in computing *ab initio* energies and gradients (II A); a demonstration that these classic techniques break down when

the screening threshold is significantly loosened (II B); and a remedy for these pathologies (II C). The screening-based MTS protocol is formulated in Section II D, and the method is shown to be phenomenologically justified by a timescale separation in the forces. Instead of proceeding immediately to numerical tests, Section II E highlights the fact that classical and *ab initio* multiple-timestep schemes are governed by different cost inequalities, such that computational savings are harder to achieve in the *ab initio* setting. Several approaches to subverting this limitation are described. In Section III, timings are presented for a model biological system, the protonated-sarcosine/glycine dipeptide embedded in a 19-water cluster. Section IV concludes with a summary of findings and a discussion of possible future directions.

## II. METHODS DEVELOPMENT

A screening-based multiple-timestep AIMD scheme is developed in this work. Briefly stated, the method consists of several standard MD steps at loose screening thresholds, followed by a correction to the momenta — in the usual r-RESPA fashion — from the difference in forces obtained using loose and tight thresholds. Although this MTS approach is conceptually straightforward, it entails several technical hurdles, which require both an explication of integral screening and an analysis of the pathologies that can arise in standard screening techniques. These pathologies can directly impact the accuracy of single-timestep AIMD as well, and, to the authors' knowledge, have not been analyzed in the literature.

### A. Review of basic screening methods

This section summarizes the simplest complete sequence of screening methods that might be employed in computing energies and gradients using an atomic-orbital-based quantum-chemistry software package.[28–31]

#### *1. Before SCF: Shell-pair economization*

Because screening is based on the asymptotic behavior of the two-electron integrals, it requires a modest computational overhead relative to the cost of computing the integrals outright. In the limit of large system size, however, the number of integrals to be screened

continues to grow quartically, even as the number of integrals *retained* grows quadratically. As a result, the screening overhead can eventually dominate the cost of the calculation. One means of reducing the overhead is to eliminate integrals from consideration by some crude (but efficient) pre-screening method. Pre-screening is *not* the focus of the methodological developments of this work, but its relationship to more rigorous screening merits brief review.

In contracted-Gaussian basis sets, each atomic orbital is composed of a sum of primitive Gaussians. For example, an $s$-type orbital has the form

$$\phi_\mu(\mathbf{r}_1;\mathbf{R}_A) = \sum_m c_{\mu m} g_m(\mathbf{r}_1;\mathbf{R}_A), \tag{5}$$

where $c_{\mu m}$ is the contraction coefficient of the normalized primitive $g_m(\mathbf{r}_1;\mathbf{R}_A)$ with width $\alpha_m$. Consequently, the charge distribution associated with an $s$-type orbital pair $\mu\nu$ is a linear combination of primitive Gaussian "shell pairs" $mn$,

$$\phi^*_\mu(\mathbf{r}_1;\mathbf{R}_A)\phi_\nu(\mathbf{r}_1;\mathbf{R}_B) = \sum_{mn} c^*_{\mu m} c_{\nu n} g^*_m(\mathbf{r}_1;\mathbf{R}_A) g_n(\mathbf{r}_1;\mathbf{R}_B). \tag{6}$$

The pre-screening approach known as "shell-pair economization" consists of computing the orbital pairs $\mu\nu$ before the SCF procedure begins, discarding shell pairs if their overlap is smaller than a threshold value $\varepsilon_{\text{shell}}$.[31] The overlap is given analytically by

$$s_{mn} = \int \mathrm{d}\mathbf{r}_1\, g_m(\mathbf{r}_1;\mathbf{R}_A) g_n(\mathbf{r}_1;\mathbf{R}_B) \tag{7}$$

$$= \left(\frac{\pi}{\alpha_{mn}}\right)^{\frac{3}{2}} \mathcal{N}_m \mathcal{N}_n e^{-\frac{\alpha_m \alpha_n}{\alpha_{mn}} R^2_{AB}}, \tag{8}$$

where $\alpha_{mn} \equiv \alpha_m + \alpha_n$ and $\mathcal{N}$ denotes a normalization constant. Shell-pair economization may therefore be written as

$$\phi_\mu(\mathbf{r}_1;\mathbf{R}_A)\phi_\nu(\mathbf{r}_1;\mathbf{R}_B) \approx \sum_{mn} c_{\mu m} c_{\nu n} \Theta\left(s_{mn} - \varepsilon_{\text{shell}}\right) g_m(\mathbf{r}_1;\mathbf{R}_A) g_n(\mathbf{r}_1;\mathbf{R}_B), \tag{9}$$

with $\Theta(x)$ the (right-continuous) Heaviside step function,

$$\Theta(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0. \end{cases} \tag{10}$$

If the internuclear distance $R_{AB}$ or threshold $\varepsilon_{\text{shell}}$ is large, all of the shell pairs constituting $\mu\nu$ might be discarded by eq 9. All of the integrals involving $\mu\nu$ would, in turn, be excluded from the calculation.

### *2. During SCF: Integral screening*

After shell-pair economization, the energy is determined via the SCF procedure. Whether the system is treated with Hartree–Fock or Kohn–Sham DFT, each SCF cycle involves the construction of a contribution to the energy from the two-electron integrals,

$$E^{2e^-} = \frac{1}{2}\mathrm{Tr}\ \mathbf{P}\cdot\mathbf{F}^{2e^-} = \frac{1}{2}\mathrm{Tr}\ \mathbf{P}\cdot(\mathbf{J}-c_\mathrm{X}\mathbf{K}) = \frac{1}{2}\mathrm{Tr}\ \mathbf{P}\cdot\mathbf{\Pi}\cdot\mathbf{P}. \tag{11}$$

Here, $\mathbf{P}$ is the density matrix; $\mathbf{F}^{2e^-}$ is the two-electron portion of the Fock matrix, comprising the Coulomb matrix $\mathbf{J}$ and an appropriate fraction $c_\mathrm{X}$ of the exact Hartree–Fock exchange matrix $\mathbf{K}$; and $\mathbf{\Pi}$ denotes the antisymmetrized two-electron integrals. (Spin indices and sums have been suppressed.) In the equations that follow, $c_\mathrm{X}$ is taken to be 1, which is the appropriate value for Hartree–Fock. The resulting screening expressions are general; calculations using hybrid functionals ($0 < c_\mathrm{X} \leq 1$) use the same screening criteria but scale $\mathbf{K}$ after it has been constructed. Calculations using pure functionals, by contrast, simply omit the exchange-related terms from any given screening prescription.

Taking permutational symmetry into account, the two-electron contributions to the energy are proportional to

$$E^{2e^-} \propto \sum_{\mu\nu\lambda\sigma} (\mu\nu|\lambda\sigma)\,(\underbrace{2P_{\mu\nu}P_{\lambda\sigma}}_{\text{Coulomb}}\underbrace{-P_{\mu\lambda}P_{\nu\sigma} - P_{\mu\sigma}P_{\nu\lambda}}_{\text{exchange}}), \tag{12}$$

The overarching goal of screening is to avoid computing integrals unless they contribute materially to the energy. An affordable estimate of the largest contribution from $(\mu\nu|\lambda\sigma)$ in eq 12 is, therefore, required. The bracket representing the integral has the properties of an inner product, so the Schwarz inequality can be invoked,[29]

$$(\mu\nu|\lambda\sigma) \leq \sqrt{(\mu\nu|\mu\nu)}\sqrt{(\lambda\sigma|\lambda\sigma)} \equiv Q_{\mu\nu}Q_{\lambda\sigma}. \tag{13}$$

The Schwarz estimate $Q_{\mu\nu}Q_{\lambda\sigma}$ provides a rigorous upper bound on $(\mu\nu|\lambda\sigma)$ in terms of the self-integrals of molecular-orbital pairs, which scale as $\mathcal{O}(N^2)$. This bound could be tightened further by taking the asymptotic $R^{-1}_{\mu\nu\lambda\sigma}$ dependence of the exact integral into account,[24,32] but the common practice in atomic-orbital-based quantum chemistry has been to use eq 13 without further refinement.

Because density-matrix elements are restricted in magnitude, $|P_{\mu\nu}| \in [0,1]$, eq 12 can be converted into a rigorous estimate for the largest unsigned contribution to the energy from

$(\mu\nu|\lambda\sigma)$,

$$\overline{E} \equiv Q_{\mu\nu} Q_{\lambda\sigma} \max\left(2P_{\mu\nu}P_{\lambda\sigma}, P_{\mu\lambda}P_{\nu\sigma}, P_{\mu\sigma}P_{\nu\lambda}\right). \tag{14}$$

The screening method suggested by this expression is to discard the integral if $\overline{E}$ is smaller than a threshold $\varepsilon_{\text{hard}}$:

$$(\mu\nu|\lambda\sigma) \to \Theta(\overline{E} - \varepsilon_{\text{hard}})\,(\mu\nu|\lambda\sigma). \tag{15}$$

In fact, such an approach would be unwise:[28] Although computation of the SCF energy is the aim, steady convergence of the SCF procedure hinges on the accuracy of the two-electron portion of the Fock matrix, $\mathbf{F}^{2e^-} = \mathbf{P} \cdot \mathbf{\Pi}$, which only includes a single factor of the density. Thus, integrals with negligible estimated contributions to the energy may nevertheless be important in the Fock matrix. Eq 14 can be adapted to reflect this observation by splitting pairs of density-matrix elements,

$$\overline{F} \equiv Q_{\mu\nu} Q_{\lambda\sigma} \max\left(2P_{\mu\nu}, 2P_{\lambda\sigma}, P_{\mu\lambda}, P_{\mu\sigma}, P_{\nu\lambda}, P_{\nu\sigma}\right). \tag{16}$$

Screening is then equivalent to the replacement

$$(\mu\nu|\lambda\sigma) \to \Theta(\overline{F} - \varepsilon_{\text{hard}})\,(\mu\nu|\lambda\sigma). \tag{17}$$

### *3. After SCF: Gradient screening*

With the converged density matrix in-hand, the gradient can now be computed. Because the density is obtained variationally, the only explicit contribution to the gradient from the two-electron integrals is the term

$$\left(\nabla_{\mathbf{R}} E\right)^{2e^-} = \frac{1}{2}\text{Tr}\ \mathbf{P} \cdot \mathbf{\Pi}^{[\mathbf{R}]} \cdot \mathbf{P}, \tag{18}$$

where the superscript $[\mathbf{R}]$ is shorthand for the Cartesian gradient over all nuclear coordinates. Any given integral depends on, at most, four nuclear coordinates, so the scaling of this term is still formally $\mathcal{O}(N^4)$.

A gradient-screening method could potentially be developed directly from eq 18 (or, equivalently, from the gradient of eq 12).[33] A rigorous, practical alternative is to take the gradient of the energy-based screening prescription of eq 15, on the grounds that only those integrals included in the energy should subsequently contribute to the gradient. The derivative of the Heaviside function is a delta function, so this contribution is given by

$$(\mu\nu|\lambda\sigma)^{[\mathbf{R}]} \to \Theta\left(\overline{E} - \varepsilon_{\text{hard}}\right)(\mu\nu|\lambda\sigma)^{[\mathbf{R}]} + \delta\left(\overline{E} - \varepsilon_{\text{hard}}\right)\overline{E}^{[\mathbf{R}]}\,(\mu\nu|\lambda\sigma). \tag{19}$$

The delta-function term in this expression has a straightforward meaning; it describes the discontinuity in $\Theta$ which admits the integral into the energy calculation when $\overline{E} = \varepsilon_{\text{hard}}$. This term can be neglected in the gradient, because (1) the delta function goes to infinity when its argument vanishes, which is unphysical, and (2) its effect is restricted to a single value of $\overline{E}$, which occurs with vanishingly small probability. Thus, gradient screening is equivalent to the replacement

$$(\mu\nu|\lambda\sigma)^{[\mathbf{R}]} \to \Theta\left(\overline{E} - \varepsilon_{\text{hard}}\right)(\mu\nu|\lambda\sigma)^{[\mathbf{R}]}. \tag{20}$$

Eqs 15 and 17 may admit somewhat different sets of integrals, especially when $\varepsilon_{\text{hard}}$ is large; hence, eq 20 is "inconsistent" gradient screening.

To summarize, a complete prescription for two-electron-integral screening in a Hartree–Fock or DFT calculation of the energy and gradient consists of (1) pre-screening shell pairs against a threshold $\varepsilon_{\text{shell}}$ using eq 9 and (2) screening contributions to the Fock matrix and gradient against a threshold $\varepsilon_{\text{hard}}$ using eqs 17 and 20. By analogy to SCF convergence criteria, large thresholds are "loose," in the sense that the resulting energy will exhibit errors of comparable size, while small thresholds are "tight."

### B. Loose thresholds reveal flaws in `hard` screening

A screening-based multiple-timestep scheme will necessarily involve loosening the threshold in the inner timesteps. Instead of varying the thresholds for shell-pair economization *and* integral screening in the calculations that follow, the pre-screening threshold $\varepsilon_{\text{shell}}$ is tightened to the smallest value allowed ($10^{-14}$) by the chosen development environment, Q-Chem.[34] This choice is motivated by the fact that shell-pair economization is not based on a rigorous integral bound; rather, it is a numerical convenience. Consequently, varying $\varepsilon_{\text{shell}}$ can lead to unpredictable changes in the two-electron integrals. This lack of rigor is in direct conflict with the intention to construct a multiple-timestep scheme based on *well-controlled* changes in screening. The base method for screening-based multiple-timestepping therefore consists of Fock-matrix-based integral screening (eq 17) and inconsistent gradient screening (eq 20) using threshold $\varepsilon_{\text{hard}} = 10^{-h}$, where $h$ is a positive integer. For the sake of brevity, we refer to this method as `hard-i-`$h$.

It may be advantageous for $\varepsilon_{\text{hard}}$ to be looser than the chosen SCF convergence criterion,

$10^{-8}$ $E_h$. In order to establish the properties of `hard` screening in this unusual regime, HF/6-31G* potential energy curves and gradients were computed for a series of diatomic molecules — CO, $He_2$, HCl, HF, $H_2$, LiF, LiH, $N_2$, and $O_2$ — with bond separations $R \in [0.5, 4.0]$ Å. The increment between points was 0.025 Å, such that each surface was constructed from 141 individual gradient calculations. For each molecule, these calculations were repeated over a range of integer thresholds, $h \in [2, 14]$. The effective floor on integral thresholds in Q-Chem is $h = 14$, because the fundamental electron-repulsion integrals are computed by modified Chebyshev interpolation from tabulated values, with a precision set to $10^{-(h+2)}$.[35]

At the loosest thresholds ($h = 2$ and 3), the calculations unsurprisingly exhibited poor convergence behavior, often failing to converge even after hundreds of SCF cycles using DIIS extrapolation.[36,37] These failures occur due to the exclusion of large integrals which are necessary to obtain a well-converged solution to the Roothaan equations. For the remaining thresholds, $h \in [4, 14]$, all of the potential curves and gradients were visually indistinguishable on the scale of the plots. Subtle features in the curves for each method were revealed by plotting unsigned energy and gradient *differences* with respect to `hard-i-14`, which is treated as reference data.

As a concrete example, energy and gradient differences for lithium hydride, treated with `hard-i-4` screening, are plotted in Figures 1A and 1C. Both the energy and the gradient exhibit discontinuities with respect to the reference curve, some of which lie very close to (or even on top of) the potential minimum, which is the region most relevant for dynamics. Discontinuities in the energy are of the same order as $\varepsilon_{\text{hard}}$, with concomitantly large discontinuities in the gradient; further, large spikes in the energy are not always matched by corresponding features in the gradient (and *vice versa*), confirming the expectation that inconsistent gradient screening might be problematic at loose thresholds.

Any discontinuity in the energy and gradient will have deleterious effects in a molecular dynamics simulation, because the system will experience unphysical, instantaneous impulses. Inconsistencies in the gradient are especially troubling because they imply that the forces used to integrate the trajectory will be fundamentally incorrect. The solution to this latter problem is straightforward: The gradients should be screened using the Fock estimate of eq 16 rather than the energy estimate of eq 14,

$$(\mu\nu|\lambda\sigma)^{[\mathbf{R}]} \rightarrow \Theta(\overline{F} - \varepsilon_{\text{hard}})\,(\mu\nu|\lambda\sigma)^{[\mathbf{R}]}\,. \tag{21}$$

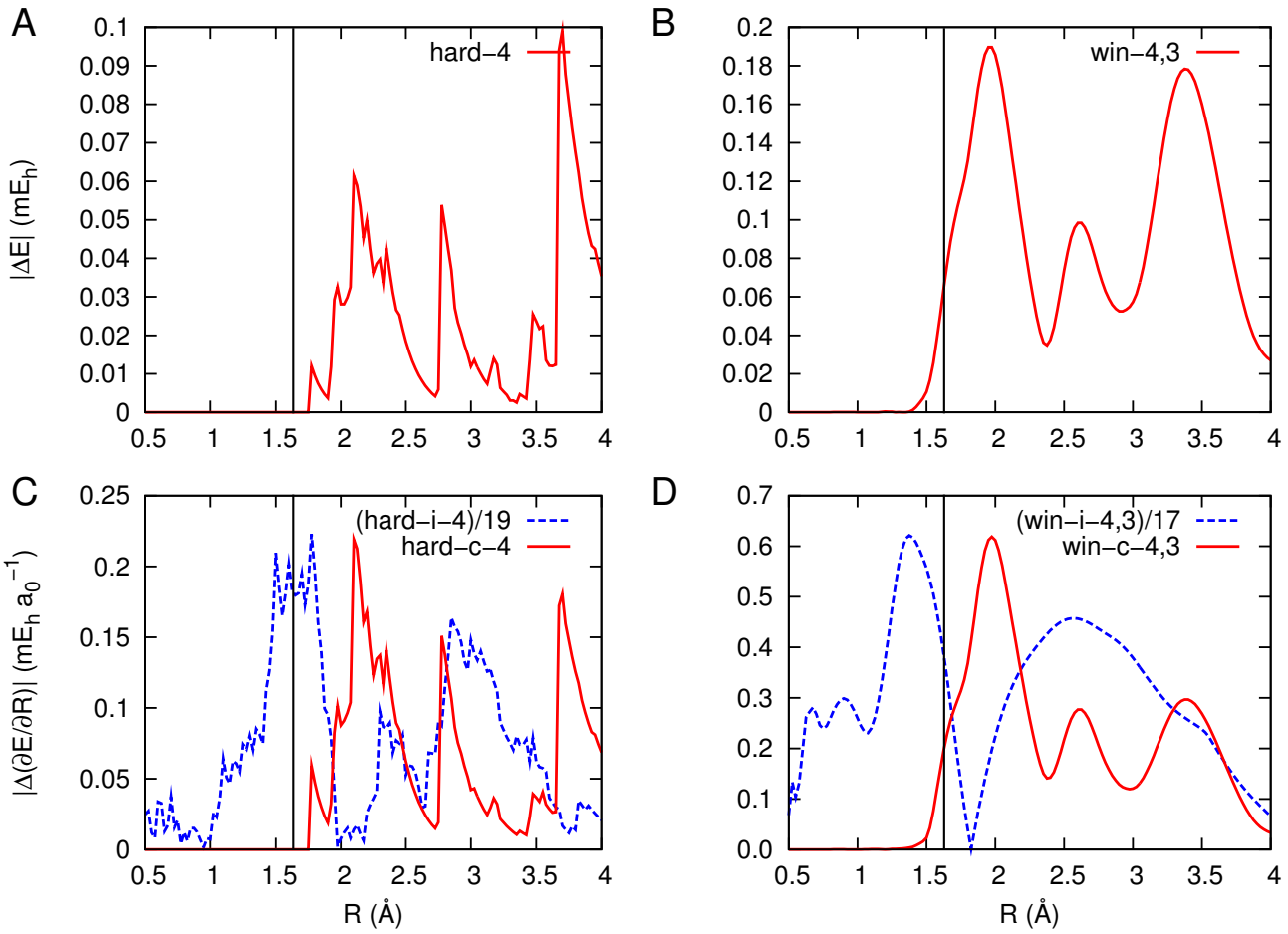



FIG. 1. Unsigned energy and gradient differences with respect to a `hard-i-14` reference, $|\Delta E|$ and $|\Delta(\partial E/\partial R)|$, obtained using various screening methods for lithium hydride at the HF/6-31G* level of theory. Panels A and B show the energy differences for `hard-4` and `win-4,3`, respectively; gradient consistency is immaterial. Panel C shows the gradient differences for `hard-i-4` (divided by 19, blue dashed line) and `hard-c-4` (red solid line); Panel D, the gradient differences for `win-i-4,3` (divided by 17, blue dashed line) and `win-c-4,3` (red solid line). The dropline in each panel indicates the optimum Li–H bond separation, $R = 1.635749$ Å. See Sections II B and II C for detailed discussion.

When the Fock-based integral screening of eq 17 is paired with this consistent gradient screening, the combined method is referred to as `hard-c-`$h$. The gradient results in Figure 1C show that, while still discontinuous, the `hard-c-4` gradient is fully consistent with the energy. Note that `hard-i-14` results remain a good reference — they are indistinguishable from `hard-c-14`, which only serves to emphasize that screening in the loose-threshold regime

is qualitatively different from tighter screening.

### C. Window screening

Remedying the remnant discontinuities that appear in `hard` energies and gradients at loose thresholds requires an understanding of their origin. `Hard` screening effectively assigns a binary weight to each integral, as dictated by the Heaviside weight function $w_{\text{hard}}(\overline{F};\varepsilon_{\text{hard}}) = \Theta\left(\overline{F}-\varepsilon_{\text{hard}}\right)$ plotted in Figure 2A. Now, consider an integral with estimated Fock-matrix contribution comparable to the threshold ($\overline{F}\approx\varepsilon_{\text{hard}}$) when the system is in some geometry $\mathbf{R}$. Whether that integral will be retained or discarded after a small perturbation to the geometry, $\mathbf{R}+\mathrm{d}\mathbf{R}$, will be determined by changes in the Schwarz estimates and density-matrix elements constituting $\overline{F}$. Depending on the sign and magnitude of these changes, therefore, the integral can "blink" on or off. When $\varepsilon_{\text{hard}}$ is very tight, blinking integrals will have an insignificant effect on the energy and gradient, but when the threshold is sufficiently loose, these integrals will lead to the observed discontinuities.

If screening-based multiple-timestepping is to be viable, a screening method must be developed which is less sensitive to small changes in the integral estimate, such that the corresponding potential energy curves and gradients will be continuous. To accomplish this aim, the hard cutoff $\varepsilon_{\text{hard}}$ is paired with a looser "soft" threshold $\varepsilon_{\text{soft}} = 10^{-s}$ to form a window spanning $\Omega \equiv h-s$ orders of magnitude. This approach is hereafter termed window screening.

Within the window, each integral is assigned a fractional weight from a smoothly-varying function satisfying two conditions: First, integrals with estimates close to $\varepsilon_{\text{hard}}$ should be assigned very small weights, such that blinking is a more modest source of error. Second, integrals with estimates close to $\varepsilon_{\text{soft}}$ should be rescaled only very delicately, with integrals above the soft threshold carrying full weight. Sigmoid functions of the *logarithm* of the estimate, $\overline{f} \equiv -\log\overline{F}$, have just these properties; a quintic spline with vanishing first and second derivatives at the boundaries of the window is chosen here, in direct analogy to smoothing functions used in force-field-based multiple-timestep approaches.[38,39]

The window-screening approach for the energy is therefore defined by the replacement

$$(\mu\nu|\lambda\sigma) \to w(\overline{f};h,s)\,(\mu\nu|\lambda\sigma)\,, \tag{22}$$

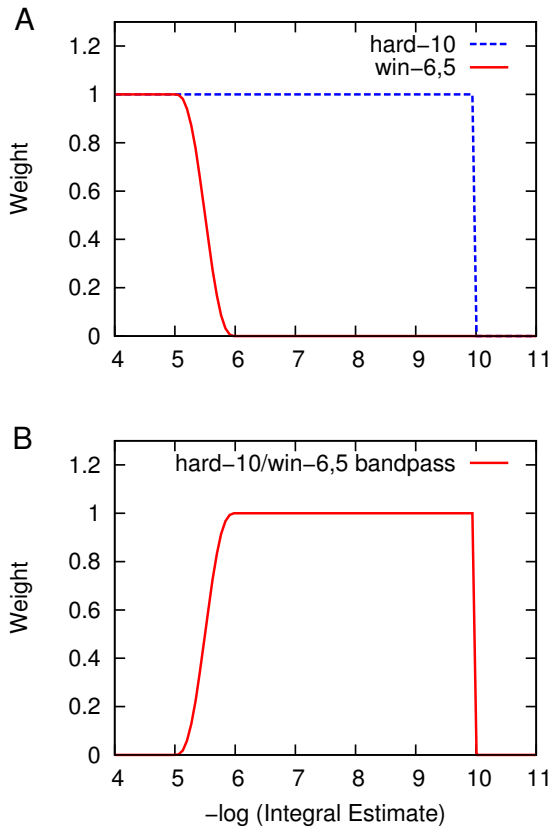



FIG. 2. Weight functions used in integral screening as a function of the logarithm of the integral estimate. Panel A depicts a `hard-10` weight function (blue dashed line) and a `win-6,5` window-screening function (red solid line), as described in Section II C. Panel B depicts the corresponding bandpass-screening function, eq 44.

where the weight function $w$ is

$$w(\overline{f};h,s) = \begin{cases} 1, & \overline{f} \le s \\ 1 + u^3(15u - 6u^2 - 10), & s < \overline{f} < h \\ 0, & \overline{f} \ge h \end{cases} \tag{23}$$

with spline variable $u$ given by

$$u(\overline{f};h,s) = \frac{\overline{f} - s}{\Omega}. \tag{24}$$

An example of this weight function is plotted in Figure 2A. As may be apparent from the figure, $w = u = \frac{1}{2}$ at the midpoint of the window, $\overline{f} = s + (\Omega/2)$. Equivalently, integrals $(\sqrt{10})^{\Omega}$ times smaller than the soft threshold — a bit more than a factor of 3, for an integer window — will be assigned half weight or less.

Because eq 23 recovers the Heaviside function when $h = s$, window screening can be understood as a `hard` screening generalization, which generates the same set of two-electron integrals but is not as susceptible to blinking errors. Additional error is incurred because the integrals have been rescaled (thereby interpolating between `hard`-$h$ and `hard`-$s$ results), but Figure 1B indicates that the largest errors only increase by a factor of 2. Importantly, these errors are associated with smooth features in the potential-energy surface, which demonstrates that the desired aim has been achieved. In practice, an integer window has been found to achieve the best compromise between accuracy and smoothing.

Of course, the discontinuities that appear in the gradient must also be smoothed. As before, the gradient of eq 22 is taken to obtain

$$(\mu\nu|\lambda\sigma)^{[\mathbf{R}]} \to w(\overline{f}; h, s)\,(\mu\nu|\lambda\sigma)^{[\mathbf{R}]} + \frac{\partial w}{\partial \overline{f}}\overline{f}^{[\mathbf{R}]}\,(\mu\nu|\lambda\sigma)\,. \tag{25}$$

The quintic spline $w$ was chosen because its derivatives vanish at the boundaries of the window, which significantly limits the effect of the weight-function derivative in eq 25. But unlike the delta function that initially appeared in the `hard` gradient (eq 19), this term is neither patently unphysical nor restricted in effect to a single value of $\overline{f}$:

$$\frac{\partial w}{\partial \overline{f}} = \begin{cases} 0, & \overline{f} \le s \\ -\frac{30}{\Omega}u^2(u-1)^2, & s < \overline{f} < h \\ 0 & \overline{f} \ge h. \end{cases} \tag{26}$$

The derivative smoothly increases from zero at either boundary to a modest value of $(\partial w/\partial \overline{f}) = -(15/8\Omega)$ at the center of window. The gradient of $\overline{f}$ is given by

$$\overline{f}^{[\mathbf{R}]} = -\left(\log \overline{F}\right)^{[\mathbf{R}]} = -\left(\frac{\ln \overline{F}}{\ln 10}\right)^{[\mathbf{R}]} = -\frac{\overline{F}^{[\mathbf{R}]}}{\overline{F}\ln 10}. \tag{27}$$

Evaluating the gradient of the Fock-matrix estimate (eq 16) and canceling terms,

$$\frac{\partial w}{\partial \overline{f}}\overline{f}^{[\mathbf{R}]}\,(\mu\nu|\lambda\sigma) = \frac{30u^2(u-1)^2}{\Omega \ln 10}\left(\frac{Q_{\mu\nu}^{[\mathbf{R}]}}{Q_{\mu\nu}} + \frac{Q_{\lambda\sigma}^{[\mathbf{R}]}}{Q_{\lambda\sigma}} + \frac{P_{\max}^{[\mathbf{R}]}}{P_{\max}}\right)(\mu\nu|\lambda\sigma)\,, \tag{28}$$

where $P_{\max}$ is shorthand for the largest of the density-matrix elements in eq 16.

Each Schwarz-gradient term in eq 28 will vanish identically for all nuclear coordinates but those of the (one or two) centers forming the relevant orbital pair. For those coordinates, the asymptotic behavior will be, for example,

$$\frac{Q_{\mu\nu}^{[\mathbf{R}_A]}}{Q_{\mu\nu}} \sim \frac{R_{AB}e^{-\alpha_{\mu\nu}R_{AB}^2}}{e^{-\alpha_{\mu\nu}R_{AB}^2}}, \tag{29}$$

which tends to zero in the limit of small *or* large internuclear separations. These terms may, therefore, be discarded with little or no effect.

The density gradient $P_{\max}^{[\mathbf{R}]}$ will be largest when perturbations to the geometry strongly mix occupied and virtual orbitals, such as when the HOMO–LUMO gap is small. To the extent that Hartree–Fock and DFT are unreliable in the limit of vanishing fundamental gap, this term is discarded, too. There is an additional practical reasons for doing so: To evaluate $P_{\max}^{[\mathbf{R}]}$ exactly would require that the coupled-perturbed Hartree–Fock or Kohn–Sham equations be solved, at significant additional cost.[40–44] As shown below, neglecting this gradient has no serious consequences for MD; in fact, the statistical properties of the dynamics improve when window-screened gradients are used.

In light of the above discussion, the consistent window screening method `win-c-`$h, s$ is implemented for gradients as

$$(\mu\nu|\lambda\sigma)^{[\mathbf{R}]} \to w(\overline{f}; h, s)\,(\mu\nu|\lambda\sigma)^{[\mathbf{R}]}\,. \tag{30}$$

Inconsistent screening (`win-i-`$h, s$) can be performed instead by replacing $\overline{f}$ in eq 30 with the negative logarithm of the energy estimate $\overline{E}$ (eq 14).

Lithium hydride gradients obtained using these methods are shown in Figure 1D. As was observed for the energy, the size of the errors in the gradient increases by a factor of 2 or 3; the associated discontinuities are fully smoothed. Moreover, the behavior of the `win-c` gradient is consistent with the behavior of the energy.

### D. Formulation and validation of screening-based multiple-timestepping

Having established that window-screening methods smooth the discontinuities engendered in potential energy surfaces and gradients by standard screening approaches, this section examines the effect of screening method on the statistical properties of single-timestep *ab initio* molecular dynamics trajectories in the microcanonical ensemble. To probe the effect of screening, the water dimer was simulated with HF/3-21G. The initial configuration of the system was obtained by reoptimizing the minimum-energy structure from the Cambridge Cluster Database[45,46] using `hard-c-14` screening; initial velocities were identical in all simulations and were sampled from the Maxwell–Boltzmann distribution at 298 K.

Trajectories were computed using a range of screening methods — `hard` and `win`; with and without gradient consistency; with hard cutoffs $h \in \{4, 6, 8, 10, 12, 14\}$ and soft thresholds

$s = h - 1$; and with timesteps $\Delta t \in \{5, 10, 15, 20, 30, 40, 50, 75, 100\}$ au. Each trajectory was integrated for a maximum of $\tau_{\text{steps}} = 10,000$ steps or to the first instance of failed SCF convergence. In this way, the longest viable timestep was found to be $\Delta t = 75$ au; when $\Delta t = 100$ au, one of the water molecules invariably autoionized, followed closely by SCF convergence failure. While this result is entirely dependent on the choice of SCF convergence algorithm, the success or failure of a series of trajectories — each employing the *same* convergence algorithm — is indicative of the robustness of the MD method. This result also highlights the fact that AIMD functions as a reactive force field, and malignancies in the forces or the molecular-dynamics integrator can lead to qualitatively incorrect behavior.

While symplectic integrators are guaranteed to yield energy-conserving trajectories for small enough timesteps,[47] step-to-step energy fluctuations are not guaranteed to vanish, and these fluctuations are indicative of the quality of the integration. Energy conservation of all complete trajectories was assessed using the time-average of these energy fluctuations relative to the initial energy, $E_0$:

$$\langle \delta \rangle = \left\langle \frac{E_\tau - E_0}{E_0} \right\rangle = \frac{1}{\tau_{\text{steps}}} \sum_{\tau=1}^{\tau_{\text{steps}}} \frac{E_\tau - E_0}{E_0}. \tag{31}$$

The smaller the value of $\langle \delta \rangle$, the more closely the trajectory conserves energy.

The fluctuation data are presented in Figure 3. Curves for methods with $h \geq 10$ are indistinguishable, and only $h = 10$ curves are shown. Several conclusions can be drawn from these data:

- Window screening makes trajectories with loose thresholds more robust against SCF convergence failures, in exchange for a modest increase in the mean fluctuation. In particular, the `win-i-6,5` curve is complete, while the `hard-i-6` curve is missing several points. Further, some of the `win-i-4,3` trajectories ran to completion, unlike those using `hard-i-4`.

- Gradient consistency has a qualitative effect on energy conservation in loosely-screened trajectories. Fluctuations in `hard-c-6` were reduced by an order of magnitude relative to `hard-i-6`, while fluctuations in `win-c-6,5` and `win-c-4,3` were reduced by a factor of 100 relative to their inconsistently-screened counterparts.

- Window screening with a consistent gradient is surprisingly stable — all of the `win-c-4,3` trajectories ran to completion.

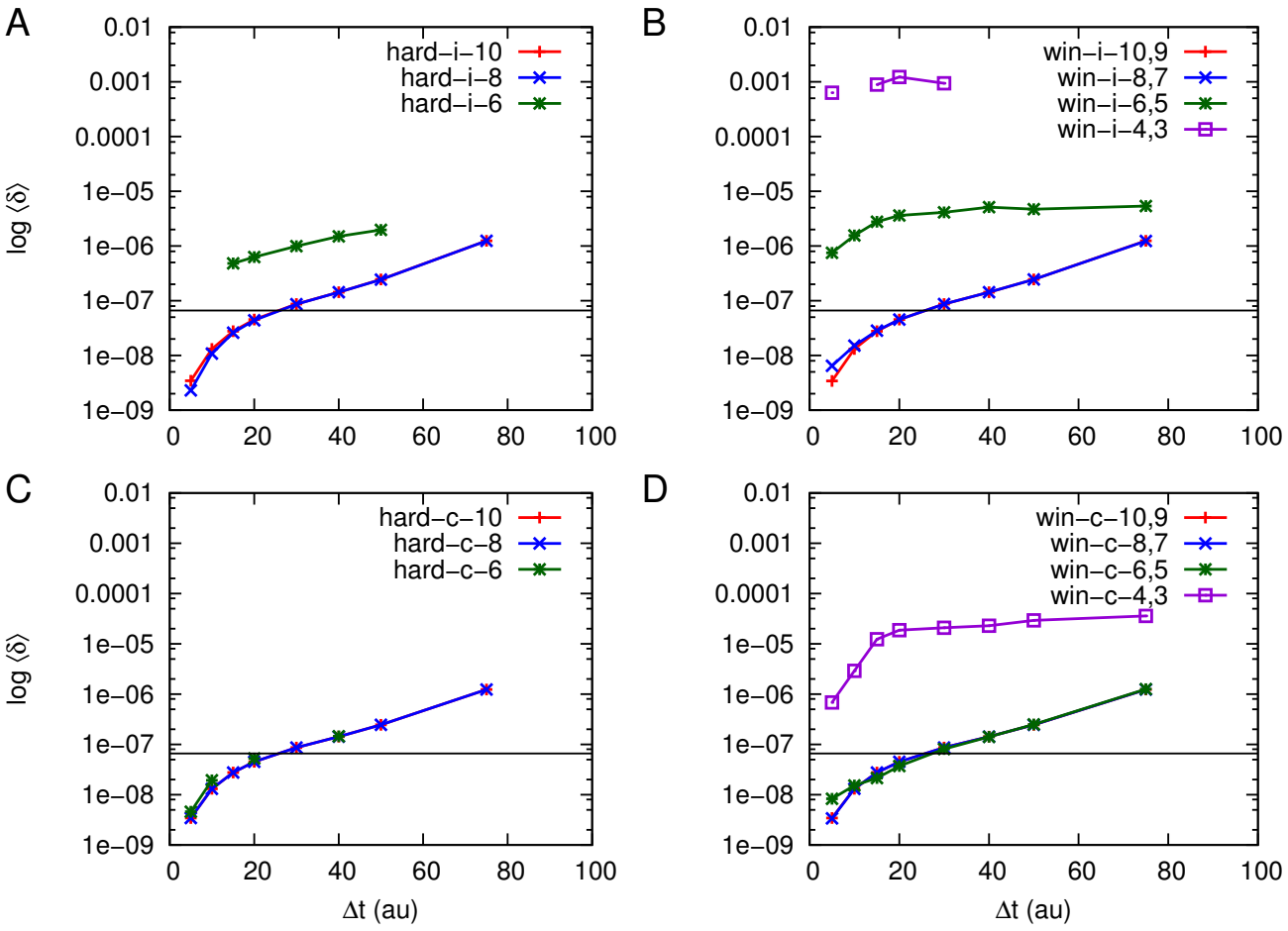



FIG. 3. Mean energy fluctuations $\langle\delta\rangle$ (eq 31) in completed HF/3-21G water-dimer trajectories as a function of timestep and screening method. Panels A and B show fluctuations for `hard` and `win` screening methods with inconsistent gradients, respectively; Panels C and D, with consistent gradients. The horizontal line in each panel indicates the threshold below which the magnitude of the mean energy fluctuation will be smaller than 10 $\mu\mathrm{E}_h$. See Section II D for detailed discussion.

The magnitude of the mean energy fluctuation provides a rational basis for selecting the inner timestep (denoted $\delta t$) in screening-based MTS: $\delta t$ is chosen such that the mean energy fluctuation will be smaller than 10 $\mu\mathrm{E_h}$, equivalent to a 1% fluctuation in the effective temperature of the microcanonical system. Keeping in mind that the total energy of this test system is roughly 151 $\mathrm{E_h}$, a timestep is desired with relative fluctuations smaller than about $7 \times 10^{-8}$, as indicated by the horizontal lines in Figure 3. Examination of the data shows that the largest timestep meeting this criterion is $\delta t = 20$ au (0.484 fs), which is a commonly-used step size for hyrdogen-containing systems.

For the inner- and outer-timestep screening protocol, a pair of methods should be identified which are expected to provide stable, physically sensible trajectories, even for large systems. Because the SCF procedure is converged strictly numerically, and because screened energies and gradients are inexact (independent of any considerations related to windowing), even symplectic integrators can lead to long-time energy drift. (This property is sometimes addressed via extended-Lagrangian MD methods,[48] which directly enforce energy conservation, at the expense of slightly perturbed dynamics.) Consequently, the estimated energy drift over the course of the trajectory is a relevant figure of merit.

To establish whether a given trajectory drifted appreciably, the fluctuation data were fit to a line, $E_t = mt + E_0$, using the Levenberg–Marquardt algorithm implemented in GNUPLOT.[49,50] The estimated unsigned drift over the length of the trajectory was then computed as a fraction of the mean energy fluctuation,

$$\zeta = \frac{|m|\, t_{\max}}{E_1 \langle \delta \rangle}. \tag{32}$$

Perfect energy conservation corresponds to vanishing $m$, or $\zeta = 0$, while $\zeta > 1$ indicates a drift exceeding the typical fluctuations. Estimated drifts for the methods depicted in Figure 3 with $\Delta t = 20$ au are listed in Table I. Note that consistent screening of the gradient leads to significant reductions in the drift, as expected.

Selection of the outer-timestep screening method should be guided by the necessity of describing the system reliably, while the inner-timestep screening should be as loose as possible while maintaining similar quality. The criterion for a high-quality trajectory is that the drift in Table I should be no larger than 1% of the mean energy fluctuation. Consequently, the `hard-i-10` method was chosen as the screening method for the outer timestep, and `win-c-6,5` was chosen for the inner timestep. Because the statistical properties of these methods are similar, the corresponding screening-based MTS scheme is broadly expected to yield stable trajectories; the effect of looser screening in the inner timestep will be explored in Section III B.

To test this expectation, the HF/3-21G water dimer was simulated for 4.84 ps, using between 1 and 10 inner timesteps of fixed size $\delta t = 20$ au. (While not strictly an MTS method, using a single inner timestep is equivalent to performing sequential momentum updates using the partitioned forces and closely tracks the single-timestep `hard-i-10` trajectory.) As Figure 4 shows, fluctuations in the resulting trajectories are effectively constant — the sub-

| Screening | Threshold(s) | $\zeta$ (%) |
|---|---|---|
| `hard-i` | 10 | 0.00 |
| | 8 | 6.41 |
| | 6 | 235.43 |
| `hard-c` | 10 | 0.00 |
| | 8 | 0.00 |
| | 6 | 15.05 |
| `win-i` | 10,9 | 0.00 |
| | 8,7 | 7.73 |
| | 6,5 | 205.53 |
| | 4,3 | 71.01 |
| `win-c` | 10,9 | 0.00 |
| | 8,7 | 0.00 |
| | 6,5 | 0.27 |
| | 4,3 | 296.69 |

TABLE I. Estimated energy drifts $\zeta$ (eq 32) as a percentage of mean energy fluctuations for HF/3-21G water-dimer trajectories integrated with timestep $\Delta t = 20$ au. Drift percentages are not strictly comparable between the `hard` and `win` methods; mean energy fluctuations are generally larger in the latter case, as shown in Figure 3.

tle oscillations are a small-system artifact — and are narrowly distributed around the value observed for the single-timestep method. This encouraging result is not as mysterious as it may at first appear: The `win-c-6,5` inner-timestep method was chosen to exhibit comparable fluctuations to `hard-i-10`, and this behavior is reproduced in the MTS data. The drift is consistent, as well, remaining smaller than $\zeta = 1\%$ until $\Delta t = 140$ au and increasing to no more than a few percent ($\zeta \approx 3.11\%$) at $\Delta t = 200$ au.

The screening-based MTS scheme is now complete — $n$ inner timesteps of length $\delta t = 20$ au are integrated using `win-c-6,5` screening, followed by an outer-timestep momentum correction using `hard-i-10` screening — and it has been shown to be stable. Now this MTS scheme will be justified on physical (rather than purely numerical) grounds, on the basis of an associated timescale separation. To this end, water-dimer configurations were extracted

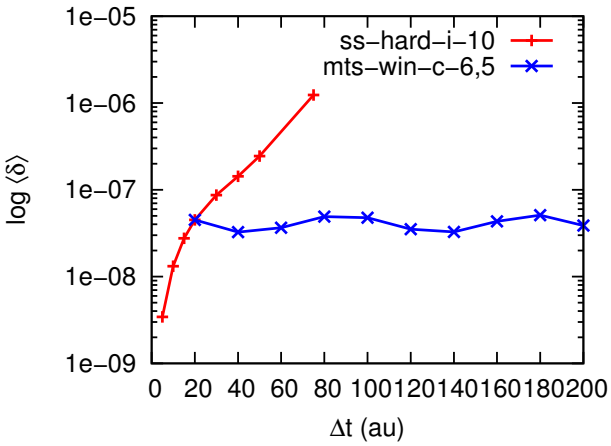



FIG. 4. Mean energy fluctuations relative to the initial energy $E_0$ in HF/3-21G water-dimer multiple-timestep trajectories as a function of outer timestep. The fluctuations are largely determined by the inner-timestep method.

from a single-timestep `hard-i-10` trajectory with $\Delta t = 20$ au; gradient calculations were then performed on these geometries using `win-c-6,5`. The difference between the `hard-i-10` and `win-c-6,5` results is shown for the energy and for a single force component in Figure 5; this difference yields the effective contribution to these quantities from those integrals with estimated magnitude between $10^{-6}$ and $10^{-10}$, with fractional contributions from integrals as large as $10^{-5}$. Thus, any higher-frequency features in the `hard-i-10` energy and force must be associated with integrals of larger magnitude. These results lead to the conclusion that a separation in timescale between larger and smaller integrals does indeed exist, consistent with the original motivation for this approach. As such, screening-based MTS is more than just an *ad hoc* method.

### E. Computational cost considerations for *ab initio* MTS

With the results of Section II D in hand, MTS could immediately be applied to realistic chemical systems. In doing so, however, an important distinction between classical and *ab initio* MTS would be neglected. Suppose that the inner timestep $\delta t$ is appropriate for integrating the dynamics of a system with single-timestep integration methods. The cost of

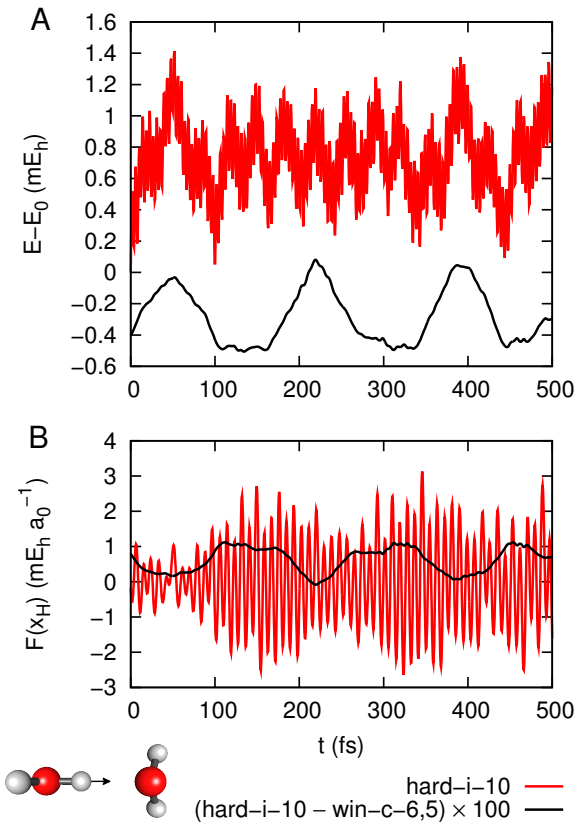



FIG. 5. Phenomenological timescale separation in the energy and a force component of the HF/3-21G water dimer. Panel A shows the energy relative to the initial energy $E_0$ for the `hard-i-10` method (red or grey line) and the energy difference between `hard-i-10` and the same trajectory reevaluated using `win-c-6,5` (scaled up by a factor of 100, black line). Panel B makes a similar comparison for the force component associated with the Cartesian $x$ coordinate of one of the hydrogens, as illustrated at lower left. See Section II D for detailed discussion.

incrementing the system by an outer timestep $\Delta t = n\delta t$ is then

$$\text{cost of increment} = \begin{cases} nS, & \text{single-timestep integrator} \\ nI + O, & \text{multiple-timestep integrator,} \end{cases} \tag{33}$$

where $I$ is the composite cost of the energy and gradient in the inner timestep, $O$ is the cost of the outer timestep, and $S$ is the equivalent cost for the single-timestep method.

In classical MTS, $S$ is simply the cost of evaluating the complete set of analytic energy and gradient functions describing the system; in the *ab initio* case, $S$ is the cost of computing

the energy and gradient at the same level of theory as in the outer timestep. Thus,

$$S = \begin{cases} I + O, & \text{classical MTS} \\ O, & \textit{ab initio} \text{ MTS.} \end{cases} \tag{34}$$

Eqs 33 and 34 imply that classical MTS will be cheaper than the single-timestep equivalent whenever $I < S$; *any* non-trivial partitioning of the system is guaranteed to meet this requirement. The *ab initio* case is qualitatively different: MTS with $n$ inner timesteps will only be cheaper than the single-timestep method when

$$I < \frac{n-1}{n} O. \tag{35}$$

In other words, the repeated work at the outer *ab initio* timestep leads to an overhead, not present in classical MD, that sets a lower bound on the number of inner timesteps that will lead to computational savings. Choosing $n$ to exceed this bound may not be practicable, due to resonance considerations[51–55] or to long physical timescales in a given system. The approach taken here is to dig further into the workings of the electronic-structure methodology in order to minimize this overhead. In particular, by carrying forward information from the inner timesteps, much of the repeated outer-timestep effort can be eliminated. As a result, eq 35 may be rewritten as

$$I < \frac{n-\kappa}{n} O \tag{36}$$

with $\kappa < 1$, subverting what appeared to be a hard inequality. The smaller the value of $\kappa$, the more favorable MTS will be.

### *1. SCF considerations*

In the trivial case that the method used to screen the energies does not change — as is true when only switching gradient consistency, or for correlation-based MTS[20] — the inner- and outer-timestep energies will be identical. Thus, the SCF procedure can be skipped altogether, and the gradient can be computed using the inner-timestep density, $\mathbf{P}_i$. Of course, the present `hard-i-10/win-c-6,5` MTS scheme is not so simple; this possibility is mentioned strictly for completeness.

A more relevant and general improvement involves the initial SCF guess. $\mathbf{P}_i$ is an approximation to the desired outer-timestep density, $\mathbf{P}_o$; it can, therefore, be used as the initial

guess for $\mathbf{P}_o$. Although the initial Fock matrix will be more expensive to construct than that associated with a sparse guess formed by superposition of atomic densities (Q-Chem's default), the total number of SCF cycles at the higher level of theory will be appreciably reduced.

### *2. DFT quadrature considerations*

In calculations using density functional theory, the exchange-correlation functional must be integrated in order to obtain the exchange-correlation energy:

$$E_{\mathrm{XC}} = \int \mathrm{d}\mathbf{r}\, \mathcal{F}\left[\rho(\mathbf{r})\right], \tag{37}$$

where $\mathcal{F}$ is the density functional and the the real-space density is

$$\rho(\mathbf{r}) = \sum_{\mu\nu} P_{\mu\nu}\phi^*_\mu(\mathbf{r})\phi_\nu(\mathbf{r}). \tag{38}$$

Because eq 37 involves a continuous integral over many atomic centers, it is difficult to evaluate straightforwardly. Thus, it is customarily decomposed into a sum of discretized, single-center integrals,[56]

$$E_{\mathrm{XC}} \approx \sum_{A}^{\text{nuclei}} \sum_{i}^{\text{electrons}} \omega_{Ai}\mathcal{F}\left[\rho(\mathbf{r}_{Ai})\right], \tag{39}$$

where grid points $\mathbf{r}_{Ai}$ and Becke weights $\omega_{Ai}$ have been introduced. The Becke weights specify the partitioning of real space among individual "fuzzy" atoms;[57] taken together, the points $\{\mathbf{r}_{Ai}\}$ form the quadrature grid. Instead of reconstructing this threshold-independent grid information (as well as the Becke-weight derivatives, which are required to construct the exchange-correlation vectors in the Kohn–Sham Fock matrix), this information can be reused in the outer timestep.

Because density functionals are typically nonlinear and often extremely complicated, a common approach to reducing the cost of Gaussian quadrature is to neglect contributions to the density $\rho(\mathbf{r})$ from density-matrix elements smaller than some threshold $\varepsilon_{\mathrm{quad}}$. This threshold, much like the threshold for shell-pair economization, remains fixed throughout the present calculations, so the exchange-correlation energy and vectors need not be recalculated when building the initial Fock matrix; the inner-timestep results can be used instead. A small Fock matrix error is incurred in this process — the readily available set of inner-timestep exchange-correlation vectors is obtained by quadrature from an *almost*-converged

density, $\mathbf{P}_a$, rather than $\mathbf{P}_i$ — but this error will be rectified in subsequent SCF cycles. A reasonable alternative to reusing grid information, then, is simply to skip quadrature altogether in the first SCF cycle of the outer timestep.

### *3. Bandpass screening*

The improved SCF guess described in Section II E 1 reduces the number of SCF cycles in the outer timestep considerably. Accordingly, the initial outer-timestep SCF cycle can be associated with an appreciable fraction of the cost of the MD step as a whole, and measures reducing that cost may be useful. When the outer timestep of a screening-based MTS increment is entered, Coulomb and exchange matrices from the lower level of theory remain on hand. If the increment in these matrices associated with the change in screening can be accurately computed — in a similar spirit to standard Fock-matrix increments[58,59] — additional savings in the initial Fock build can be obtained. The procedure by which this aim will be accomplished is called *bandpass screening*.

At convergence in the last inner timestep, the Coulomb and exchange matrices are given by the contraction

$$\mathbf{F}_i^{2e^-} = \mathbf{P}_i \cdot \mathbf{\Pi}_i^{(i)}, \tag{40}$$

where $\mathbf{\Pi}_i^{(i)}$ is the weighted set of two-electron integrals that survive screening of $\mathbf{P}_i \cdot \mathbf{\Pi}$ by the chosen inner-timestep method. (Note that when $\mathbf{\Pi}$ appears without labels, it represents the complete, *unweighted* set of two-electron integrals.) Because $\mathbf{P}_i$ is used as an initial guess in the subsequent outer timestep, the Coulomb and exchange contributions to the initial Fock matrix are

$$\text{initial } \mathbf{F}_o^{2e^-} = \mathbf{P}_i \cdot \mathbf{\Pi}_i^{(o)}, \tag{41}$$

where $\mathbf{\Pi}_i^{(o)}$ is the weighted set of two-electron integrals that survive screening of $\mathbf{P}_i \cdot \mathbf{\Pi}$ by the chosen *outer*-timestep method.

A formal connection between these sets can be made,

$$\mathbf{\Pi}_i^{(o)} \equiv \mathbf{\Pi}_i^{(i)} + \Delta\mathbf{\Pi}_i^{(oi)}, \tag{42}$$

such that

$$\text{initial } \mathbf{F}_o^{2e^-} = \mathbf{P}_i \cdot \mathbf{\Pi}_i^{(i)} + \mathbf{P}_i \cdot \Delta\mathbf{\Pi}_i^{(oi)}. \tag{43}$$

$\Delta\mathbf{\Pi}_i^{(oi)}$ accounts for the reweighting of integrals when the screening method is changed, subject to a fixed density. To be precise, each member of $\mathbf{\Pi}_i^{(i)}$ has a weight $w_i^{(i)} \equiv w(\overline{f}_i; h^{(i)}, s^{(i)})$, where $\overline{f}_i \equiv -\log \overline{F}(\mathbf{P}_i)$, and similarly for members of $\mathbf{\Pi}_i^{(o)}$. The weight of a member of $\Delta\mathbf{\Pi}_i^{(oi)}$ is therefore given by

$$\Delta w_i^{(oi)} = \begin{cases} 0, & \overline{f}_i \leq \min(s^{(i)}, s^{(o)}) \\ w_i^{(o)} - w_i^{(i)}, & \min(s^{(i)}, s^{(o)}) < \overline{f}_i < \max(h^{(i)}, h^{(o)}) \\ 0, & \overline{f}_i \geq \max(h^{(i)}, h^{(o)}), \end{cases} \tag{44}$$

where $\{h^{(i)}, s^{(i)}\}$ and $\{h^{(i)}, s^{(i)}\}$ are the hard cutoff and soft threshold of the inner- and outer-timestep screening methods, respectively. As shown in Fig. 2B, this weight function excludes integrals estimated to be above the larger soft threshold *or* below the smaller hard cutoff — in other words, it is a bandpass filter designed to retain only those integrals weighted differently by the two screening methods. Bandpassing will be most beneficial when the inner- and outer-timestep screening methods assign different weights to a relatively small range of integrals.

In developing this bandpass idea, $\mathbf{P}_i \cdot \mathbf{\Pi}_i^{(i)}$ was assumed to be readily available in the outer timestep. In Q-Chem, the last available Coulomb and exchange matrices are instead those associated with the *almost-converged* density from the penultimate inner-timestep SCF cycle, $\mathbf{P}_a$. The almost- and fully-converged densities (and the corresponding integral sets) can be related using the definitions

$$\mathbf{P}_i \equiv \mathbf{P}_a + \Delta\mathbf{P}_{ia} \tag{45a}$$

$$\mathbf{\Pi}_i^{(i)} \equiv \mathbf{\Pi}_a^{(i)} + \Delta\mathbf{\Pi}_{ia}^{(i)}. \tag{45b}$$

$\Delta\mathbf{\Pi}_{ia}^{(i)}$ accounts for the reweighting of integrals when the density changes with fixed screening method, and its members have bandpass weights different from those given in eq 44, namely,

$$\Delta w_{ia}^{(i)} = \begin{cases} 0, & \overline{f}_i, \overline{f}_a \leq s^{(i)} \\ w_i^{(i)} - w_a^{(i)}, & \text{in all other cases} \\ 0, & \overline{f}_i, \overline{f}_a \geq h^{(i)}. \end{cases} \tag{46}$$

These weights are *not* the same as those obtained by screening $\Delta\mathbf{P}_{ia} \cdot \mathbf{\Pi}$ — the weight function is nonlinear.

Combining eqs 43 and 45, the outer-timestep Coulomb and exchange matrices can be rewritten as

$$\text{initial } \mathbf{F}_o^{2e^-} = \mathbf{P}_a \cdot \mathbf{\Pi}_a^{(i)} + \mathbf{P}_i \cdot \Delta\mathbf{\Pi}_{ia}^{(i)} + \Delta\mathbf{P}_{ia} \cdot \mathbf{\Pi}_a^{(i)} + \mathbf{P}_i \cdot \Delta\mathbf{\Pi}_i^{(oi)}. \tag{47}$$

As long as SCF convergence parameters are sufficiently tight (such as the $10^{-8}$ $E_h$ setting used in this work), the density difference between the final SCF cycles of the inner timestep will be small enough that the cross terms involving both densities can be discarded. The resulting bandpass prescription is very similar to eq 43,

$$\text{initial } \mathbf{F}_o^{2e^-} = \mathbf{P}_a \cdot \mathbf{\Pi}_a^{(i)} + \mathbf{P}_i \cdot \Delta\mathbf{\Pi}_i^{(oi)}, \tag{48}$$

and can be implemented with minor modifications to existing SCF routines.

Finally, a bandpass-screening approach for the two-electron-integral contribution to the gradient is formally possible, but this approach is not cost-effective. Because this bandpassing would involve contractions with the difference in *converged* densities from the inner and outer timesteps — which, in general, are not sufficiently small to discard — this approach would lead to prohibitive additional cost.

## III. RESULTS

With the formulation, validation, and cost analysis of the screening-based MTS scheme completed, application of the method to a challenging chemical system is now possible. Here, the MTS scheme — as implemented in a development version of Q-Chem[34] — is applied to a biological model, the protonated sarcosine-glycine dipeptide embedded in a 19-water cluster. Although this choice of model was inspired by recent experiments using gas-phase $SarGlyH^+$ as a testbed for understanding the structural and dynamical effects of peptide methylation,[60,61] these phenomena are not addressed in this work. Rather, a simple demonstration is made that screening-based MTS realizes significant computational savings for a reasonably large system containing 29 heavy atoms and 268 electrons.

The ground-state dynamics of $SarGlyH^+(H_2O)_{19}$ were simulated at both the Hartree–Fock and B3LYP levels of theory using the 6-31G** basis set (corresponding to a total of 680 basis functions) and a serial implementation of the underlying electronic structure methodology. The initial configuration of the system, shown in Figure 6, was obtained by

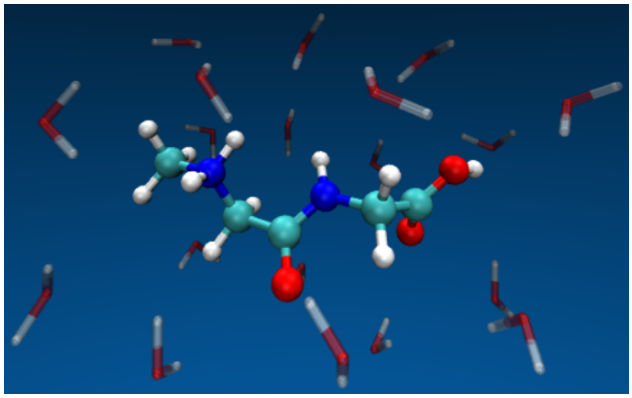


FIG. 6. Protonated-sarcosine/glycine dipeptide embedded in a 19-water cluster.

building SarGlyH$^+$ in IQmol;[62] optimizing the structure with the MMFF94s force field;[63] inserting the optimized dipeptide in the HF/6-31G** minimum-energy structure of $(H_2O)_{19}$ from the Cambridge Cluster Database;[45,46] and reoptimizing with MMFF94s. Initial velocities were sampled from the Maxwell–Boltzmann distribution at 298 K. As before, the base inner timestep for the MTS trajectories was $\delta t = 20$ au. Timesteps were screened using `hard-i-10` (outer) and `win-c-6,5` (inner) methods. All trajectories were propagated for at least 480 au (11.6 fs); for purposes of timing comparison, single-timestep simulations using `hard-i-10` and `win-c-6,5` were also performed.

### A. Outer-timestep efficiencies are large

In order to assess the usefulness of the strategies proposed in Section II E for reducing the cost of the `hard-i-10` outer timestep, a series of simulations was performed in which the algorithmic improvements were added hierarchically. Consequently, three types of Hartree–Fock calculation were performed:

1. no-band: A Hartree–Fock calculation with no improvements.

2. only-r: The inner-timestep density $\mathbf{P}_i$ is used as a guess.

3. both-rb: Bandpass screening is applied, in addition to only-r.

Because DFT includes additional efficiencies related to the quadrature grid, there were five types of B3LYP calculation:

1. no-band: A B3LYP calculation with no improvements.

2. only-g: Quadrature-grid information is reused.

3. both-gr: $\mathbf{P}_i$ is used as a guess, in addition to only-g.

4. gr-and-b: Bandpass screening is applied, in addition to both-gr.

5. full-grbq: DFT quadrature results are reused, in addition to gr-and-b.

As the comparisons in Figure 7 show, reusing the density overwhelms all other efficiencies, reducing the cost of the outer timestep roughly by half for both Hartree–Fock and B3LYP. Additional savings accrue in B3LYP calculations when quadrature information is reused, reducing the cost by a further $\approx 12\%$. While bandpassing produces more modest savings, the aggregate effect of a few seconds' saving per outer timestep can be significant for long trajectories.

Figure 7 shows that an HF/both-rb outer timestep is comparable in cost to an inner timestep. Further, B3LYP/full-grbq outer timesteps are actually *cheaper*, such that the inner timesteps become the computational bottleneck. In both cases, these steep reductions stem from the fact that much of the effort required to converge the `hard-i-10` calculation has already been done in the preceding inner timestep. Therefore, the goal of addressing outer-timestep "overhead," as discussed in Section II E, has largely been met.

## B. Minimal multiple-timestepping realizes significant savings

Table II lists the average CPU minutes required to simulate the $\mathrm{SarGlyH^+(H_2O)_{19}}$ system for a femtosecond as a function of the number of inner timesteps, $n$. The ratio between single-timestep timings for `hard-i-10` and `win-c-6,5` provides an estimate of the maximum theoretically-achievable acceleration in these calculations — roughly 87% for Hartree–Fock and 35% for B3LYP. The smaller percentage for B3LYP reflects the fact that DFT quadrature adds what amounts to a fixed cost in each and every SCF cycle (apart from the initial Fock build in the outer timestep).

Figure 8 presents the same data graphically, clearly showing that roughly a fifth of the achievable speed-up for Hartree–Fock is realized with a minimal MTS scheme ($n = 2$), and more than half is realized when $n = 4$. (The same trend holds for B3LYP.) These results are due entirely to the overhead-trimming interventions in the outer timestep, as evidenced

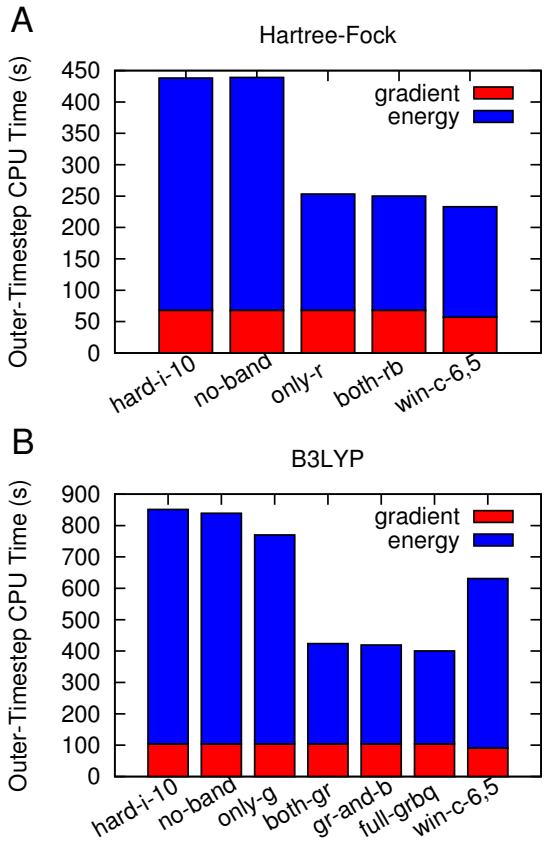



FIG. 7. Average CPU seconds required to compute the `hard-i-10` outer-timestep energy and gradient in AIMD simulations of the $\mathrm{SarGlyH^+(H_2O)_{19}}$ system using Hartree–Fock and B3LYP in 6-31G** basis set. `hard-i-10` and `win-c-6,5` denote costs from single-timestep trajectories, while the other labels indicate the cost of the outer timestep in screening-based MTS trajectories exploiting efficiencies to differing degrees. See Section III A for detailed discussion.

by naïve MTS calculations costing more than a single-timestep calculation until $n = 3$ for HF or $n = 4$ for B3LYP. Reusing information in the outer timestep has effectively skirted the cost inequality that governs *ab initio* MTS (eq 35); the effective values of the constant $\kappa$ in eq 36 derived from our simulations are $\kappa_{\mathrm{HF}} \approx 0.57$ and $\kappa_{\mathrm{B3LYP}} \approx 0.47$. These constants are not necessarily transferable between systems, but future analysis will establish whether they are typical.

| $n$ | Method | |
|---|---|---|
| | Hartree–Fock | B3LYP |
| 1h | 15.10 | 29.37 |
| 2 | 12.44 | 28.49 |
| 3 | 11.32 | 26.22 |
| 4 | 10.23 | 25.10 |
| 5 | 9.82 | 24.21 |
| 6 | 9.55 | 23.79 |
| 7 | 9.36 | 23.19 |
| 8 | 9.22 | 23.25 |
| 9 | 9.10 | 22.53 |
| 10 | 9.06 | 22.18 |
| 11 | 8.97 | 21.91 |
| 12 | 8.96 | 22.35 |
| 1w | 8.07 | 21.73 |

TABLE II. Average CPU minutes per simulated fs as a function of the number of inner timesteps $n$ for fully-bandpassed (HF/both-rb or B3LYP/full-grbq) screening-based MTS simulations of the $SarGlyH^+(H_2O)_{19}$ system in the 6-31G** basis set. "1h" and "1w" denote single-timestep simulations using `hard-i-10` and `win-c-6,5`, respectively.

### C. Looser inner-timestep screening can provide further savings

The `hard-i-10/win-c-6,5` MTS scheme was constructed to satisfy fairly stringent criteria for statistical properties of the corresponding single-timestep trajectories on a picosecond timescale. As noted in Section II D, looser screening in the inner timestep may still yield quality results to the extent that inaccuracies in the inner timestep are corrected by frequent-enough outer timesteps. Here, the degree to which loosening the threshold affects the savings achievable for $SarGlyH^+(H_2O)_{19}$ is examined.

Looser screening reduces the time required for each SCF cycle, but the errors introduced into the density can make it harder to achieve SCF convergence, such that the average number of cycles increases. For any given system, then, a regime may exist in which loos-

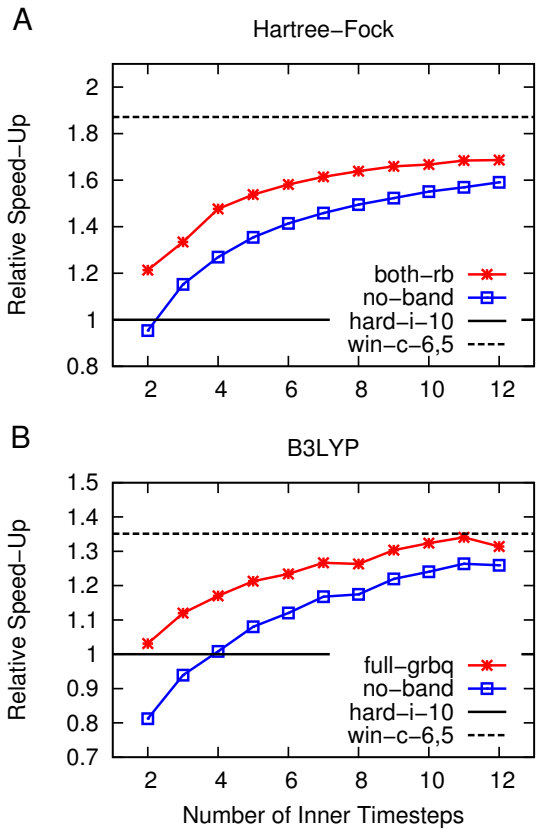



FIG. 8. The speed-up of screening-based MTS simulations of the SarGlyH$^+$(H$_2$O)$_{19}$ system in the 6-31G** basis set. Red stars indicate the fully-bandpassed results (HF/both-rb or B3LYP/full-grbq, cf. Table II), while blue boxes are data for simulations without any use of efficiencies in the outer timestep. The solid and dashed black lines indicate the single-timestep cost of the `hard-i-10` and `win-c-6,5` methods, respectively. Results below the black line indicate cases where the cost of the MTS simulation would be more expensive than a single-timestep without bandpassing.

ening the threshold makes the calculation more expensive. As a case in point, SCF did not converge when a `win-c-4,3` was used for SarGlyH$^+$(H$_2$O)$_{19}$. Therefore, narrower windows were tried — $s \in \{3.12, 3.3, 3.6\}$, corresponding roughly to $2.5 \times 10^{-4}$ steps in integral magnitude between the hard cutoff and soft threshold. (A narrower window reduces errors associated with integral rescaling, which could restore convergence, but it also reduces robustness against blinking.) The narrow-window B3LYP calculations never converged, while the HF calculations were costlier than `win-c-6,5`, owing to a sharp increase in the number of SCF cycles required (as much as a factor of 6.5). The aggressive `win-c-4,3` method was ultimately abandoned for the present application, and `win-c-5,4` was tested instead. While

| $n$ | Method | |
|---|---|---|
| | Hartree–Fock | B3LYP |
| 1h | 15.10 | 29.37 |
| 2 | 11.78 | 32.62 |
| 3 | 10.27 | 29.64 |
| 4 | 9.09 | 27.91 |
| 5 | 8.52 | 26.83 |
| 6 | 8.21 | 26.24 |
| 7 | 7.93 | 25.49 |
| 8 | 7.73 | 25.39 |
| 9 | 7.66 | 24.86 |
| 10 | 7.48 | 24.33 |
| 11 | 7.36 | 24.16 |
| 12 | 7.29 | 24.58 |
| 1w | 6.42 | 23.34 |

TABLE III. Average CPU minutes per simulated fs as a function of the number of inner timesteps $n$ for fully-bandpassed (HF/both-rb or B3LYP/full-grbq) screening-based MTS simulations of the $SarGlyH^+(H_2O)_{19}$ system in the 6-31G** basis set. "1h" and "1w" denote single-timestep simulations using `hard-i-10` and `win-c-5,4`, respectively.

both HF and B3LYP calculations converged, only the HF trajectories were cheaper than the corresponding `win-c-6,5` calculations — by about 20%, as shown in Table III.

Another consequence of looser screening in the inner timestep is that the various efficiencies discussed in Section II E can be less useful, as shown in Figure 9. In particular, because the density from the last inner timestep is not as good of a guess for the outer-timestep density, the number of SCF cycles required is not reduced as sharply. These more modest savings — equivalent to $\kappa_{\mathrm{HF}} \approx 0.67$ and $\kappa_{\mathrm{B3LYP}} \approx 0.59$ — are compensated in the case of Hartree–Fock by the fact that the inner timestep cost is significantly reduced. As a result, `win-c-5,4/hard-i-10` MTS always involves some speed-up, as shown in Figure 10A, and about half of the roughly 135% theoretically-achievable speed-up is realized when $n = 4$. By

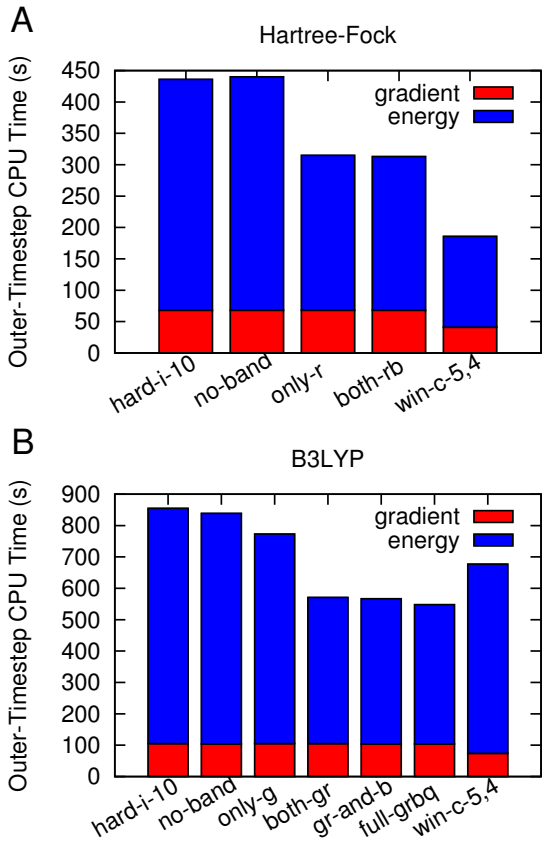



FIG. 9. Average CPU seconds required to compute the `hard-i-10` outer-timestep energy and gradient in AIMD simulations of the SarGlyH$^+$(H$_2$O)$_{19}$ system using Hartree–Fock and B3LYP in the 6-31G** basis set. `hard-i-10` and `win-c-5,4` denote costs from single-timestep trajectories, while the other labels indicate the cost of the outer timestep in screening-based MTS trajectories exploiting efficiencies to differing degrees. See Section III A for detailed discussion.

contrast, $n = 4$ is the minimum number of inner timesteps required for B3LYP to achieve even a modest speed-up, as shown in Figure 10B. In light of these results, a wise practice would be to compare timings for a small number of possible inner-timestep methods and single increments of the corresponding MTS protocol *before* committing to a specific scheme.

Finally, it is worth noting that these timing assessments are fairly conservative, due to the fact that screening-based multiple-timestepping has been evaluated with all other parameters fixed. Algorithmic improvements in the outer timestep led to a situation in which the inner timesteps are cost-dominant — an inversion of the usual relationship. Because a much looser threshold is employed when screening these inner steps, a looser SCF convergence criterion could likely be employed as well, reducing the inner-timestep cost and increasing

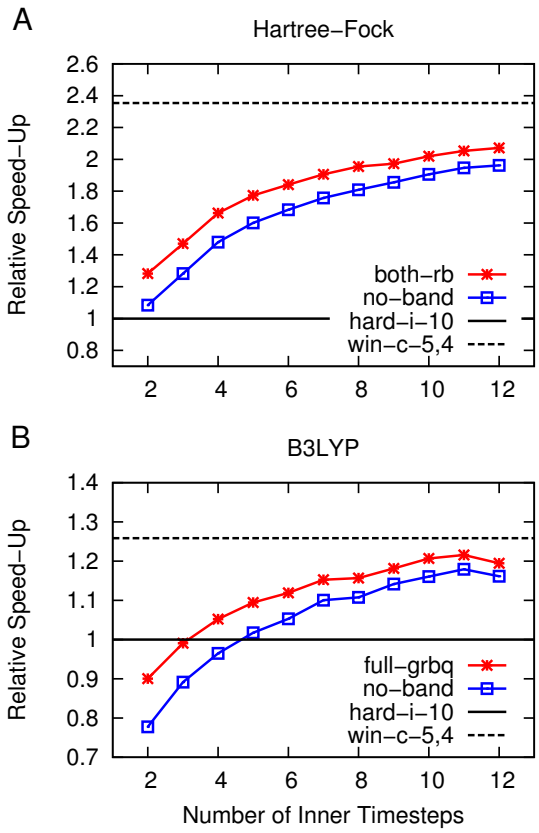



FIG. 10. The speed-up of screening-based MTS simulations of the SarGlyH$^+$(H$_2$O)$_{19}$ system in the 6-31G** basis set. Red stars indicate the fully-bandpassed results (HF/both-rb or B3LYP/full-grbq, cf. Table III), while blue boxes are data for simulations without any use of efficiencies in the outer timestep. The solid and dashed black lines indicate the single-timestep cost of the `hard-i-10` and `win-c-5,4` methods, respectively. Results below the black line indicate cases where the cost of the MTS simulation is more expensive than the single-timestep method.

the theoretically-achievable speed-up. Established AIMD extrapolation techniques (such as Fock-matrix extrapolation[64,65]) could also potentially be applied to reduce the cost of the inner timesteps. Finally, the shorter spatial extent of the inner-timestep thresholding suggests that linearly scaling SCF algorithms would likely become effective sooner, with respect to system size, than the analogous tight-threshold outer step. In this large-system regime, relative timings would likely improve further.

## IV. CONCLUDING REMARKS

This work addressed the question of whether multiple-timestep methods could be applied to Hartree–Fock or DFT molecular dynamics on a firmer-than-*ad-hoc* basis. This question was answered affirmatively by showing that two-electron integrals of disparate magnitude are associated with forces that vary on disparate timescales. In the process, a multiple-timestep scheme was formulated using classic hard-cutoff screening techniques and a window-screening method of our own devising. Once efficiencies in the outer-timestep were identified and exploited, this multiple-timestep protocol realized computational savings with minimal or small numbers of inner timesteps ($n = 2 - -4$).

The Schwarz-based integral estimates in this work highlight only one possibility for screening-based multiple-timestep AIMD; the same principles should apply to more sophisticated screening protocols, such as those based on multipole expansions. In particular, screening methods which explicitly account for the asymptotic decay of interactions between $\mu\nu$ and $\lambda\sigma$ orbital pairs are likely to introduce additional opportunities for dynamics acceleration: Because the corresponding inequalities provide tighter bounds on integral magnitudes, inner timesteps may be significantly cheaper; at the same time, the corresponding range and timescale separations may be more pronounced.

The field of multiple-timestep Born–Oppenheimer molecular dynamics is still quite young, and many possible correspondences between electronic-structure parameters and physical timescale separations remain to be tested. Varying these parameters is expected to lead to a new appreciation of hidden pathologies in commonly used AIMD protocols, just as varying the threshold highlighted blinking integrals and gradient inconsistency. For example, the SCF procedure is not iterated to exact convergence; as a result, the dynamics do not strictly preserve phase-space area, leading to drift.[3,66] Varying the SCF convergence criterion is likely to exacerbate this problem, and stable trajectories may only be achievable within the rigorously time-reversible approach of Niklasson *et al.*[66]

As the number of validated resources for multiple-timestep Hartree–Fock and DFT increases, combining these methods directly may become useful. Combined methods may be constructed by varying several parameters simultaneously, or in hierarchical fashion, with inner timesteps subject to further Trotter splitting. The latter scenario requires careful implementation, because the MTS framework must be flexible enough to combine different

protocols to an arbitrary hierarchical depth while also allowing for the use of thermostats, which are fully compatible with the r-RESPA formalism[9,67,68].

Omnibus methods of either type may be viable for outer timesteps well above the onset of resonance ($\Delta t \gtrsim 160$ au $\approx 4$ fs).[51–55] Indeed, Figure 4 indicates that screening-based multiple-timestepping already fits this description, although any resonance-related drift remains modest. Translating existing resonance-resistant MTS methods[69–78] into the *ab initio* context may open the way to efficient, accurate, genuinely long-timescale *ab initio* molecular dynamics.

## ACKNOWLEDGMENTS

Support and resources from the Center for High Performance Computing at the University of Utah are gratefully acknowledged; the authors also used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation Grant #ACI-1053575. RPS further acknowledges start-up research funds from the University of Utah. The authors thank Peter Gill, Martin Head-Gordon, John Herbert, Daniel Lambrecht, Yihan Shao, and Mark Tuckerman for helpful discussions.

## REFERENCES

[1]Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*, 1st ed.; Clarendon Press: Oxford, 1989.

[2]Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed.; Academic Press: San Diego, CA, 2002.

[3]Marx, D.; Hutter, J. *Modern Methods and Algorithms of Quantum Chemistry*; NIC Series; John von Neumann Institute of Computing: Jülich, Germany, 2000; Vol. 1; p 301.

[4]Tuckerman, M. E. *Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms*; NIC Series; John von Neumann Institute of Computing: Jülich, Germany, 2002; Vol. 10; p 299.

[5]Sagui, C.; Darden, T. A. Molecular dynamics simulations of biomolecules: long-range electrostatic effects. *Annu. Rev. Bioph. Biom.* **1999**, *28*, 155.

[6]Zhou, R.; Berne, B. J. A new molecular dynamics method combining the reference system

propagator algorithm with a fast multipole method for simulating proteins and other complex systems. *J. Chem. Phys.* **1995**, *103*, 9444.

[7]Procacci, P.; Marchi, M. Taming the Ewald sum in molecular dynamics simulations of solvated proteins via a multiple time step algorithm. *J. Chem. Phys.* **1996**, *104*, 3003.

[8]Zhou, R.; Harder, E.; Xu, H.; Berne, B. J. Efficient multiple time step method for use with Ewald and particle mesh Ewald for large biomolecular systems. *J. Chem. Phys.* **2001**, *115*, 2348.

[9]Tuckerman, M.; Berne, B. J.; Martyna, G. J. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.* **1992**, *97*, 1990.

[10]Streett, W. B.; Tildesley, D. J.; Saville, G. Multiple time-step methods in molecular dynamics. *Mol. Phys.* **1977**, *35*, 639.

[11]Swindoll, R. D.; Haile, J. M. A multiple time-step method for molecular dynamics simulations of fluids of chain molecules. *J. Comput. Phys.* **1984**, *53*, 289.

[12]Grubmüller, H.; Heller, H.; Windemuth, A.; Schulten, K. Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions. *Mol. Simulat.* **1991**, *6*, 121.

[13]Tuckerman, M. E.; Martyna, G. J.; Berne, B. J. Molecular dynamics algorithm for condensed systems with multiple time scales. *J. Chem. Phys.* **1990**, *93*, 1287.

[14]Tuckerman, M. E.; Berne, B. J.; Rossi, A. Molecular dynamics algorithm for multiple time scales: systems with disparate masses. *J. Chem. Phys.* **1991**, *94*, 1465.

[15]Tuckerman, M. E.; Berne, B. J.; Martyna, G. J. Molecular dynamics algorithm for multiple time scales: systems with long range forces. *J. Chem. Phys.* **1991**, *94*, 6811.

[16]Anglada, E.; Junquera, J.; Soler, J. M. Efficient mixed-force first-principles molecular dynamics. *Phys. Rev. E* **2003**, *68*, 055701(R).

[17]Guidon, M.; Schiffmann, F.; Hutter, J.; VandeVondele, J. *Ab initio* molecular dynamics using hybrid density functionals. *J. Chem. Phys.* **2008**, *128*, 214104.

[18]Gibson, D. A.; Carter, E. A. Time-reversible multiple time scale ab initio molecular dynamics. *J. Chem. Phys.* **1993**, *97*, 13429–13434.

[19]Tuckerman, M. E.; Parrinello, M. Integrating the Car-Parrinello equations. II. Multiple time scale techniques. *J. Chem. Phys.* **1994**, *101*, 1316–1329.

[20]Steele, R. P. Communication: Multiple-timestep *ab initio* molecular dynamics with electron correlation. *J. Chem. Phys.* **2013**, *139*, 011102.

[21]Møller, C.; Plesset, M. S. Notes on an approximation treatment for many-electron systems. *Phys. Rev.* **1934**, *46*, 618.

[22]Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry*; Dover Publishing: Mineola, NY, 1996.

[23]Head-Gordon, M.; Pople, J. A.; Frisch, M. J. MP2 energy evaluation by direct methods. *Chem. Phys. Lett.* **1988**, *153*, 503.

[24]Ochsenfeld, C.; Kussmann, J.; Lambrecht, D. S. Linear-scaling methods in quantum chemistry. *Rev. Comp. Ch.* **2007**, *23*, 1.

[25]Luehr, N.; Markland, T. E.; Martínez, T. J. Multiple time step integrators in *ab initio* molecular dynamics. *J. Chem. Phys.* **2014**, *140*, 084116.

[26]Adamson, R. D.; Dombroski, J. P.; Gill, P. M. W. Chemistry without Coulomb tails. *Chem. Phys. Lett.* **1996**, *254*, 329.

[27]Adamson, R. D.; Dombroski, J. P.; Gill, P. M. W. Efficient calculation of short-range Coulomb energies. *J. Comput. Chem.* **1999**, *20*, 921.

[28]Almlöf, J.; Faegri, Jr., K.; Korsell, K. Principles for a direct SCF approach to LCAO-MO *ab-initio* calculations. *J. Comput. Chem.* **1982**, *3*, 385.

[29]Häser, M.; Ahlrichs, R. Improvements on the direct SCF method. *J. Comput. Chem.* **1989**, *10*, 104.

[30]Panas, I.; Almlöf, J.; Feyereisen, M. W. Ab initio methods for large systems. *Int. J. Quantum Chem.* **1991**, *40*, 797.

[31]Adamson, R. D. Shell-pair economization. B.Sc. honors thesis, Massey University, 1995.

[32]Gill, P. M. W.; Johnson, B. G.; Pople, J. A. A simple yet powerful upper bound for Coulomb integrals. *Chem. Phys. Lett.* **1994**, *217*, 65.

[33]Horn, H.; Weiß, H.; Häser, M.; Ehrig, M.; Ahlrichs, R. Prescreening of two-electron integral derivatives in SCF gradient and Hessian calculations. *J. Comput. Chem.* **1991**, *12*, 1058.

[34]Shao, Y. et al. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Mol. Phys.* **2014**, DOI: `http://dx.doi.org/10.1080/00268976.2014.952696`.

[35]Gill, P. M. W.; Johnson, B. G.; Pople, J. A. Two-electron repulsion integrals over Gaussian *s* functions. *Int. J. Quantum Chem.* **1991**, *40*, 745.

[36]Pulay, P. Convergence acceleration of iterative sequences. The case of SCF iteration. *Chem. Phys. Lett.* **1980**, *73*, 393.

[37]Pulay, P. Improved SCF convergence acceleration. *J. Comput. Chem.* **1982**, *3*, 556.

[38]Lau, K. F.; Alper, H. E.; Thacher, T. S.; Stouch, T. R. Effects of switching functions on the behavior of liquid water in molecular dynamics simulations. *J. Phys. Chem.* **1994**, *98*, 8785.

[39]Morrone, J. A.; Zhou, R.; Berne, B. J. Molecular dynamics with multiple time scales: avoid pitfalls. *J. Chem. Theory Comput.* **2010**, *6*, 1798.

[40]Gerratt, J.; Mills, I. M. Force constants and dipole-moment derivatives of molecules from perturbed Hartree–Fock calculations. I. *J. Chem. Phys.* **1968**, *49*, 1719.

[41]Pople, J. A.; Krishnan, R.; Schlegel, H. B.; Binkley, J. S. Derivative studies in Hartree–Fock and Møller–Plesset theories. *Int. J. Quantum Chem.* **1979**, *16 S13*, 225.

[42]Fournier, R. Second and third derivatives of the linear combination of Gaussian type orbitals–local spin density energy. *J. Chem. Phys.* **1990**, *92*, 5422.

[43]Komornicki, A.; Fitzgerald, G. Molecular gradients and hessians implemented in density functional theory. *J. Chem. Phys.* **1993**, *98*, 1398.

[44]Maurice, D. Single electron theories of excited states. Ph.D. thesis, University of California, Berkeley, 1998.

[45]Wales, D. J.; Doye, J. P. K.; Dullweber, A.; Hodges, M. P.; Naumkin, F. Y.; Calvo, F.; Hernández-Rojas, J.; Middleton, T. F. The Cambridge Cluster Database. `http://www-wales.ch.cam.ac.uk/CCD.html`, date accessed: October 9, 2014.

[46]Maheshwary, S.; Patel, N.; Sathyamurthy, N.; Kulkarni, A. D.; Garde, S. R. Structure and stability of water clusters $(H_2O)_n$, $n = 8$–20: an *ab initio* investigation. *J. Phys. Chem. A* **2001**, *105*, 10525.

[47]Scuro, S. R.; Chin, S. A. Forward symplectic integrators and the long-time phase error in periodic motions. *Phys. Rev. E* **2005**, *71*, 056703.

[48]Schlegel, H. B. Ab Initio Molecular Dynamics with Born-Oppenheimer and Extended Lagrangian Methods Using Atom Centered Basis Functions. *Bull. Korean Chem. Soc.* **2003**, *24*, 837–842.

[49]Marquardt, D. W. An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Ind. Appl. Math.* **1963**, *11*, 431.

[50]Williams, T.; Kelley, C. GNUPLOT. `http://www.gnuplot.info`, date accessed: October 9, 2014.

[51]Stuart, S. J.; Zhou, R.; Berne, B. J. Molecular dynamics with multiple time scales: the

selection of efficient reference system propagators. *J. Chem. Phys.* **1996**, *105*, 1426.

[52]Littell, T. R.; Skeel, R. D.; Zhang, M. Error analysis of symplectic multiple timestepping. *SIAM J. Numer. Anal.* **1997**, *34*, 1792.

[53]Bishop, T. C.; Skeel, R. D.; Schulten, K. Difficulties with multiple time stepping and fast multipole algorithm in molecular dynamics. *J. Comput. Chem.* **1997**, *18*, 1785.

[54]Barash, D.; Yang, L.; Qian, X.; Schlick, T. Inherent speedup limitations in multiple time step/particle mesh Ewald algorithms. *J. Comput. Chem.* **2003**, *24*, 77.

[55]Ma, Q.; Izaguirre, J. A.; Skeel, R. D. Verlet-I/R-RESPA/Impulse is limited by nonlinear instabilities. *SIAM J. Sci. Comput.* **2003**, *24*, 1951.

[56]Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*; WILEY-VCH Verlag: Weinheim, Germany, 2001.

[57]Becke, A. D. A multicenter numerical integration scheme for polyatomic molecules. *J. Chem. Phys.* **1988**, *88*, 2547.

[58]Cremer, D.; Gauss, J. An unconventional SCF method for calculations on large molecules. *J. Comput. Chem.* **1986**, *7*, 274.

[59]Schwegler, E.; Challacombe, M.; Head-Gordon, M. Linear scaling computation of the Fock matrix. II. Rigorous bounds on exchange integrals and incremental Fock build. *J. Chem. Phys.* **1997**, *106*, 9708.

[60]Leavitt, C. M.; Wolk, A. B.; Kamrath, M. Z.; Garand, E.; Van Stipdonk, M. J.; Johnson, M. A. Characterizing the intramolecular H-bond and secondary structure in methylated GlyGlyH$^+$ with $H_2$ predissociation spectroscopy. *J. Am. Soc. Mass Spectr.* **2011**, *22*, 1941.

[61]Johnson, C. J.; Wolk, A. B.; Fournier, J. A.; Sullivan, E. N.; Weddle, G. H.; Johnson, M. A. Communication: He-tagged vibrational spectra of the SarGlyH$^+$ and $H^+(H_2O)_{2,3}$ ions: quantifying tag effects in cryogenic ion vibrational predissociation (CIVP) spectroscopy. *J. Chem. Phys.* **2014**, *140*, 221101.

[62]Gilbert, A. T. B. IQmol molecular viewer. `http://www.iqmol.org`, date accessed: October 9, 2014.

[63]Halgren, T. A. MMFF VI. MMFF94s option for energy minimization studies. *J. Comput. Chem.* **1999**, *20*, 720.

[64]Herbert, J. M.; Head-Gordon, M. Accelerated, energy-conserving Born-Oppenheimer molecular dynamics via Fock matrix extrapolation. *Phys. Chem. Chem. Phys.* **2005**, *7*,

3269–3275.

[65]Niklasson, A. M. N.; Tymczak, C. J.; Challacombe, M. Time-reversible ab initio molecular dynamics. *J. Chem. Phys.* **2007**, *126*, 144103.

[66]Niklasson, A. M. N.; Tymczak, C. J.; Challacombe, M. Time-reversible Born-Oppenheimer molecular dynamics. *Phys. Rev. Lett.* **2006**, *97*, 123001.

[67]Martyna, G. J.; Klein, M. L.; Tuckerman, M. Nosé–Hoover chains: the canonical ensemble via continuous dynamics. *J. Chem. Phys.* **1992**, *97*, 2635.

[68]Tuckerman, M. E.; Berne, B. J.; Martyna, G. J.; Klein, M. L. Efficient molecular dynamics and hybrid Monte Carlo algorithms for path integrals. *J. Chem. Phys.* **1993**, *99*, 2796.

[69]Izaguirre, J. A.; Reich, S.; Skeel, R. D. Longer time steps for molecular dynamics. *J. Chem. Phys.* **1999**, *110*, 9853.

[70]Ma, Q.; Izaguirre, J. A. Targeted mollified impulse: a multiscale stochastic integrator for long molecular dynamics simulations. *Multiscale Model. Simul.* **2003**, *2*, 1.

[71]Chin, S. A. Dynamical multiple-time stepping methods for overcoming resonance instabilities. *J. Chem. Phys.* **2004**, *120*, 8.

[72]Minary, P.; Tuckerman, M. E.; Martyna, G. J. Long time molecular dynamics for enhanced conformational sampling in biomolecular systems. *Phys. Rev. Lett.* **2004**, *93*, 150201.

[73]Janežič, D.; Praprotnik, M.; Merzel, F. Molecular dynamics integration and molecular vibrational theory. I. New symplectic integrators. *J. Chem. Phys.* **2005**, *122*, 174101.

[74]Jia, Z.; Leimkuhler, B. Geometric integrators for multiple time-scale simulation. *J. Phys. A: Math. Gen.* **2006**, *39*, 5379.

[75]Kräutler, V.; Hünenberger, P. H. A multiple timestep algorithm compatible with a large number of distance classes and an arbitrary distance dependence of the time step size for the fast evaluation of nonbonded interactions in molecular simulations. *J. Comput. Chem.* **2006**, *27*, 1163.

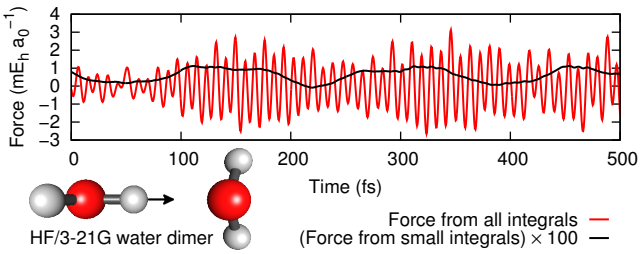
[76]Omelyan, I. P. Advanced multiple time scale molecular dynamics. *J. Chem. Phys.* **2009**, *131*, 104101.

[77]Morrone, J. A.; Markland, T. E.; Ceriotti, M.; Berne, B. J. Efficient multiple time scale molecular dynamics: using colored noise thermostats to stabilize resonances. *J. Chem. Phys.* **2011**, *134*, 014103.

[78]Leimkuhler, B.; Margul, D. T.; Tuckerman, M. E. Stochastic, resonance-free multiple time-step algorithm for molecular dynamics with very large time steps. *Mol. Phys.* **2013**, *111*,

3579.

For Table of Contents Use Only




**Title:** Multiple-timestep *ab initio* molecular dynamics based on two-electron integral screening

**Authors:** Shervin Fatehi and Ryan P. Steele

For Table of Contents Use Only