

# Energy window stochastic density functional theory

Cite as: J. Chem. Phys. 151, 114116 (2019); doi: 10.1063/1.5114984

Submitted: 13 June 2019 • Accepted: 27 August 2019 •

Published Online: 20 September 2019



View Online



Export Citation



CrossMark

Ming Chen,<sup>1,a)</sup> Roi Baer,<sup>2,b)</sup>  Daniel Neuhauser,<sup>3,c)</sup> and Eran Rabani<sup>1,4,d)</sup> 

## AFFILIATIONS

<sup>1</sup>Department of Chemistry, University of California and Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

<sup>2</sup>Fritz Haber Center of Molecular Dynamics and Institute of Chemistry, The Hebrew University of Jerusalem, Jerusalem 91904, Israel

<sup>3</sup>Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095, USA

<sup>4</sup>The Raymond and Beverly Sackler Center of Computational Molecular and Materials Science, Tel Aviv University, Tel Aviv 69978, Israel

<sup>a)</sup>Electronic mail: [mingchen.chem@berkeley.edu](mailto:mingchen.chem@berkeley.edu)

<sup>b)</sup>Electronic mail: [roi.baer@huji.ac.il](mailto:roi.baer@huji.ac.il)

<sup>c)</sup>Electronic mail: [dxn@chem.ucla.edu](mailto:dxn@chem.ucla.edu)

<sup>d)</sup>Electronic mail: [eran.rabani@berkeley.edu](mailto:eran.rabani@berkeley.edu)

## ABSTRACT

Linear scaling density functional theory is important for understanding electronic structure properties of nanometer scale systems. Recently developed stochastic density functional theory can achieve linear or even sublinear scaling for various electronic properties without relying on the sparsity of the density matrix. The basic idea relies on projecting stochastic orbitals onto the occupied space by expanding the Fermi-Dirac operator and repeating this for  $N_\chi$  stochastic orbitals. Often, a large number of stochastic orbitals are required to reduce the statistical fluctuations (which scale as  $N_\chi^{-1/2}$ ) below a tolerable threshold. In this work, we introduce a new stochastic density functional theory that can efficiently reduce the statistical fluctuations for certain observable and can also be integrated with an embedded fragmentation scheme. The approach is based on dividing the occupied space into energy windows and projecting the stochastic orbitals with a single expansion onto all windows simultaneously. This allows for a significant reduction of the noise as illustrated for bulk silicon with a large supercell. We also provide theoretical analysis to rationalize why the noise can be reduced only for a certain class of ground state properties, such as the forces and electron density.

Published under license by AIP Publishing. <https://doi.org/10.1063/1.5114984>

## I. INTRODUCTION

Accurate modeling of the electronic structure is important for understanding the thermal, optical, and magnetic properties of molecules and condensed matters. Density functional theory (DFT)<sup>1</sup> within the Kohn-Sham (KS) scheme<sup>2</sup> is often the method of choice due to its relatively low computational cost, which formally scales as  $N_e^3$ , where  $N_e$  is the number of electrons in the system. However, even this mild power-law scaling limits the applicability of DFT to study complex systems that require modeling of more than  $10^4$  electrons.<sup>3–5</sup> In order to reduce the scaling, much effort has been devoted to the development of linear scaling DFT algorithms.<sup>6–17</sup> The

majority of these algorithms either divide a system into subsystems<sup>14–17</sup> and attempt to solve the electronic structure problem in each subsystem or rely on the sparsity of the density matrix.<sup>6,10,12</sup> Although linear scaling DFT has achieved many successes, difficulties exist in both approaches. Algorithms based on solving subsystem such as like embedding methods require a proper treatment of boundaries between subsystems<sup>16,18–22</sup> while methods based on the locality of the density matrix often suffer from slow convergence when the fundamental band gap is small.

Recently, we have introduced an alternative scheme for linear-scaling DFT based on the notion of stochastic orbitals, termed stochastic DFT (sDFT).<sup>23</sup> The idea is to compute the electron

density and ground state electronic properties with stochastic orbitals that are random linear combinations of the deterministic occupied orbitals. Importantly, the sDFT approach depends on neither the sparsity of density matrix nor how the boundaries between fragments in an embedding scheme are treated. This has been shown for a variety of systems, including insulators and semiconductors.<sup>23–28</sup> Although sDFT is very efficient for large systems, lowering the statistical noise of ground state electronic properties to below tolerable thresholds often requires hundreds to thousands of stochastic orbitals, leading to significant computational costs.<sup>23</sup> Various techniques based on embedded fragmentation algorithms have been successfully introduced to reduce the noise in the electron density for specific systems, such as molecular systems<sup>24</sup> and covalently bonded materials,<sup>25,26</sup> but more robust noise reduction schemes are needed, particularly for systems for which it is nontrivial to construct embedded fragments.

Here, we propose a general improvement to sDFT that reduces the noise in the electron density and nuclear forces by breaking the electron density into components corresponding to orbital energy windows and using sDFT to sample each window. We refer to this new method as “energy window stochastic density functional theory” (ew-sDFT). We find that ew-sDFT successfully reduces the statistical noise in the electron density and in certain other ground state electronic structure properties (e.g., forces).

This paper is organized as follows: We briefly review the sDFT approach and then present the energy-window modification followed by a numerical illustration for bulk silicon with a large supercell. Finally, we provide a discussion of the noise reduction mechanism in ew-sDFT.

## II. STOCHASTIC DENSITY FUNCTIONAL THEORY

We begin by briefly reviewing the stochastic density functional theory (for more details, consult Refs. 23–26 and 28). Without loss of generality, we limit the discussion to a system in a box of volume  $V$  with periodic boundary condition. Electronic properties are calculated by KS-DFT with Hamiltonian  $\hat{h}_{\text{KS}}$  given by

$$\hat{h}_{\text{KS}} = \hat{t} + \hat{v}_{\text{nl}} + \hat{v}_{\text{loc}} + \hat{v}_{\text{h}}[\rho] + \hat{v}_{\text{xc}}[\rho], \quad (1)$$

where  $\hat{t}$ ,  $\hat{v}_{\text{nl}}$ ,  $\hat{v}_{\text{loc}}$ ,  $\hat{v}_{\text{h}}[\rho]$ , and  $\hat{v}_{\text{xc}}[\rho]$  are kinetic, nonlocal pseudopotential, local pseudopotential, Hartree, and exchange correlation terms. In KS-DFT, the Hartree and exchange correlation potentials explicitly depend on the electron density  $\rho(\mathbf{r})$ , which is given by  $\rho(\mathbf{r}) = 2 \sum_{i=1}^{N_{\text{occ}}} |\psi_i(\mathbf{r})|^2$  on  $N_{\text{g}}$  real-space grid points, where  $\psi_i(\mathbf{r})$  are the KS orbitals of  $\hat{h}_{\text{KS}}$  and  $N_{\text{occ}}$  is the number of occupied orbitals.  $\rho(\mathbf{r})$  can also be expressed as the trace of an operator, i.e.,

$$\rho(\mathbf{r}) = 2\text{Tr} \hat{\rho} \delta(\mathbf{r} - \hat{\mathbf{r}}), \quad (2)$$

where  $\hat{\rho} = \lim_{\beta \rightarrow \infty} \theta_{\beta}(\hat{h}_{\text{KS}}; \mu)$  is the one-body density matrix,  $\theta_{\beta}(x; \mu) = 1/(1 + e^{\beta(x-\mu)})$  is the Fermi-Dirac distribution function with inverse temperature  $\beta$  and chemical potential  $\mu$ , and  $\delta(\mathbf{r} - \hat{\mathbf{r}})$  is the Dirac-delta function. In practice,  $\beta$  is chosen to be large enough to converge the ground state properties and  $\mu$  is tuned such that  $\text{Tr} \hat{\rho} = N_e/2$  is equal to the number of electron pairs.

In sDFT,  $\rho(\mathbf{r})$  in Eq. (2) is determined by using a set of  $N_{\chi}$  stochastic orbitals  $|\chi\rangle$  (rather than finding all KS orbitals),

$$\rho(\mathbf{r}) = 2\langle\langle \chi | \hat{\rho} \delta(\mathbf{r} - \hat{\mathbf{r}}) | \chi \rangle\rangle_{\chi} = 2\langle\langle |\xi(\mathbf{r})|^2 \rangle\rangle_{\chi}, \quad (3)$$

where  $|\xi\rangle = \sqrt{\hat{\rho}}|\chi\rangle$  and  $\langle\langle \dots \rangle\rangle_{\chi}$  imply averaging over an ensemble of stochastic orbitals. One possible choice of  $|\chi\rangle$  in real space is  $\chi(\mathbf{r}) = \pm 1/\sqrt{\Delta V}$  with equal random probabilities of positive or negative values for each grid point, and  $\Delta V = V/N_{\text{g}}$  represents the volume element. In practice, a set of  $N_{\chi}$  stochastic orbitals  $\{|\chi_i\rangle\}$  is used in sDFT and a sample average leads to  $\bar{\rho}_{\chi}(\mathbf{r}) = \frac{2}{N_{\chi}} \sum_{i=1}^{N_{\chi}} |\xi_i(\mathbf{r})|^2 \cdot \bar{\rho}_{\chi}(\mathbf{r})$  is then used to evaluate quantities that are functionals of the density, such as the local pseudopotential energy, Hartree energy, exchange-correlation energy, and the local pseudopotential nuclei force. A ground state quantity corresponding to a one-body operator,  $\hat{O}$ , can also be obtained via a trace formula,

$$O = 2\text{Tr}(\hat{\rho}\hat{O}) = 2\langle\langle \xi | \hat{O} | \xi \rangle\rangle_{\chi}. \quad (4)$$

Equation (4) is used to evaluate kinetic energy ( $E_{\text{K}}$ ) and nonlocal pseudopotential energy ( $E_{\text{nl}}$ ) and nonlocal pseudopotential nuclei force. Applying Eq. (4) with a finite number of samples results in a statistical approximation of the expectation value of  $\hat{O}$ .

Achieving linear (or sublinear) scaling in sDFT relies on two requirements. First, applying  $\sqrt{\hat{\rho}} = \sqrt{\theta_{\beta}(\hat{h}_{\text{KS}}; \mu)}$  to a stochastic orbital  $|\chi\rangle$  should scale linearly with respect to system size. This is achieved by expanding the Fermi-Dirac operator in a Chebyshev polynomial,<sup>29,30</sup>

$$\sqrt{\hat{\rho}}|\chi\rangle = \sqrt{\theta_{\beta}(\hat{h}_{\text{KS}}; \mu)}|\chi\rangle = \sum_{n=0}^{N_c} a_n(\mu, \beta) T_n(\hat{h}_{\text{KS}})|\chi\rangle, \quad (5)$$

where  $a_n(\mu, \beta)$  are the expansion coefficients of the Fermi-Dirac operator<sup>31</sup> and  $N_c$  is the highest degree of polynomials. Linear scaling then relies on the sparsity of the KS Hamiltonian and not on the density matrix. Second, for many local observables like the energy per particle and the forces on the atoms, the statistical noise does not increase with increasing system size and often decreases with the system size, leading to sublinear scaling.<sup>23–26,28</sup>

Within the sDFT formulation, the exact electron density  $\rho(\mathbf{r})$  can only be recovered by averaging infinitely many stochastic orbitals. Thus, for a finite set of stochastic orbitals,  $\bar{\rho}_{\chi}(\mathbf{r})$  contains a statistical noise. With the choice of random state  $|\chi\rangle$  described above, the variance of density calculated from one stochastic orbital,  $\rho_{\chi}(\mathbf{r}) = 2\langle\chi | \theta_{\beta}(\hat{h}_{\text{KS}}; \mu) \delta(\mathbf{r} - \hat{\mathbf{r}}) | \chi\rangle$ , is approximately given by<sup>25</sup>

$$\text{Var}_{\chi}(\rho_{\chi}(\mathbf{r})) = 2\rho^2(\mathbf{r}), \quad (6)$$

where  $\text{Var}_{\chi}(\dots)$  represents variance in  $\rho(\mathbf{r})$ . From the central limit theorem, the variance of the electron density  $\bar{\rho}_{\chi}(\mathbf{r})$  evaluated with  $N_{\chi}$  stochastic orbitals is  $2\rho^2(\mathbf{r})/N_{\chi}$ . Equation (6) indicates that the standard deviation of the density is proportional to the density itself. Thus, by dividing  $\rho(\mathbf{r})$  into multiple components and using sDFT to sample each of them, it is possible to reduce the statistical noise, as discussed in Sec. III.

## III. ENERGY WINDOW STOCHASTIC DENSITY FUNCTIONAL THEORY

The central idea behind the energy window stochastic density functional theory (ew-sDFT) is to divide the space spanned by the

occupied orbitals into subspaces and perform sDFT calculation in each subspace before combining contributions from all subspaces. We will illustrate ew-sDFT with two energy windows; however, generalizing the results to multiple windows is straightforward. First, we divide the occupied orbitals,  $\{\psi_i\}$ , into two subsets: one with orbital energy ( $\epsilon_i$ ) smaller than a given energy  $\epsilon^*$  while orbitals in the other subset have orbital energies higher than  $\epsilon^*$ , where  $\epsilon^* < \mu$ . The orbital indexes in the two subsets are  $w_1 = \{i|\epsilon_i \leq \epsilon^*\}$  and  $w_2 = \{i|\epsilon^* < \epsilon_i < \mu\}$ . The corresponding projection operators onto each energy window are  $\hat{\rho}_1 = \theta_\beta(\hat{h}_{\text{KS}}; \epsilon^*)$  and  $\hat{\rho}_2 = \theta_\beta(\hat{h}_{\text{KS}}; \mu) - \theta_\beta(\hat{h}_{\text{KS}}; \epsilon^*)$ . Similar to sDFT, applying  $\sqrt{\hat{\rho}_1}$  and  $\sqrt{\hat{\rho}_2}$  to a stochastic orbital  $|\chi\rangle$  generates filtered stochastic orbitals  $|\xi^{(1)}\rangle = \sqrt{\hat{\rho}_1}|\chi\rangle$  and  $|\xi^{(2)}\rangle = \sqrt{\hat{\rho}_2}|\chi\rangle$ . The deterministic density  $\rho(\mathbf{r})$  can be recovered by averaging  $2|\xi^{(1)}(\mathbf{r})|^2 + 2|\xi^{(2)}(\mathbf{r})|^2$ , i.e.,

$$\rho(\mathbf{r}) = 2\left(|\xi^{(1)}(\mathbf{r})|^2 + |\xi^{(2)}(\mathbf{r})|^2\right). \quad (7)$$

We evaluate both  $\sqrt{\hat{\rho}_1}$  and  $\sqrt{\hat{\rho}_2}$  using a Chebyshev polynomial expansion given by Eq. (5). We emphasize that ew-sDFT barely increases the computational wall time compared to sDFT because the same Chebyshev polynomials of  $\hat{h}_{\text{KS}}$  can be used to simultaneously project a random stochastic orbital on both energy windows, generating  $|\xi^{(1)}\rangle$  and  $|\xi^{(2)}\rangle$  with a similar effort to one window sDFT.

Within ew-sDFT, a general ground state observable can be evaluated according to

$$O = 2\text{Tr}(\hat{\rho}\hat{O}) = 2\left(\langle\xi^{(1)}|\hat{O}|\xi^{(1)}\rangle + \langle\xi^{(2)}|\hat{O}|\xi^{(2)}\rangle\right) \quad (8)$$

and by  $\rho_\chi(\mathbf{r}) = 2|\xi^{(1)}(\mathbf{r})|^2 + 2|\xi^{(2)}(\mathbf{r})|^2$ , the variance of  $\rho_\chi(\mathbf{r})$  in ew-sDFT is given by

$$\text{Var}_\chi(\rho_\chi(\mathbf{r})) = 2(\rho_1^2(\mathbf{r}) + \rho_2^2(\mathbf{r})), \quad (9)$$

where  $\rho_1(\mathbf{r}) = 2\sum_{i\in w_1}|\psi_i(\mathbf{r})|^2$  and  $\rho_2(\mathbf{r}) = 2\sum_{i\in w_2}|\psi_i(\mathbf{r})|^2$  are densities corresponding to the two energy windows. Recalling that the variance of  $\tilde{\rho}_\chi(\mathbf{r})$  from sDFT is  $2\rho^2(\mathbf{r}) = 2(\rho_1(\mathbf{r}) + \rho_2(\mathbf{r}))^2$ , the variance is thus reduced by  $4\rho_1(\mathbf{r})\rho_2(\mathbf{r})$ . Therefore, with similar computational wall time, ew-sDFT achieves smaller fluctuations in the electron density compared to sDFT.

The implementation of ew-sDFT is similar to sDFT and can be implemented in plane waves, real space grids, or Gaussian basis sets. Starting from the initial guess of the electron density, ew-sDFT first generates  $N_\chi$  random vectors in real space as described in Sec. II. Next, the chemical potential  $\mu$  is determined by solving the following equation:

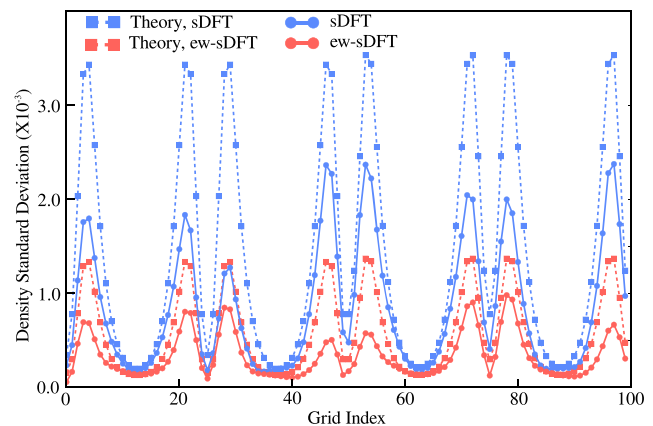
$$\frac{2}{N_\chi} \sum_{i=1}^{N_\chi} \langle\chi_i|f(\hat{h}_{\text{KS}}; \beta, \mu)|\chi_i\rangle = N_e. \quad (10)$$

Note that Eq. (10) is the same as the one used in sDFT since  $\langle\chi_i|\hat{\rho}_1|\chi_i\rangle + \langle\chi_i|\hat{\rho}_2|\chi_i\rangle = \langle\chi_i|\hat{\rho}|\chi_i\rangle$  for each  $|\chi_i\rangle$ . After determining  $\mu$ , each  $|\chi_i\rangle$  is filtered by  $\sqrt{\hat{\rho}_1}$  and  $\sqrt{\hat{\rho}_2}$  using a Chebyshev polynomial expansion as described by Eq. (5) to generate  $|\xi_i^{(1)}\rangle$  and  $|\xi_i^{(2)}\rangle$ . The filtered stochastic orbitals are then used to calculate the density,  $\tilde{\rho}_\chi(\mathbf{r})$ , and other electronic properties via Eqs. (7) and (8). The KS Hamiltonian is then updated with  $\tilde{\rho}_\chi(\mathbf{r})$ , and the above calculations

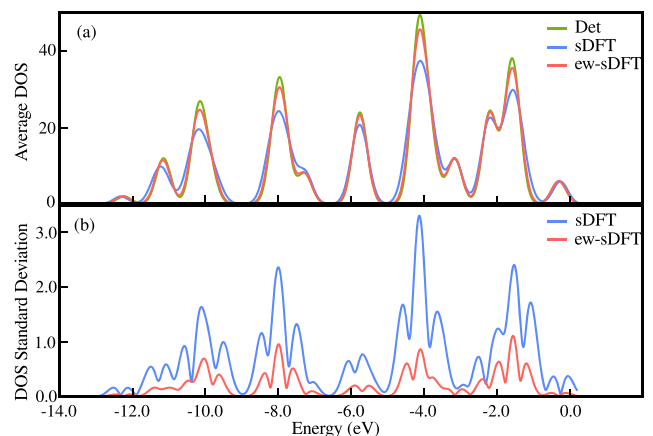
repeat until  $\tilde{\rho}_\chi(\mathbf{r})$  converges. The same random number seed is used throughout the self-consistency iterations to ensure convergence.

## IV. RESULTS AND DISCUSSION

We assessed the performance of ew-sDFT on silicon crystal with various supercell size. Silicon crystal is a challenging system for linear scaling DFT algorithms due to its small band gap within the local density approximation (LDA) to DFT (fundamental gap of  $\sim 0.6$  eV). We used  $\beta \approx 600$  in inverse Hartree, sufficient to converge the ground state properties with respect to the electron temperature. The same value of  $\beta \approx 600$  was also sufficient to converge the ground state properties within ew-sDFT, regardless of the number of



**FIG. 1.** Standard deviation of densities at selected grid points are evaluated from sDFT (solid blue) and ew-sDFT with 10 windows (solid red) calculations for  $\text{Si}_{64}$  system. Theoretical estimated density standard deviation from Eqs. (6) and (9) are plotted as dashed blue line for sDFT and dashed red line for ew-sDFT.



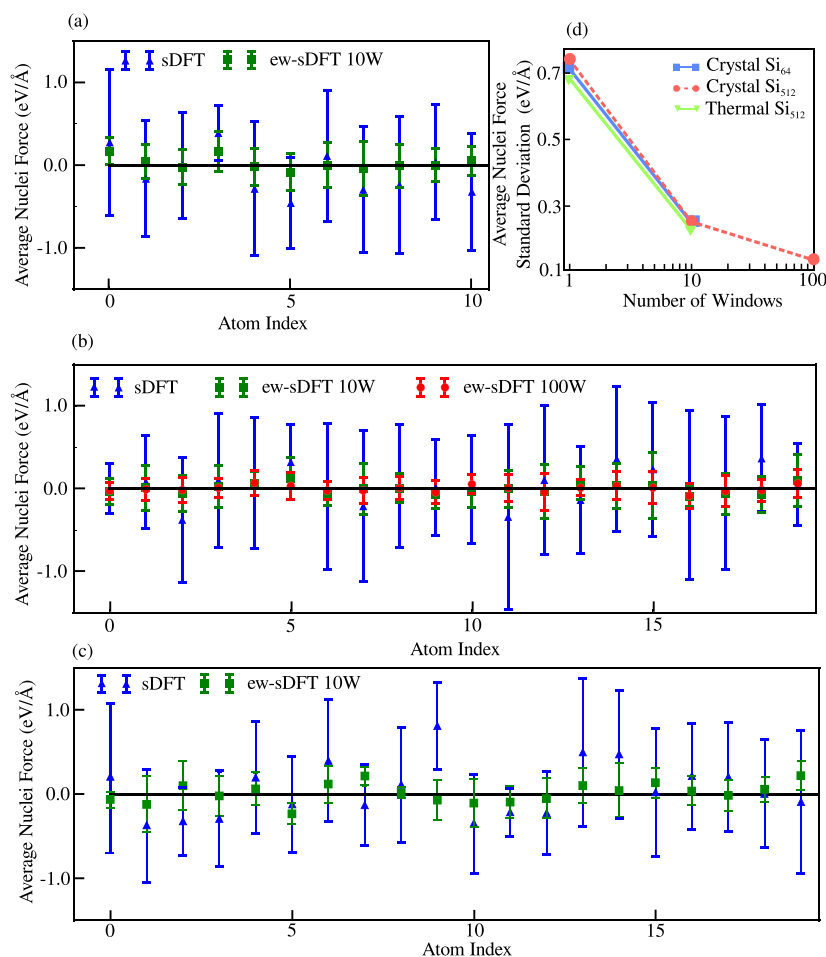
**FIG. 2.** (a) Density of states (DOS) of  $\text{Si}_{64}$  by using densities calculated from sDFT (blue line) and ew-sDFT with 10 windows (red line) are compared to deterministic DOS (green line). DOS from sDFT and ew-sDFT are averaged from 10 independent runs. (b) Standard deviations of DOS are shown as blue line for sDFT and red line for ew-sDFT.

**TABLE I.** The average values of energies per electron, including kinetic energy ( $E_k$ ), nonlocal pseudopotential energy ( $E_{nl}$ ), external energy/local pseudopotential energy ( $E_{loc}$ ), the summation of Hartree ( $E_h$ ) and exchange-correlation ( $E_{xc}$ ) energy and total energy ( $E_{tot}$ ), are presented in the table. The standard deviations in the last digits of energies are listed in parentheses.  $N_w$  is the of energy windows used for ew-sDFT calculation.

System	Method	$N_w$	$E_k$	$E_{nl}$	$E_{loc}$	$E_h + E_{xc}$	$E_{tot}$
Crystal $Si_{64}$	sDFT		10.450(40)	6.081(8)	-8.633(22)	-6.270(6)	-26.937(37)
	ew-sDFT	10	10.450(40)	6.081(8)	-8.641(19)	-6.271(6)	-26.946(36)
	Deterministic		10.450	6.083	-8.660	-6.266	-26.958
Crystal $Si_{512}$	sDFT		10.364(7)	6.105(3)	-8.581(6)	-6.291(2)	-26.969(7)
	ew-sDFT	10	10.363(6)	6.104(3)	-8.587(5)	-6.293(2)	-26.978(7)
	ew-sDFT	100	10.363(7)	6.104(3)	-8.589(5)	-6.293(2)	-26.980(7)
	Deterministic		10.360	6.105	-8.594	-6.292	-26.985
Thermal $Si_{512}$	sDFT		10.301(5)	6.104(3)	-8.790(5)	-6.190(1)	-26.942(8)
	ew-sDFT	10	10.298(6)	6.104(2)	-8.800(2)	-6.190(1)	-26.956(6)
	Deterministic		10.298	6.104	-8.804	-6.190	-26.959

energy windows. The wave function cutoff was 30 Ry and the density cutoff was 60 Ry, which corresponds to real space grid spacing of 0.22 Å. Troullier-Martins norm conserving pseudopotentials<sup>32</sup> were used with Kleinman-Bylander separable form.<sup>33</sup> The application of

the nonlocal pseudopotential was performed in real space to reduce the computational costs.<sup>34</sup> We first tested a rather small supercell containing 64 silicon atoms as an illustrative example of ew-sDFT. 512 stochastic orbitals were used in both sDFT and ew-sDFT and



**FIG. 3.** Selected average nuclei forces along x axis and standard deviations (error-bars) are shown for (a) crystal  $Si_{64}$ , (b) crystal  $Si_{512}$ , and (c) thermal  $Si_{512}$ . Panel (a): Blue and green symbols are nuclei forces from sDFT and ew-sDFT with 10 windows, respectively. Panel (b): Blue, green, and red dots are nuclei forces from sDFT, ew-sDFT with 10 windows, and ew-sDFT with 100 windows, respectively. Panel (c): The nuclei forces from for a thermally equilibrated  $Si_{512}$  structure using sDFT (blue symbols) and ew-sDFT (green symbols). Since the forces are significantly larger than in the perfectly ordered case, we subtract the corresponding deterministic value from each force. (d) The standard deviations of nuclei forces along the x axis averaged over all nuclei as a function of number of energy windows. The results for  $Si_{64}$  crystal,  $Si_{512}$  crystal, and thermally equilibrated  $Si_{512}$  are shown as blue, red, and green symbols, respectively. sDFT results correspond to 1 energy window.

10 energy windows were used in ew-sDFT. All statistical quantities were evaluated from 10 independent runs.

In addition to the perfect crystalline silicon structures, we have also compared the performance of sDFT and ew-sDFT on a thermally equilibrated silicon structure. The structure was obtained by equilibrating Si<sub>512</sub> crystal at 1000 K and ambient pressure with molecular dynamics simulation using the Stillinger and Weber force-field.<sup>35</sup> Due to smaller fundamental gap of thermal equilibrium silicon, we used  $\beta \approx 800$  in inverse Hartree to converge the ground state properties in both sDFT and ew-sDFT calculations. All the other computational details were not changed, and all statistical quantities were averaged over 5 independent runs.

In Fig. 1, we plot the standard deviation of  $\bar{\rho}_\chi(\mathbf{r})$  for sDFT (blue curves) and ew-sDFT (red curves). In addition, we plot the estimate of the standard deviation given by Eqs. (6) and (9) for sDFT and ew-sDFT (dashed lines), respectively. We find that the standard deviation of the electron density is roughly proportional to the value of  $\rho(\mathbf{r})$ , consistent with the theoretical predictions. Both theory and numerical calculations demonstrate that ew-sDFT reduces the statistical noise in the density. With 10 windows, the density fluctuation from ew-sDFT is reduced by nearly a factor of 3 compared to sDFT. Since the standard deviation scales as  $N_\chi^{-1/2}$ , it implies that the sDFT requires about 10 times more stochastic orbitals to achieve the same level of statistical noise in the density.

A similar picture emerges for the density of states (DOS), as shown in Fig. 2. We constructed the KS Hamiltonian  $\hat{h}_{\text{KS}}[\bar{\rho}_\chi(\mathbf{r})]$  from  $\bar{\rho}_\chi(\mathbf{r})$  and diagonalized it in the subspace of occupied orbitals. The DOS of  $\hat{h}_{\text{KS}}[\bar{\rho}_\chi(\mathbf{r})]$  with  $\bar{\rho}_\chi(\mathbf{r})$  from ew-sDFT is in better agreement with the deterministic result (upper panel of Fig. 2) and the fluctuations in the DOS are smaller (lower panel of Fig. 2).

Reducing the noise in the electron density should lead to a reduction of the noise of *all* ground state properties. To test this hypothesis, we calculated the energy per electron and nuclei forces for Si<sub>64</sub> and Si<sub>512</sub> supercells at  $T = 0$  K and for Si<sub>512</sub> supercell at  $T = 1000$  K and compared the results of ew-sDFT to sDFT. 10 energy windows were used for all systems. In addition, to explore the effect of the number of windows, we have used 100 energy windows for the crystal Si<sub>512</sub>. The energies per electron calculated by sDFT and ew-sDFT agree well with deterministic results, as presented in Table I. Surprisingly, the standard deviations are similar comparing sDFT and ew-sDFT. Increasing the number of windows to 100 does not change this picture.

Contrary to the case of the total energy, the standard deviation in the nuclei forces is significantly reduced by the introduction of energy windows, as shown in Fig. 3. Increasing the number of windows in ew-sDFT to 100 further reduces the noise in the nuclei forces [Fig. 3 panel (d)], however, the standard deviation seems to converge to a lower limit. This suggests that there should exist an optimal window size (close to 10) balancing between the small additional computational overhead associated with increasing the number of windows and the reduction in the statistical noise. In addition to the role of the number of energy windows, Fig. 3 panel (d) clearly demonstrates that the noise in the nuclei forces does not depend on the system size, implying that ew-sDFT scales linearly with the system size for nuclei forces.

Turning to discuss the role of thermal fluctuations and the introduction of disorder, comparing the standard deviation of the

nuclei forces in Fig. 3 panels (c) and (d), we find that the reduction in the noise in the nuclei forces using 10 energy windows is not limited to a perfectly ordered structure at  $T = 0$  and a similar reductions are observed for the thermally equilibrated structure at  $T = 1000$  K. This signifies that the sort of self-averaging observed here is not a consequence of the perfectly ordered structure.

## V. FLUCTUATIONS OF OBSERVABLES AND DENSITY FUNCTIONALS

In this section, we analyze the fluctuations in observables described by the trace over a generic operator  $\hat{O}$  and rationalize the behavior depicted in Table I and Fig. 3, where reduction in the fluctuations was observed for the forces but not the total energies per particle. We begin by focusing on a simple case in which the operator  $\hat{O}$  does not depend on the electron density,  $\bar{\rho}_\chi(\mathbf{r})$ . In the stochastic formulation [cf. Eq. (4)], the expectation value is approximated by  $O_\chi = 2\langle \xi | \hat{O} | \xi \rangle$ . The variance of  $O_\chi$  can be calculated by the following equation:

$$\text{Var}_\chi(O_\chi) = 8 \sum_{i,j=1}^{N_{\text{occ}}} \langle \psi_i | \hat{O} | \psi_j \rangle^2, \quad (11)$$

where  $\psi_i$  and  $\psi_j$  are deterministic KS occupied orbitals. We need to generate the KS orbital for the analysis, but of course, ew-sDFT does not require any knowledge of the KS orbitals. The variance given by Eq. (11) is determined by the Frobenius norm of the matrix representation of  $\hat{O}$  in the occupied subspace. In the case of ew-sDFT, the variance is given by the sum of variances for each window (for simplicity we take only two windows)

$$\text{Var}_\chi(O_\chi) = 8 \left( \sum_{i,j \in w_1} \langle \psi_i | \hat{O} | \psi_j \rangle^2 + \sum_{i,j \in w_2} \langle \psi_i | \hat{O} | \psi_j \rangle^2 \right). \quad (12)$$

Recalling that  $w_1$  ( $w_2$ ) is the set of orbital indices for the first (second) energy window, Eq. (12) suggests that the variance of  $O_\chi$  is proportional to the Frobenius norms of diagonal blocks of  $\langle \psi_i | \hat{O} | \psi_j \rangle$  (see Fig. 4). If  $\langle \psi_i | \hat{O} | \psi_j \rangle$  is diagonal dominant, the variance calculated from Eqs. (11) and (12) is nearly identical. In this case, ew-sDFT is less efficient at reducing the statistical fluctuations in  $O_\chi$ . In the limiting case in which  $\psi_i$  are eigenfunctions of  $\hat{O}$ , the noise level of  $O_\chi$  calculated from ew-sDFT is identical to that of sDFT

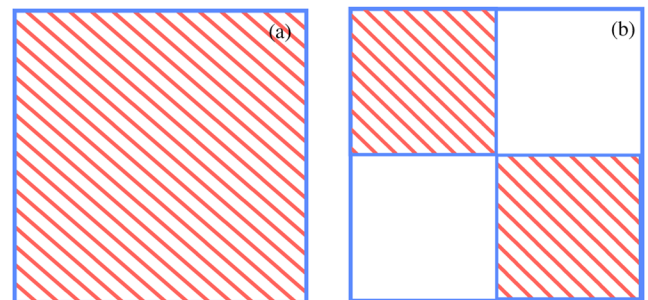


FIG. 4. The variance of  $O_\chi$  depends on (a) the Frobenius norm of the whole matrix (shaded zone) in sDFT and (b) the Frobenius norm of the diagonal blocks (shaded zone) in ew-sDFT.

(a single window ew-sDFT). On the hand, if  $\langle \psi_i | \hat{O} | \psi_j \rangle$  is off-diagonal dominant, significant noise reduction can be achieved by ew-sDFT, as suggested by comparing Eqs. (11) and (12).

To illustrate this, consider the stochastic fluctuations in the kinetic energy, nonlocal pseudopotential energy, and local pseudopotential energy, which are described by the operators  $\hat{t}$ ,  $\hat{v}_{\text{nl}}$ , and  $\hat{v}_{\text{loc}}$  [independent on the electron density,  $\bar{\rho}_\chi(\mathbf{r})$ ]. The corresponding matrix elements calculated in the KS eigenstates representation are shown in Fig. 5 panels (a)–(c), respectively. The matrices are clearly diagonal dominant for all three cases [panels (a)–(c)]. Thus, it is not surprising that ew-sDFT does not reduce the noise for the corresponding observables,  $E_k$ ,  $E_{\text{nl}}$ , and  $E_{\text{loc}}$  (as discussed above, see also Table I). On the contrary, matrices of nuclei forces are not diagonal dominant, as illustrated in Fig. 5 panels (d)–(f), consistent with the ew-sDFT results shown in Fig. 3. We note in passing that for amorphous materials, off-diagonal terms in the matrices of  $\hat{t}$ ,  $\hat{v}_{\text{nl}}$ , and  $\hat{v}_{\text{loc}}$  become important, as shown in the [supplementary material](#). However, these matrices are still diagonal dominant and thus noise reduction in energy terms are limited, as illustrated in the thermal equilibrium Si<sub>512</sub> example in Table I. On the other hand, for these amorphous materials, the structure of the corresponding matrices for the force remain nondiagonal dominant and ew-sDFT can achieve excellent noise reduction in atom forces, which is also demonstrated by the thermal equilibrium Si<sub>512</sub> example in Fig. 3.

Next, we consider the case where the operator  $\hat{O}$  depends on the electron density,  $\bar{\rho}_\chi(\mathbf{r})$ . This is important for observables like the Hartree energy and the exchange-correlation energy required to fully understand the behavior with respect to the number of energy

windows observed for the total energy. Consider the variance of a density functional  $f[\bar{\rho}_\chi(\mathbf{r})]$ , where  $\bar{\rho}_\chi(\mathbf{r})$  is calculated from  $N_\chi$  stochastic orbitals in sDFT or ew-sDFT. Approximating  $\bar{\rho}_\chi(\mathbf{r})$  as  $\bar{\rho}_\chi(\mathbf{r}) \approx \rho(\mathbf{r})(1 + \alpha(\mathbf{r}))$ , where  $\alpha(\mathbf{r})$  is a small random perturbation of order  $\sqrt{\epsilon}$  where  $\epsilon = 1/N_\chi$ , and retaining terms to order  $\epsilon$  leads to

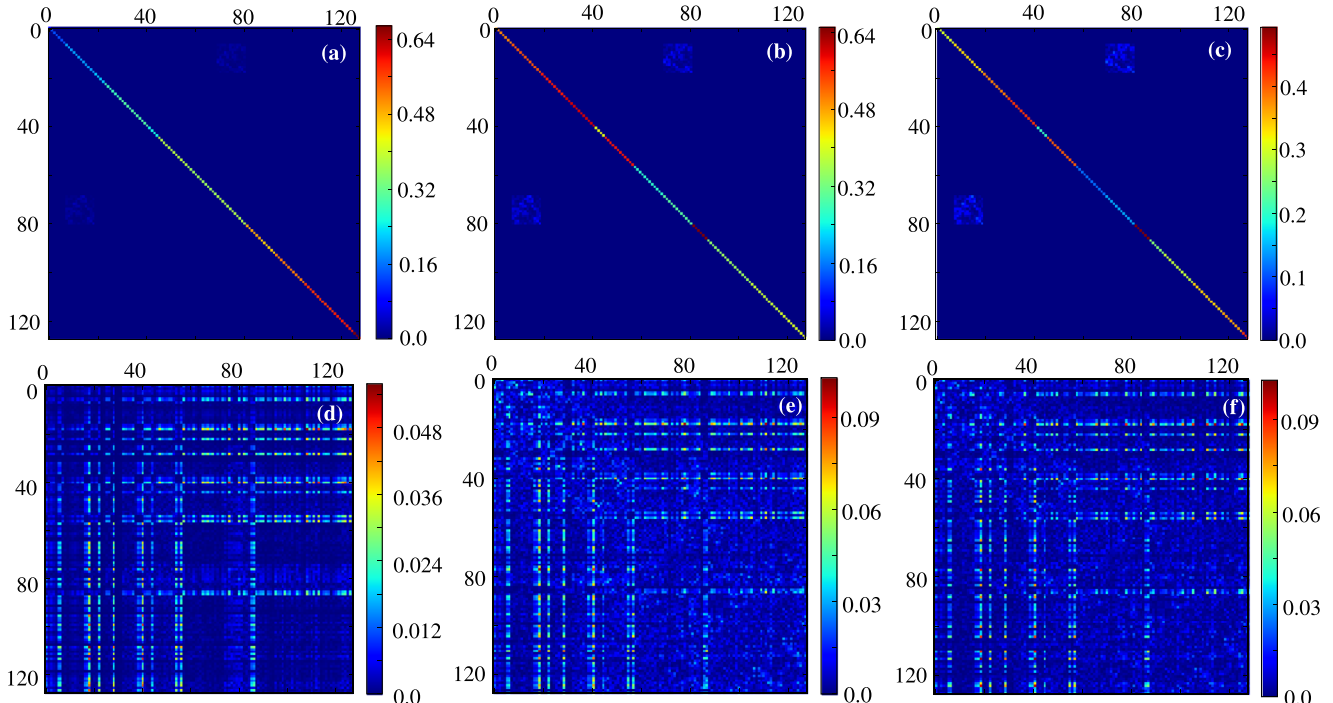
$$\langle f[\bar{\rho}_\chi(\mathbf{r})] \rangle = f[\rho(\mathbf{r})] + \epsilon \iint \frac{\delta^2 f}{\delta \rho(\mathbf{r}) \delta \rho(\mathbf{r}')} \rho^2(\mathbf{r}, \mathbf{r}') d\mathbf{r} d\mathbf{r}' + O(\epsilon^{3/2}), \quad (13)$$

$$\text{Var}(f[\bar{\rho}_\chi(\mathbf{r})]) = 8\epsilon \sum_{i,j=1}^{N_{\text{occ}}} \left( \left\langle \psi_i \left| \frac{\delta f}{\delta \rho} \right| \psi_j \right\rangle \right)^2 + O(\epsilon^{3/2}), \quad (14)$$

where, as before,  $\rho(\mathbf{r}, \mathbf{r}') = 2\langle \mathbf{r} | \hat{\rho} | \mathbf{r}' \rangle$  is the single particle density matrix. For linear functional like the local pseudopotential energy,  $\langle f[\bar{\rho}_\chi(\mathbf{r})] \rangle = f[\rho(\mathbf{r})]$  and the functional variance given by Eq. (14) is identical to applying the central limit theorem to the variance given by Eq. (11) with  $\hat{O}$  replaced by a local effective operator  $\frac{\delta f}{\delta \rho}[\rho(\mathbf{r})](\mathbf{r})$ .

For a nonlinear functional,  $\iint \frac{\delta^2 f}{\delta \rho(\mathbf{r}) \delta \rho(\mathbf{r}')} \rho^2(\mathbf{r}, \mathbf{r}') d\mathbf{r} d\mathbf{r}'$  is nonzero, corresponding to “bias” which decays as  $1/N_\chi$  as recently discussed by Cytter *et al.*<sup>27</sup> and Fabian *et al.*<sup>28</sup> In ew-sDFT, the average and variance of  $f[\bar{\rho}_\chi(\mathbf{r})]$  becomes

$$\langle f[\bar{\rho}_\chi(\mathbf{r})] \rangle = f[\rho(\mathbf{r})] + \epsilon \iint \frac{\delta^2 f}{\delta \rho(\mathbf{r}) \delta \rho(\mathbf{r}')} (\rho_1^2(\mathbf{r}, \mathbf{r}') + \rho_2^2(\mathbf{r}, \mathbf{r}')) d\mathbf{r} d\mathbf{r}' + O(\epsilon^{3/2}), \quad (15)$$



**FIG. 5.** The matrices  $\langle \psi_i | \hat{O} | \psi_j \rangle$  are plotted for (a)  $\hat{O} = \hat{t}$ , (b)  $\hat{O} = \hat{v}_{\text{nl}}$ , and (c)  $\hat{O} = \hat{v}_{\text{loc}}$ , (d)  $\hat{O} = \frac{\partial \hat{v}_{\text{loc}}}{\partial X_1}$ , (e)  $\hat{O} = \frac{\partial \hat{v}_{\text{nl}}}{\partial X_1}$ , and (f)  $\hat{O} = \frac{\partial \hat{v}_{\text{loc}}}{\partial X_1} + \frac{\partial \hat{v}_{\text{nl}}}{\partial X_1}$ , where  $X_1$  is the  $x$  coordinate of the first atom.

$$\text{Var}(f[\tilde{\rho}_X(\mathbf{r})]) = 8\epsilon \left( \sum_{i,j \in \omega_1} \left( \left\langle \psi_i \left| \frac{\delta f}{\delta \rho} \right| \psi_j \right\rangle \right)^2 + \sum_{i,j \in \omega_2} \left( \left\langle \psi_i \left| \frac{\delta f}{\delta \rho} \right| \psi_j \right\rangle \right)^2 \right) + O(\epsilon^{3/2}). \quad (16)$$

Using the above relations in Eqs. (11)–(16) results in the following expression for the variance of total energy:

$$\text{Var}(E_{\text{tot}}) = 8\epsilon \sum_{i,j=1}^{N_{\text{occ}}} \left( \langle \psi_i | \hat{h}_{\text{KS}} | \psi_j \rangle \right)^2 = 8\epsilon \sum_{i=1}^{N_{\text{occ}}} \epsilon_i^2 + O(\epsilon^{3/2}) \quad (17)$$

for sDFT and

$$\begin{aligned} \text{Var}(E_{\text{tot}}) &= 8\epsilon \left( \sum_{i,j \in \omega_1} \left( \langle \psi_i | \hat{h}_{\text{KS}} | \psi_j \rangle \right)^2 + \sum_{i,j \in \omega_2} \left( \langle \psi_i | \hat{h}_{\text{KS}} | \psi_j \rangle \right)^2 \right) + O(\epsilon^{3/2}) \\ &= 8\epsilon \sum_{i=1}^{N_{\text{occ}}} \epsilon_i^2 + O(\epsilon^{3/2}) \end{aligned} \quad (18)$$

for ew-sDFT. The expressions for the total energy variance given by Eqs. (17) and (18) are identical, suggesting that the statistical noise in ew-sDFT is identical to that of sDFT, consistent with the results shown above. We wish to emphasize that the results in Eqs. (17) and (18) also support the observed sublinear scaling of the energy per electron with both sDFT and ew-sDFT, since the occupied orbital energies are bounded, and thus, the variance of the total energy is proportional to the system size. This implies that the standard deviation of the energy per electron is inverse proportional to the square root of the system size, explaining our previous observations.<sup>23,24</sup>

## VI. CONCLUSION

In this study, we developed a stochastic density functional theory scheme which allows us to significantly reduce the statistical noise in certain ground state observable with DFT. This was achieved by dividing the occupied space into energy windows and using correlated sampling in each window. Our new scheme (ew-sDFT) is able to reduce the statistical noise in both the electron density and nuclei forces without significantly increasing the computational time, but not for the total energy or total energy per electron. We showed that noise reduction of an observable in ew-sDFT correlates with whether or not the corresponding operator is diagonal dominant in KS eigenstate representation. The total energy is diagonal dominant, and no noise reduction was observed. However, for the forces and electron density, this was not the case and we observed a significant reduction in the noise that results in nearly an order of magnitude reduction in the computational effort compared to sDFT (since the noise is proportional to  $1/\sqrt{N_X}$ ). This reduction allows for a more rapid sampling of the canonical distribution using methods that do not rely on the total energy, such as Langevin dynamics.<sup>26</sup> The developed ew-sDFT provides an additional approach to noise reduction within the family of sDFT methods and can be combined with embedded fragmentation schemes to further improve the accuracy and efficiency of stochastic density functional theories.

## SUPPLEMENTARY MATERIAL

Detailed derivations for variance of ground state properties in sDFT and ew-sDFT are provided in the [supplementary material](#). The [supplementary material](#) also includes operator matrices for disordered Si<sub>64</sub> system.

## ACKNOWLEDGMENTS

R.B. gratefully acknowledges support from the Israel Science Foundation, Grant No. 800/19. D.N. and E.R. are grateful for support by the Center for Computational Study of Excited State Phenomena in Energy Materials (C2SEPEM) at the Lawrence Berkeley National Laboratory, which is funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division under Contract No. DEAC02-05CH11231 as part of the Computational Materials Sciences Program. Resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231 are greatly acknowledged.

## REFERENCES

- 1 P. Hohenberg and W. Kohn, *Phys. Rev.* **136**, B864 (1964).
- 2 W. Kohn and L. J. Sham, *Phys. Rev.* **140**, A1133 (1965).
- 3 S. Baroni, S. de Gironcoli, A. Dal Corso, and P. Giannozzi, *Rev. Mod. Phys.* **73**, 515 (2001).
- 4 T. Frauenheim, G. Seifert, M. Elstner, T. Niehaus, C. Kohler, M. Amkreutz, M. Sternberg, Z. Hajnal, A. D. Carlo, and S. Suhai, *J. Phys.: Condens. Matter* **14**, 3015 (2002).
- 5 C. Freysoldt, B. Grabowski, T. Hickel, J. Neugebauer, G. Kresse, A. Janotti, and C. G. Van de Walle, *Rev. Mod. Phys.* **86**, 253 (2014).
- 6 F. Mauri, G. Galli, and R. Car, *Phys. Rev. B* **47**, 9973 (1993).
- 7 P. Ordejón, D. A. Drabold, M. P. Grumbach, and R. M. Martin, *Phys. Rev. B* **48**, 14646 (1993).
- 8 S. Goedecker, *J. Comput. Phys.* **118**, 261 (1995).
- 9 E. Hernández and M. J. Gillan, *Phys. Rev. B* **51**, 10157 (1995).
- 10 W. Kohn, *Phys. Rev. Lett.* **76**, 3168 (1996).
- 11 A. H. R. Palser and D. E. Manolopoulos, *Phys. Rev. B* **58**, 12704 (1998).
- 12 R. Baer and M. Head-Gordon, *Phys. Rev. Lett.* **79**, 3962 (1997).
- 13 S. Goedecker, *Rev. Mod. Phys.* **71**, 1085 (1999).
- 14 W. Yang, *Phys. Rev. Lett.* **66**, 1438 (1991).
- 15 P. Cortona, *Phys. Rev. B* **44**, 8454 (1991).
- 16 T. Zhu, W. Pan, and W. Yang, *Phys. Rev. B* **53**, 12713 (1996).
- 17 F. Shimajo, R. K. Kalia, A. Nakano, and P. Vashishta, *Phys. Rev. B* **77**, 085103 (2008).
- 18 W. Yang and T. Lee, *J. Chem. Phys.* **103**, 5674 (1995).
- 19 A. W. Götz, S. M. Beyhan, and L. Visscher, *J. Chem. Theory Comput.* **5**, 3161 (2009).
- 20 T. A. Wesolowski, S. Shedge, and X. Zhou, *Chem. Rev.* **115**, 5891 (2015).
- 21 J. D. Goodpaster, T. A. Barnes, and T. F. Miller, *J. Chem. Phys.* **134**, 164108 (2011).
- 22 C. Huang and E. A. Carter, *J. Chem. Phys.* **135**, 194104 (2011).
- 23 R. Baer, D. Neuhauser, and E. Rabani, *Phys. Rev. Lett.* **111**, 106402 (2013).
- 24 D. Neuhauser, R. Baer, and E. Rabani, *J. Chem. Phys.* **141**, 041102 (2014).
- 25 M. Chen, R. Baer, D. Neuhauser, and E. Rabani, *J. Chem. Phys.* **150**, 034106 (2019).
- 26 E. Arnon, E. Rabani, D. Neuhauser, and R. Baer, *J. Chem. Phys.* **146**, 224111 (2017).
- 27 Y. Cytter, E. Rabani, D. Neuhauser, and R. Baer, *Phys. Rev. B* **97**, 115207 (2018).

<sup>28</sup>M. D. Fabian, B. Shpiro, E. Rabani, D. Neuhauser, and R. Baer, "Stochastic density functional theory," *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* (published online).

<sup>29</sup>R. Kosloff, *J. Phys. Chem.* **92**, 2087 (1988).

<sup>30</sup>R. Kosloff, *Annu. Rev. Phys. Chem.* **45**, 145 (1994).

<sup>31</sup>R. Baer and M. Head-Gordon, *J. Chem. Phys.* **107**, 10003 (1997).

<sup>32</sup>N. Troullier and J. L. Martins, *Phys. Rev. B* **43**, 1993 (1991).

<sup>33</sup>L. Kleinman and D. M. Bylander, *Phys. Rev. Lett.* **48**, 1425 (1982).

<sup>34</sup>R. D. King-Smith, M. C. Payne, and J. S. Lin, *Phys. Rev. B* **44**, 13063 (1991).

<sup>35</sup>F. H. Stillinger and T. A. Weber, *Phys. Rev. B* **31**, 5262 (1985).