The Journal of Chemical Physics

Efficient Langevin dynamics for "noisy" forces

Cite as: J. Chem. Phys. **152**, 161103 (2020); https://doi.org/10.1063/5.0004954
Submitted: 19 February 2020 . Accepted: 06 April 2020 . Published Online: 23 April 2020

Eitam Arnon, Eran Rabani, Daniel Neuhauser, and Roi Baer

The Journal of Chemical Physics COMMUNICATION scitation.org/journal/jcp

Efficient Langevin dynamics for "noisy" forces

Cite as: J. Chem. Phys. 152, 161103 (2020); doi: 10.1063/5.0004954
Submitted: 19 February 2020 • Accepted: 6 April 2020 •
Published Online: 23 April 2020

Eitam Arnon,[1] Eran Rabani,[2,a)] Daniel Neuhauser,[3,b)] and Roi Baer[1,c)]

AFFILIATIONS

[1] Fritz Haber Research Center for Molecular Dynamics, Institute of Chemistry, The Hebrew University of Jerusalem, Jerusalem 91904, Israel

[2] Department of Chemistry, University of California and Materials Science Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA and The Raymond and Beverly Sackler Center for Computational Molecular and Materials Sciences, Tel Aviv University, Tel Aviv 69978, Israel

[3] Department of Chemistry and Biochemistry, University of California at Los Angeles, Los Angeles, California 90095, USA

[a)] **Electronic mail:** eran.rabani@berkeley.edu
[b)] **Electronic mail:** dxn@chem.ucla.edu
[c)] **Author to whom correspondence should be addressed:** roi.baer@huji.ac.il

ABSTRACT

Efficient Boltzmann-sampling using first-principles methods is challenging for extended systems due to the steep scaling of electronic structure methods with the system size. Stochastic approaches provide a gentler system-size dependency at the cost of introducing "noisy" forces, which could limit the efficiency of the sampling. When the forces are deterministic, the first-order Langevin dynamics (FOLD) offers efficient sampling by combining a well-chosen preconditioning matrix S with a time-step-bias-mitigating propagator [G. Mazzola and S. Sorella, Phys. Rev. Lett. **118**, 015703 (2017)]. However, when forces are noisy, *S* is set equal to the force-covariance matrix, a procedure that severely limits the efficiency and the stability of the sampling. Here, we develop a new, general, optimal, and stable sampling approach for FOLD under noisy forces. We apply it for silicon nanocrystals treated with stochastic density functional theory and show efficiency improvements by an order-of-magnitude.

Prediction of the equilibrium properties of extended systems using atomistic models often requires sampling from the Boltzmann distribution of a series of configurations.[1–6] Most common sampling methods implicitly assume that either the potential energy surface[7] or the forces on the nuclei[8–12] are accessible, either through deterministic *ab initio* methods such as density functional theory (DFT) or other quantum chemistry methods (for small-medium sized systems)[13] or through empirical force-fields. For extended systems, *ab initio* methods often rely on stochastic techniques such as Quantum Monte Carlo (QMC)[14–17] or stochastic DFT (sDFT).[18–23] For example, in sDFT, the forces are calculated using a relatively small number of *stochastic orbitals* instead of using the full set of deterministic Kohn–Sham eigenstates. Therefore, the forces calculated within sDFT are noisy with fluctuating values. Such noisy forces can also occur with partially converged self-consistent field approaches to deterministic DFT.[24,25]

Langevin dynamics (LD) often serves to generate a series of thermally distributed nuclear configurations, based on the calculated forces on the nuclei. The balance between accuracy, which favors small time steps, and efficiency, which requires large time steps (to reduce the correlations between consecutive configurations in the series), determines the overall complexity and accuracy of this class of approaches. A common form of Langevin dynamics is the so-called second-order LD (SOLD),[15,17,19,26–28] in which the Newton equation of motion includes a friction term and a noisy force obeying the fluctuation–dissipation relation. An alternative is the first-order Langevin dynamics (FOLD),[15,29,30] which is conceptually simpler than SOLD because it does not have inertia, and therefore, only nuclear configurations are Boltzmann-sampled. FOLD is amenable to the introduction of a *preconditioning matrix,* which, by proper choice, dramatically increases the configurational sampling efficiency without affecting the accuracy.[29] Unfortunately, when the forces are noisy, this preconditioning matrix must be set equal to the force covariance matrix[16] and, thus, cannot be used for obtaining optimal sampling efficiency. Therefore, it seems that noisy forces, used in conjunction with FOLD, are inherently less efficient than

deterministic ones. An additional complication appears as numerical instabilities due to the singular nature of the force covariance matrix.

In this Communication, we develop an approach that enables the use of noisy forces within FOLD, lifting the constraints on the preconditioning matrix. Furthermore, we demonstrate the approach for silicon nanocrystals within sDFT and show an order of magnitude increase in sampling efficiency compared to state of the art methods for noisy forces. The solution lies in *adding* random noise that combines with the preconditioning matrix to complement the noise in the force coming from the stochastic electronic structure method.

In its simplest form, the time-discretized first-order Langevin dynamics produces a set of M configurations $\boldsymbol{R}_\tau \equiv \left(R_\tau^1, \ldots, R_\tau^{3N}\right)^\dagger$, $\tau = 1,\ldots,M$, for an N nuclei system, through a random walk described by[16]

$$\boldsymbol{R}_{\tau+1} = \boldsymbol{R}_\tau + \sqrt{2k_BT\Delta_t}\boldsymbol{\zeta}_\tau + \Delta_t S^{-1}\boldsymbol{f}(\boldsymbol{R}_\tau), \tag{1}$$

where $\boldsymbol{f}(\boldsymbol{R}) \equiv \left(f^1(\boldsymbol{R}), \ldots, f^{3N}(\boldsymbol{R})\right)^\dagger = -\nabla V(\boldsymbol{R})$ is the force acting on the nuclear degrees of freedom $\boldsymbol{R}$, Δ_t is a unit-less time step parameter, and S is an arbitrary $3N \times 3N$ symmetric positive-definite matrix. The random vector $\boldsymbol{\zeta}_\tau = \left(\zeta_\tau^1, \ldots, \zeta_\tau^{3N}\right)^\dagger$, with which thermal fluctuations are introduced, is distributed such that $\langle\boldsymbol{\zeta}^\tau\rangle = 0$ and

$$\left\langle \boldsymbol{\zeta}_\tau \boldsymbol{\zeta}_{\tau'}^\dagger \right\rangle = S^{-1}\delta_{\tau\tau'}. \tag{2}$$

For *any* choice of the preconditioning matrix S, the generated trajectory samples the Boltzmann distribution at temperature T in the $\Delta_t \to 0$ and $M \to \infty$ limits.[30] For finite values of M and Δ_t, the configurations can then be used to produce estimates of the thermal average of quantities A: $\langle A\rangle_T \approx \langle A_M\rangle \equiv \left\langle \frac{1}{M}\sum_{\tau=1}^M A(\boldsymbol{R}_\tau)\right\rangle$. One would expect that the variance of A_M is $\sigma_{A,T,M}^2 = \frac{\sigma_{A,T}^2}{M}$, where $\sigma_{A,T}^2$ is the thermal variance in A at temperature T. However, since configurations R_τ and $R_{\tau+\tau'}$ correlate, the actual variance is much larger, $\sigma_{A,T,M}^2 = \frac{\sigma_{A,T}^2}{M}\tau_c$, where τ_c is the number of correlated time steps. The smaller the τ_c, the more efficient the Langevin dynamics for sampling.

Consider now the efficiency of the method in the $T \to 0$ limit for the $3N$-dimensional harmonic oscillator $V(\boldsymbol{R}) = \frac{1}{2}\boldsymbol{R}^\dagger H\boldsymbol{R}$, where H is the Hessian matrix $[H_{ij} = \frac{\partial^2 V(\boldsymbol{R})}{\partial R^i \partial R^j}]$. In this limit, the trajectory generated by Eq. (1) with $\boldsymbol{f}(\boldsymbol{R}) = -H\boldsymbol{R}$ is given by $\boldsymbol{R}_\tau = (1-\Delta_t U)^\tau \boldsymbol{R}_0$, where $U = S^{-1}H$ and $\tau = 0, 1, 2, \ldots$ enumerate the time steps. The correlation between displacements after many time steps decays as $e^{-\Delta_t u_{\min}\tau}$, where $u_{\min} > 0$ is the smallest eigenvalue of U, so $\tau_c \approx (\Delta_t u_{\min})^{-1}$. Furthermore, the trajectory R_τ remains stable as long as $u_{\max}\Delta_t < 2$, where $u_{\max}$ is the largest eigenvalue of U. Thus, τ_c is limited from below by

$$\tau_c > \frac{1}{2}\frac{u_{\max}}{u_{\min}} \equiv \frac{1}{2}\mathrm{cond}(U). \tag{3}$$

It is now evident how preconditioning is important. Without it (say $S = I_{3N}$, where I_{3N} is the $3N \times 3N$ unit matrix), we find $\tau_c > \frac{1}{2}\mathrm{cond}(U)$, which in typical problems can easily exceed 10^3, making the random walk very inefficient. Optimal preconditioning involves taking $S = H$, enabling τ_c to be as low as 1. However, in this case, one would have $\tau_c \approx \Delta_t^{-1}$, and since Δ_t has to be kept small to avoid bias, τ_c is often quite large *even under preconditioning*. This is where a method that reduces the time step bias, thus allowing Δ_t to grow, is required. Such a random walk was proposed in Ref. 30, based on the exact solution for a harmonic potential. It involves the following process

$$\boldsymbol{R}_{\tau+1} = \boldsymbol{R}_\tau + \sqrt{2k_BT\Delta_2}\boldsymbol{\zeta}_\tau + \Delta_1 S^{-1}\boldsymbol{f}(\boldsymbol{R}_\tau), \tag{4}$$

employing two time steps

$$\Delta_n = \frac{1}{n}\left(1 - e^{-n\Delta_t}\right),\ n = 1, 2, \tag{5}$$

and it was shown to lead to significantly lower time step biases. We refer to this type of random walk as "reduced-bias FOLD" (RB-FOLD).

What happens when the forces are random? Can we still use RB-FOLD and have efficient sampling? The random forces $\boldsymbol{\phi}(\boldsymbol{R}_\tau) = \boldsymbol{f}(\boldsymbol{R}_\tau) + \boldsymbol{\eta}_\tau$ coming from sDFT or QMC will give the deterministic force $\boldsymbol{f}(\boldsymbol{R}_\tau) = \langle\boldsymbol{\phi}(\boldsymbol{R}_\tau)\rangle$ on the average but will also involve random inseparable fluctuations $\boldsymbol{\eta}_\tau = \left(\eta_\tau^1, \ldots, \eta_\tau^{3N}\right)^\dagger$. Simply plugging the random force $\boldsymbol{\phi}(\boldsymbol{R})$ into the FOLD equation will give the wrong effective dynamics $\boldsymbol{R}_{\tau+1} = \boldsymbol{R}_\tau + \sqrt{2k_BT\Delta_t}\left(\boldsymbol{\zeta}_\tau + \sqrt{\frac{\Delta_t}{2k_BT}}S^{-1}\boldsymbol{\eta}_\tau\right) + \Delta_t S^{-1}\boldsymbol{f}(\boldsymbol{R}_\tau)$ since the noise fluctuations $\boldsymbol{\eta}_\tau$ clearly cause additional heating, violating the fluctuation–dissipation relation. Hence, whenever one replaces $\boldsymbol{f}(\boldsymbol{R})$ by $\boldsymbol{\phi}(\boldsymbol{R})$ in Eq. (1), one also needs to replace $\boldsymbol{\zeta}$ of Eq. (2) by a "smaller" fluctuation $\tilde{\boldsymbol{\zeta}}$, so the FOLD is now

$$\boldsymbol{R}_{\tau+1} = \boldsymbol{R}_\tau + \sqrt{2k_BT\Delta_t}\tilde{\boldsymbol{\zeta}}_\tau + \Delta_t S^{-1}\boldsymbol{\phi}(\boldsymbol{R}_\tau), \tag{6}$$

where

$$\left\langle \tilde{\boldsymbol{\zeta}}_\tau \tilde{\boldsymbol{\zeta}}_{\tau'}^\dagger \right\rangle = \left[S^{-1} - \frac{\Delta_t}{2k_BT}S^{-1}\mathrm{cov}\boldsymbol{\phi}(\boldsymbol{R}_\tau)S^{-1}\right]\delta_{\tau,\tau'}. \tag{7}$$

The present formulation is similar to the approach previously developed in Ref. 26 for using random forces in second-order Langevin dynamics. Here, $\mathrm{cov}\boldsymbol{\phi}(\boldsymbol{R}_\tau) = \left\langle \boldsymbol{\eta}_\tau \boldsymbol{\eta}_\tau^\dagger \right\rangle$ is the force covariance matrix, and it is proportional to $\frac{1}{I}$, where I is the number of stochastic iterations in the electronic structure calculation. Note, however, that the term on the right-hand side must be positive-definite, a condition that can be achieved by a sufficient reduction in either the time step Δ_t or the random force covariance. In both cases, this requires additional computational work. In Ref. 16, the specific choice $S = \alpha \times \mathrm{cov}\boldsymbol{\phi}(\boldsymbol{R}_\tau)$ (where α is a properly chosen constant) was made, which had the appeal that $\left\langle \tilde{\boldsymbol{\zeta}}_\tau \tilde{\boldsymbol{\zeta}}_\tau^\dagger \right\rangle$, like $\left\langle \boldsymbol{\zeta}_\tau \boldsymbol{\zeta}_\tau^\dagger \right\rangle$ of Eq. (2), was proportional to S^{-1}. However, this choice has the following shortcomings: (a) S is now time-dependent and requires special treatment in the equation of motion;[16] (b) it straddles S, leaving no room for using it as a preconditioning matrix for optimizing the efficiency; and (c) it assumes implicitly that the covariance matrix is invertible, which is not always the case. In light of these limitations, we advocate leaving S in its original form as an *optimal preconditioning matrix* (e.g., $S \approx H$) and using Eq. (1) with $\boldsymbol{\phi}(\boldsymbol{R}_\tau)$ replacing $\boldsymbol{f}(\boldsymbol{R}_\tau)$ and with $\tilde{\boldsymbol{\zeta}}$ of Eq. (7) replacing $\boldsymbol{\zeta}$ of Eq. (2). We refer to this method as "noisy-FOLD" since it is an extension of the FOLD method to noisy forces. A similar treatment in the case of the random force counterpart of RB-FOLD [Eq. (4)], to which

we henceforth refer to as "noisy-RB-FOLD," leads to the following FOLD:

$$\boldsymbol{R}_{\tau+1} = \boldsymbol{R}_\tau + \sqrt{2k_BT\Delta_2}\hat{\boldsymbol{\zeta}}_\tau + \Delta_1 S^{-1}\boldsymbol{\phi}(\boldsymbol{R}_\tau), \tag{8}$$

where

$$\left\langle \hat{\boldsymbol{\zeta}}_\tau \hat{\boldsymbol{\zeta}}_{\tau'}^\dagger \right\rangle = \left[\left(1 - \frac{\Delta_1^2}{2k_BT\Delta_2} S^{-1}\mathrm{cov}\boldsymbol{\phi}(\boldsymbol{R}_\tau)\right)S^{-1}\right]\delta_{\tau\tau'}. \tag{9}$$

These two equations form the main result of this Communication since this noisy-RB-FOLD preserves much of the flexibility in choosing the matrix S as in the RB-FOLD solution while allowing for stochastic forces. As noted above for noisy FOLD, here too, the right-hand side of Eq. (9) must be positive-definite. To enforce this condition, additional numerical work is required, either by decreasing the time step or the force covariance. The first measure, decreasing the time step, increases the sample correlations, so additional time steps are needed as a compensation. The second measure, reducing cov$\boldsymbol{\phi}$, calls for a step-up in the number of stochastic electronic-structure iterations.

We use the Harmonic potential discussed above to demonstrate the theory in Fig. 1. We plot the fluctuation σ_V and the bias ΔV for various sampling procedures within FOLD, comparing the non-optimal preconditioning choice, $S = \alpha\mathrm{cov}\boldsymbol{\phi}$ (with $\alpha = 1$ in the units of the Harmonic oscillator, triangles), discussed in Ref. 16 and the optimal preconditioning $S = H$ (squares) advocated here. It is evident from Fig. 1 that whether one uses noisy-FOLD [blue symbols, Eqs. (6) and (7)] or noisy-RB-FOLD [red symbols, Eqs. (8) and (9)], the bias ΔV can be reduced only by decreasing the time steps Δ_t. However, the abovementioned

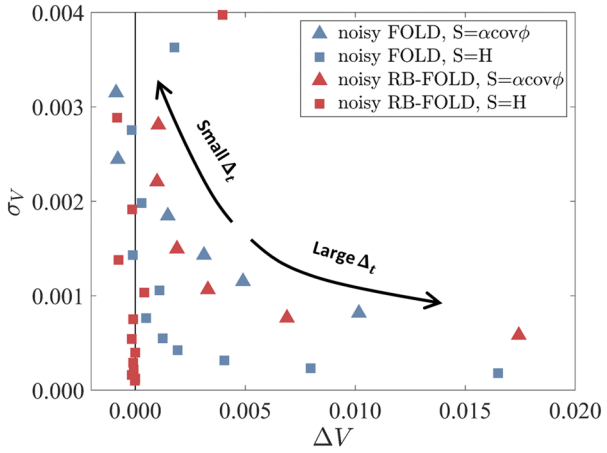

FIG. 1. The bias ($\Delta V = \langle V\rangle - \frac{3}{2}k_BT$, where $k_BT = 0.1$) and the fluctuation σ_V in the average potential energy estimate $\langle V\rangle$ (determined using binning analysis[30]) for noisy-RB-FOLD and RB-FOLD calculations on a 3D harmonic oscillator with a random force $\boldsymbol{\phi}$, $\langle\boldsymbol{\phi}\rangle = H\boldsymbol{R}$, and cov$\boldsymbol{\phi} = 0.02I_{3N}$, where the Hessian H is diagonal with values of 0.1, 1, and 10. We show results for $S = \alpha\mathrm{cov}\boldsymbol{\phi}$ (with $\alpha = 1$, triangles) and $S = H$ (squares). The blue symbols correspond to noisy-FOLD [Eqs. (6) and (7)], while the red symbols correspond to noisy-RB-FOLD [Eqs. (8) and (9)]. The points are differentiated by a time step parameter Δ_t (not specified). The results are calculated using trajectories of 5×10^7 steps.

analysis of τ_c showed that as Δ_t decreases, the fluctuation σ_V grows. Under optimal preconditioning $S = H$ (squares), we see that noisy-FOLD [blue symbols, Eqs. (6) and (7)] biases are reduced, yet the error control is still unsatisfactory since any attempt to reduce the bias further (by decreasing Δ_t) increases once again the fluctuation σ_V. This problem does not arise for noisy-RB-FOLD results [red squares, Eqs. (8) and (9)], where Δ_t can grow to lower σ_V without a bias penalty. Note that to within small fluctuations, the same results shown here for noisy forces also appear for deterministic ones (obtained by taking $\boldsymbol{\phi} = f$ and cov$\boldsymbol{\phi} = 0$), not shown here.

We expect the noisy-RB-FOLD calculations to be highly efficient not only for the Harmonic model but also for more realistic systems. To demonstrate this, we apply the method to the problem of determining the structural properties of a realistic atomistic system such as the $Si_{35}H_{36}$ nanocrystal at $T = 300$ K, described with DFT at the local density approximation level.[31] Our purpose is to validate the noisy-RB-FOLD sampling approaches based on sDFT forces using calculations based on sampling methods that employ dDFT forces (RB-FOLD, FOLD, and SOLD[19]) and to compare the efficiencies of these methods. Note that all the FOLD methods in Fig. 2 are based on optimal sampling, with $S = H$, where a finite-difference approximation for estimating the Hessian was used within deterministic DFT (see the supplementary material for additional details). We could not show results for the choice $S = \alpha\mathrm{cov}\boldsymbol{\phi}$ of Ref. 16 because of numerical problems stemming from the fact that the sDFT forces have a force-covariance matrix that is nearly singular (see the supplementary material). In Fig. 2, (left panel) we show that, indeed, our new noisy-RB-FOLD method as well as the other methods predicts the same first peak of the pair distribution function $g(r)$ (to within statistical fluctuations).[32] In order to study the efficiency, we plot in the right panel the pair-distance correlation function in terms of the distance r_{ij} between a pair of silicon atoms, numbered i and j,

$$C_\tau = \frac{\left\langle \sum_{t=1}^{N_\tau-\tau} r_{ij}^t r_{ij}^{t+\tau}\right\rangle_{\{i,j\}}}{\left\langle \sum_{t=1}^{N_\tau-\tau} r_{ij}^t r_{ij}^t\right\rangle_{\{i,j\}}}, \tag{10}$$

where $\langle\rangle_{\{i,j\}}$ represents an average over these pairs and N_τ is the total number of steps in the Langevin trajectory. C_τ has the initial value of one at $\tau = 0$, and then, it decays non-monotonically until it settles upon a steady fluctuation around zero. We define the time scale τ_c for this decay as the earliest time for which $C_{\tau_c} = 0.1$. Consider first the correlation functions for the FOLD and the SOLD trajectories; both are seen to have a concave structure at small values of τ, which delays decay and turns convex only at much longer times, and both trajectories exhibit a slow decay with $\tau_c \approx 100$. Next, consider the correlation functions for RB-FOLD: the deterministic RB-FOLD with $\Delta_1 = 1$ ($\Delta_t = 10$, $\Delta_2 = 0.5$) and the noisy-RB-FOLD with $\Delta_1 = 0.5$ ($\Delta_2 = 0.375$) with ($\Delta_t = 0.7$, $\Delta_2 = 0.375$). In Fig. 2, τ_c is twice as large when $\Delta_1 = 0.5$ than when $\Delta_1 = 1$, in agreement with our analysis above, and both functions have a similar convex form. We have verified that the correlations of the noisy- and deterministic RB-FOLD trajectories for $\Delta_1 = 0.5$ are identical (not shown here), and we see that they represent an order of magnitude improvement on the previously used SOLD approach for sDFT.

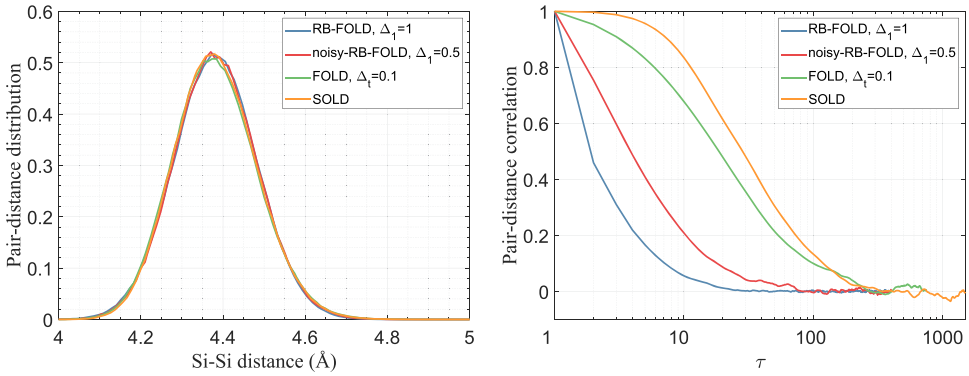

FIG. 2. Comparison of four sampling methods applied to $Si_{35}H_{36}$ at T = 300 K. The left panel shows the average of the Si–Si pair-distance distribution function $g(r)$, while the right panel displays the pair-distance correlation functions C_τ. The methods are noisy-RB-FOLD, based on sDFT, and FOLD, RB-FOLD, and SOLD, based on dDFT. All FOLD-based methods use optimal preconditioning $S = H$ (see the supplementary material concerning calculation of the Hessian H). For each of the methods, we produced a 3000-step trajectory starting from the same configuration, and the shown results are based on them.

Summarizing, in previous work,[19] we used SOLD to address the problem of noisy forces in sDFT calculations but found that thousands of time steps were required to shake off the correlations. Here, we developed a radically more efficient method for sampling system configurations under stochastic forces. It capitalizes on a recently proposed method[16] but makes critical changes in the Langevin force sampling, which restore optimal preconditioning. The final procedure is to perform a random walk following Eq. (8) while sampling the Langevin forces from Eq. (9).

Using a purely Harmonic model system, we compared noisy-FOLD and noisy-RB-FOLD and showed that the latter is much more efficient and insensitive to the time step. We further showed that the noisy-RB-FOLD has similar characteristics also when applied to the real atomistic system using sDFT forces. One notable difference between RB-DFT and noisy-RB-DFT concerns with the increase in the time step. One must assure that the left-hand side of Eq. (9) is positive-definite; hence, at some point, any increase in Δ_1^2/Δ_2 will necessitate a reduction in cov$\boldsymbol{\phi}$. This is especially important at low temperatures. The results of this work provide a general recipe for efficient and stable Boltzmann sampling under the presence of stochastic forces. Our approach is efficient for systems in which the Hessian and cov$\boldsymbol{\phi}$ do not change much over time, as is typical of calculations in solids and nanocrystals. In high temperatures, when treating gases and liquids or when bond breaking situations occur, these assumptions do not hold and an occasional update of the matrices is required. In this case, further considerations along the lines depicted in Ref. 16 may be necessary.

As explained, any positive-definite matrix S can be used without bias (in the vanishing time step limit); however, there is a great advantage in choosing $S = H$, where H is the Hessian, because this allows for large time steps and small correlation times. To demonstrate this principle, we used a finite-difference approximation for estimating the Hessian within deterministic DFT. Clearly, such an approach is not scalable for large systems. In Refs. 15 and 16, it was found that the QMC force-covariance, covϕ, is approximately proportional to the Hessian. Unfortunately, this is not the case for sDFT. Still, sufficiently good Hessian approximations can probably be obtained from empirical force-fields or the embedded-fragment sDFT procedure.[19,22,23,33] Further work researching this topic is required.

The supplementary material is given concerning the calculation of the Hessian matrix and the properties of the covariance matrix of the $Si_{35}H_{36}$ system.

RB gratefully acknowledges the support from the US-Israel Binational Science Foundation (BSF) under Grant No. 2018368. D.N. acknowledges support from the National Science Foundation, Grant No. CHE-1763176. E.R. acknowledges support from the Department of Energy, *Photonics at Thermodynamic Limits Energy Frontier Research Center*, under Grant No. DE-SC0019140.

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

REFERENCES

[1] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed. (Academic Press, San Diego, 2002).
[2] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, Oxford, 1987).
[3] D. C. Rapaport, *The Art of Molecular Dynamics Simulation* (Cambridge University Press, 2004).
[4] R. D. Skeel and Y. Fang, "Comparing Markov chain samplers for molecular simulation," Entropy **19**(10), 561 (2017).
[5] C. Y. Gao and D. T. Limmer, "Transport coefficients from large deviation functions," Entropy **19**(11), 571 (2017).
[6] T. Schlick, *Molecular Modeling and Simulation: An Interdisciplinary Guide* (Springer, 2010), Vol. 21.
[7] D. M. Ceperley, "Metropolis methods for quantum Monte Carlo simulations," AIP Conf. Proc. **690**, 85–98 (2003).
[8] D. Marx and J. Hutter, "*Ab initio* molecular dynamics: Theory and implementation," in *Modern Methods and Algorithms of Quantum Chemistry, Proceedings*, Volume 3 of NIC Series, edited by J. Grotendorst (John von Neumann Institute for Computing, Julich, 2000), p. 329.

[9]M. E. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulation* (Oxford University Press, Oxford, New York, 2010).
[10]M. Ceriotti, G. Bussi, and M. Parrinello, "Langevin equation with colored noise for constant-temperature molecular dynamics simulations," Phys. Rev. Lett. **102**(2), 020601 (2009).
[11]J. Dai and J. Yuan, "Large-scale efficient Langevin dynamics, and why it works," EPL (Europhys. Lett.) **88**(2), 20001 (2009).
[12]G. Bussi and M. Parrinello, "Accurate sampling using Langevin dynamics," Phys. Rev. E **75**(5), 056707 (2007).
[13]D. Marx and J. Hutter, *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods* (Cambridge University Press, 2009).
[14]D. M. Ceperley and M. Dewing, "The penalty method for random walks with uncertain energies," J. Chem. Phys. **110**(20), 9812–9820 (1999).
[15]Y. Luo, A. Zen, and S. Sorella, "*Ab initio* molecular dynamics with noisy forces: Validating the quantum Monte Carlo approach with benchmark calculations of molecular vibrational properties," J. Chem. Phys. **141**(19), 194112 (2014).
[16]G. Mazzola and S. Sorella, "Accelerating *ab initio* molecular dynamics and probing the weak dispersive forces in dense liquid hydrogen," Phys. Rev. Lett. **118**(1), 015703 (2017).
[17]F. R. Krajewski and M. Parrinello, "Linear scaling electronic structure calculations and accurate statistical mechanics sampling with noisy forces," Phys. Rev. B **73**(4), 041105 (2006).
[18]R. Baer, D. Neuhauser, and E. Rabani, "Self-averaging stochastic Kohn-Sham density-functional theory," Phys. Rev. Lett. **111**(10), 106402 (2013).
[19]E. Arnon, E. Rabani, D. Neuhauser, and R. Baer, "Equilibrium configurations of large nanostructures using the embedded saturated-fragments stochastic density functional theory," J. Chem. Phys. **146**(22), 224111 (2017).
[20]D. Neuhauser, E. Rabani, Y. Cytter, and R. Baer, "Stochastic optimally tuned range-separated hybrid density functional theory," J. Phys. Chem. A **120**(19), 3071–3078 (2016).
[21]Y. Cytter, E. Rabani, D. Neuhauser, and R. Baer, "Stochastic density functional theory at finite temperatures," Phys. Rev. B **97**, 115207 (2018).
[22]M. Chen, R. Baer, D. Neuhauser, and E. Rabani, "Overlapped embedded fragment stochastic density functional theory for covalently-bonded materials," J. Chem. Phys. **150**(3), 034106 (2019).
[23]M. D. Fabian, B. Shpiro, E. Rabani, D. Neuhauser, and R. Baer, "Stochastic density functional theory," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **9**(6), e1412 (2019).
[24]T. D. Kühne, M. Krack, F. R. Mohamed, and M. Parrinello, "Efficient and accurate car-parrinello-like approach to Born-Oppenheimer molecular dynamics," Phys. Rev. Lett. **98**(6), 066401 (2007).
[25]E. Martínez, M. J. Cawkwell, A. F. Voter, and A. M. N. Niklasson, "Niklasson. Thermostating extended Lagrangian Born-Oppenheimer molecular dynamics," J. Chem. Phys. **142**(15), 154120 (2015).
[26]C. Attaccalite and S. Sorella, "Stable liquid hydrogen at high pressure by a novel *ab initio* molecular-dynamics calculation," Phys. Rev. Lett. **100**(11), 114501 (2008).
[27]F. Tassone, F. Mauri, and R. Car, "Acceleration schemes for *ab initio* molecular-dynamics simulations and electronic-structure calculations," Phys. Rev. B **50**(15), 10561–10573 (1994).
[28]C. H. Bennett, "Mass tensor molecular dynamics," J. Comput. Phys. **19**(3), 267–279 (1975).
[29]H. Risken. *The Fokker-Planck Equation: Methods of Solution and Applications*. Springer series in synergetics, 2nd ed. (Springer-Verlag, Berlin; New York, 1989).
[30]F. Becca and S. Sorella, *Quantum Monte Carlo Approaches for Correlated Systems*, 1st ed. (Cambridge University Press, 2017).
[31]All calculations in this work use real-space grids of spacing $\Delta x = 0.5_{a0}$ and Troullier–Martins norm-conserving pseudopotentials [N. Troullier and J. L. Martins, "Efficient pseudopotentials for plane-wave calculations," Phys. Rev. B **43**(3), 1993–2006 (1991)] within the Kleinman–Bylander approximation [L. Kleinman and D. M. Bylander, "Efficacious form for model pseudopotentials," Phys. Rev. Lett. **48**, 1425 (1982)]. Fast Fourier transforms were used for applying the kinetic energy operator and determining the Hartree potentials, and the method of G. J. Martyna and M. E. Tuckerman ["A reciprocal space based method for treating long range interactions in ab initio and force field-based calculations in clusters," J. Chem. Phys. **110**(6), 2810–2821 (1999)] was used for treating the long range Coulomb interactions in a finite simulation cell with periodic boundary conditions. DFT calculations were preformed under the local density approximation (LDA) using the PW92 functional [J.P. Perdew and Y. Wang, "Accurate and Simple Analytic Representation of the Electron-Gas Correlation-Energy," Phys. Rev. B **45**(23), 13244–13249 (1992)].
[32]$g(r) = \frac{\Delta_n(r)}{4\pi r^2 \Delta_r \rho_0}$, where $\Delta_n(r)$ is the number of silicon pairs at distance $[r, r + \Delta_r]$ and ρ_0 is the average silicon atom density.
[33]D. Neuhauser, R. Baer, and E. Rabani, "Communication: Embedded fragment stochastic density functional theory," J. Chem. Phys. **141**(4), 041102 (2014).