Overlapped embedded fragment stochastic density functional theory for covalently-bonded materials

Ming Chen, Roi Baer, Daniel Neuhauser, and Eran Rabani

ARTICLES YOU MAY BE INTERESTED IN

Direct evaluation of the force constant matrix in quantum Monte Carlo
The Journal of Chemical Physics 150, 034104 (2019); https://doi.org/10.1063/1.5070138

Wavefunctions of macroscopic electron systems
The Journal of Chemical Physics 150, 030901 (2019); https://doi.org/10.1063/1.5050329

Prediction of many-electron wavefunctions using atomic potentials: Refinements and extensions to transition metals and large systems
The Journal of Chemical Physics 150, 034107 (2019); https://doi.org/10.1063/1.5064781
Overlapped embedded fragment stochastic density functional theory for covalently-bonded materials

Cite as: J. Chem. Phys. 150, 034106 (2019); doi: 10.1063/1.5064472
Submitted: 4 October 2018 · Accepted: 26 December 2018 · Published Online: 18 January 2019

Ming Chen,1 Roi Baer,1,4 Daniel Neuhauser,3 and Eran Rabani1,4

AFFILIATIONS
1Department of Chemistry, University of California, and Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA
2Fritz Haber Center for Molecular Dynamics and Institute of Chemistry, The Hebrew University of Jerusalem, Jerusalem 91904, Israel
3Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095, USA
4The Raymond and Beverly Sackler Center of Computational Molecular and Materials Science, Tel Aviv University, Tel Aviv 69978, Israel

ABSTRACT
The stochastic density functional theory (DFT) [R. Baer et al., Phys. Rev. Lett. 111, 106402 (2013)] is a valuable linear-scaling approach to Kohn-Sham DFT that does not rely on the sparsity of the density matrix. Linear (and often sub-linear) scaling is achieved by introducing a controlled statistical error in the density, energy, and forces. The statistical error (noise) is proportional to the inverse square root of the number of stochastic orbitals and thus decreases slowly; however, by dividing the system into fragments that are embedded stochastically, the statistical error can be reduced significantly. This has been shown to provide remarkable results for non-covalently-bonded systems; however, the application to covalently bonded systems had limited success, particularly for delocalized electrons. Here, we show that the statistical error in the density correlates with both the density and the density matrix of the system and propose a new fragmentation scheme that elegantly interpolates between overlapped fragments. We assess the performance of the approach for bulk silicon of varying supercell sizes (up to \(N_e = 16\,384\) electrons) and show that overlapped fragments reduce significantly the statistical noise even for systems with a delocalized density matrix.

Published under license by AIP Publishing. https://doi.org/10.1063/1.5064472

I. INTRODUCTION

An accurate description of the electronic properties is often a prerequisite for understanding the behavior of complex materials. Density functional theory (DFT) within the Kohn-Sham (KS) formulation provides an excellent framework that balances computational complexity and accuracy and thus has been the method of choice for extended, large-scale systems. Solving the KS equations scales formally as \(O(N_e^3)\), where \(N_e\) is the number of electrons in the system. Traditional implementations of KS-DFT are often limited to relative systems containing \(N_e < 10^4\) electrons, even with the high-performance computer architectures of today. Clearly, there is a need for low-scaling KS-DFT approaches.

A natural approach to linear-scaling KS-DFT relies on the "locality" of the density matrix, that is, the elements of the density matrix \(\rho(r, r')\) decay with increasing \(\|r - r'\|\). The reliance on Kohn's "nearsightedness" principle makes these approaches sensitive to the dimensionality and the character of the system. They work extremely well for low dimensional structures or systems with a large fundamental gap, but in 3-dimensions (3D), linear-scaling is achieved only for very large systems, typically for \(N_e > 10^4\). An alternative to Kohn's "nearsightedness" principle is based on "divide" and "conquer," where the density or the density matrix is partitioned into fragments and the KS equations are solved for each fragment separately using localized basis sets. The complexity of solving the KS equations is shifted...
to computing the interactions between the fragments, and the accuracy (systematic errors that decrease with the fragment size) and scaling depend on the specific implementation.

Recently, we have proposed a new scheme for linear-scaling DFT based approach. Stochastic DFT (sDFT) sidesteps the calculation of the density matrix and is therefore not directly sensitive to its evasive sparseness. Instead, the density is given in terms of a trace formula and is evaluated using stochastic occupied orbitals generated by a Chebyshev or Newton expansion of the occupation operator. Random fluctuations of local properties, e.g., energy per atom, forces, and density of states, are controlled by the number of stochastic orbitals and are often independent of the system size, leading to a computational cost that scales linearly (and sometimes sub-linearly).

The statistical error is substantially reduced when, instead of sampling the full density, one samples the difference between the full system density and that of reference fragments. Indeed, embedded fragment stochastic DFT (efs-DFT) has been extremely successful in reducing the noise level for non-covalently bonded fragments. However, for covalently bonded fragments, efsDFT needs to be structured appropriately, particularly for periodic boundary conditions (PBCs).

In this manuscript, we have developed a new fragmentation approach suitable for covalently bonded systems with open or periodic boundary conditions (PBCs). Similar to efs-DFT, we divide the system into core fragments, but in addition, for each core fragment, we add a buffer zone allowing for fragments to overlap in the embedding procedure. Similar in spirit to an earlier work on divide and conquer, we demonstrate that allowing the fragments to overlap results in significant improvements of the reference density matrix, reducing significantly the level of statistical noise. This manuscript is organized as follows: In Sec. II, we briefly review the theory of stochastic orbital DFT and its embedded-frAGMENTED version. In Sec. III, we outline the new overlapped embedded fragment stochastic DFT (o-efsDFT). Section V presents the results of numerical calculations on a set of silicon crystals of varying super-cell size, up to 4096 atoms, and discusses the significance of the overlapped fragments on reducing the statistical error.

II. STOCHASTIC DENSITY FUNCTIONAL THEORY AND EMBEDDED FRAGMENT STOCHASTIC DENSITY FUNCTIONAL THEORY

We consider a system with periodic boundary conditions in a super-cell of volume \( V \) with an electron density \( \rho(\mathbf{r}) \) represented on a real-space grid with \( N_0 \) grid points. The electronic properties are described within Kohn-Sham (KS) density functional theory, where the density is given by

\[ \rho(\mathbf{r}) = \sum_{\alpha=1}^{N_{\text{occ}}} \langle \psi_{\alpha} | \Psi_{\alpha} | \mathbf{r} \rangle \psi_{\alpha}(\mathbf{r}) \]

Here, \( \langle \mathbf{r} | \psi_{\alpha} \rangle \) are the KS orbitals and \( N_{\text{occ}} \) is the number of occupied states. The KS density can also be expressed as a trace over the density matrix, \( \rho = \lim_{\beta \to \infty} \theta_{\beta}(\mu - \hat{h}_{\text{KS}}) \),

\[ \rho(\mathbf{r}) = \text{Tr} \theta_{\beta}(\mu - \hat{h}_{\text{KS}}) \delta(\mathbf{r} - \mathbf{r}'), \]

where \( \delta(\mathbf{r} - \mathbf{r}') \) is Dirac’s delta function and \( \theta_{\beta}(\xi) = \left(1 + e^{-\beta \xi}\right)^{-1} \) is a step function in the limit \( \beta \to \infty \) (in practice, \( \beta \) is chosen to be large enough to converge the ground state properties). The chemical potential, \( \mu \), is determined by imposing the relation \( N_0(\mu) = \text{Tr} \theta_{\beta}(\mu - \hat{h}_{\text{KS}}) \), where \( N_0(\mu) \) is the total number of electrons in the system. In Eq. (1), \( \hat{h}_{\text{KS}} \) is the KS Hamiltonian, given by

\[ \hat{h}_{\text{KS}} = \hat{t} + \hat{v}_{\text{loc}} + \hat{v}_{\text{H}} + \hat{v}_{\text{XC}}, \]

where \( \hat{t} \) is the kinetic energy, \( \hat{v}_{\text{loc}} \) and \( \hat{v}_{\text{H}} \) are the local and nonlocal pseudopotentials, respectively, \( \hat{v}_{\text{XC}} \) is the Hartree potential, and \( \hat{v}_{\text{XC}} \) is the exchange-correlation potential.

Equation (1) is the starting point for the derivation of the stochastic DFT. The trace is performed using \( N_\chi \) stochastic orbitals, \( |\chi\rangle \), leading to

\[ \rho(\mathbf{r}) = \langle \chi | \theta_{\beta}(\mu - \hat{h}_{\text{KS}}) \delta(\mathbf{r} - \mathbf{r}') | \chi \rangle = \langle \xi | \theta_{\beta}(\mu - \hat{h}_{\text{KS}}) | \xi \rangle = \langle \xi | \chi \rangle^2 = \langle \xi | \chi \rangle^2, \]

where \( |\xi\rangle = \sqrt{\theta_{\beta}(\mu - \hat{h}_{\text{KS}})} |\chi\rangle \) is a projected stochastic orbital. In Eq. (3), \( \langle \cdot | \cdot \rangle = \langle \cdot | \cdot \rangle_{\chi} \), where \( \Delta V = V/N_\chi \) is the volume element of the real-space grid and \( \hat{\chi} \) is a projected stochastic orbital. Equation (3) becomes exact only in the limit that \( N_\chi \to \infty \), and otherwise it is a stochastic approximation to the electron density, with a random error whose magnitude decays as \( \sqrt{N_\chi} \). Equation (3) is solved self-consistently, since the KS Hamiltonian depends on \( \rho(\mathbf{r}) \) itself.

Obtaining the total energy within the sDFT formalism is tricky. The Hartree, local pseudopotential, and local exchange-correlation energy terms can be obtained directly from the stochastic density, similar to deterministic DFT. The calculation of the kinetic energy and nonlocal pseudopotential energy requires the Kohn–Sham orbitals in deterministic DFT. In sDFT, these can be evaluated using the relations

\[ E_k = \text{Tr} \hat{t} \hat{n}_\chi = \langle \xi | \hat{\chi} | \xi \rangle, \]

\[ E_{\text{nl}} = \text{Tr} \hat{v}_{\text{nl}} = \langle \xi | \hat{\chi}_{\text{nl}} | \xi \rangle, \]

The advantage of the stochastic DFT formalism is that the density can be calculated with linear (and often sub-linear) scaling at the cost of introducing a controlled statistical error, which is controlled by \( \sqrt{N_\chi} \), where \( N_\chi \) is the number of stochastic orbitals. Linear-scaling is achieved by taking advantage of the sparsity of \( \hat{h}_{\text{KS}} \) in real- and reciprocal-space representations and approximating \( \hat{\chi} = \sqrt{\theta_{\beta}(\mu - \hat{h}_{\text{KS}})} \hat{\chi} \) in Eq. (3) using Chebyshev or Newton interpolation polynomials, cf. \( \sqrt{\theta_{\beta}(\mu - \hat{h}_{\text{KS}})} = \sum_{n=0}^{N_c} a_n(\mu) T_n(\hat{h}_{\text{KS}}) \), where \( N_c \) is the length of the polynomial chain and \( T_n(\hat{h}_{\text{KS}}) \) are the Chebyshev-
polynomials. Instead of directly evaluating $T_n(h_{KS})$ from powers of $h_{KS}$, sDFT only requires the application of $T_n(h_{KS})$ to $|\chi\rangle$ via the recurrence relation of the Chebyshev polynomial, i.e., $|\chi^{(n+1)}\rangle = T_n(h_{KS})|\chi^{(n)}\rangle - 2h_{KS}|\chi^{(n-1)}\rangle$. In situations where the statistical noise does not increase with the system size, sDFT scales linearly with the system size. Often, for certain local observables, we find that sDFT scales better than linear (sub-linear) due to self-averaging. This has been recently discussed in detail in Ref. 27.

The decay of the magnitude of the statistical error with $\sqrt{N_x}$ implies that in order to reduce the error by an order of magnitude, one has to increase the number of stochastic orbitals by two orders of magnitude. Thus, small statistical errors required to converge local and single particle observables often require a huge set of stochastic orbitals, making sDFT impractical. Since the standard deviation of the density is proportional to $\sqrt{\rho}$ itself (as shown in the supplementary material), the statistical noise can be reduced by decomposing the density into a reference density, $\rho_{ref}(r)$, and correction terms $\Delta \rho(r) \ll \rho(r)$ such that $\rho(r) = \rho_{ref}(r) + \Delta \rho(r)$. This is the central idea behind embedded-fragmented stochastic DFT (efsDFT).27,29

The system is divided into non-overlapping fragments, and the reference density is decomposed as a sum of the fragment densities. This has led to a significant reduction of the noise in efsDFT compared to sDFT when fragments are closed-shell molecules in non-covalently bonded systems.28 However, for covalently bonded materials, it is necessary to break covalent bonds when fragmenting the system. This leads to an increasingly large statistical error of the density at the boundaries between the fragments (see Fig. 1 for an illustration of this effect for Si$_{28}$ with PBCs where the system was divided into two non-overlapping Si$_{64}$ fragments), leading to small improvements of efsDFT with respect to sDFT.27,29

II. OVERLAPPED EMBEDDED FRAGMENTED STOCHASTIC DFT

To overcome the limitation of the efsDFT in covalently bonded fragments, we propose to use overlapping fragments, as sketched in Fig. 2. The basic idea behind the proposed overlapped embedded fragmented stochastic DFT (o-efsDFT) is that each fragment density and density matrix is calculated with a buffer region, but the projection of the fragment density is limited to the core region only. This allows for continuous densities and density matrices across the boundaries between the core fragments, thereby resolving the issue discussed above and allowing for a significant reduction of the statistical error, even for systems with periodic boundary conditions and a delocalized density matrix.

As sketched in Fig. 2, the system is divided into non-overlapping fragments that are composed of unit cells and/or small super cells referred to as “core regions,” labeled $C_f$ for fragment $f$. Each core region is then wrapped with a “buffer” region ($B_f$), and a dressed fragment (labeled $D_f = C_f \cup B_f$) is defined as the sum of core and buffered regions. For any $r \in C_f$, the dressed fragment density matrix $\langle r | \hat{\rho} | r' \rangle$ is given by

![Buffer Region (B_f)](image)

![Core Region (C_f)](image)
\[ \langle r | \hat{\rho}_i | r' \rangle = \begin{cases} \sum_{j=1}^{N_{\text{occ}}} \langle r | \varphi_j \rangle \langle \varphi_j | r' \rangle & r' \in D_f \\ 0 & r' \notin D_f \end{cases} . \] (6)

In Eq. (6), \( \varphi_j(r) \) are KS occupied orbitals for the dressed fragment \( f \) obtained by a deterministic DFT in region \( D_f \), where we impose periodic boundary conditions for each dressed region separately. The closest distance between the boundaries \( C_f \) and \( D_f \) (\( \Delta r \), see Fig. 2) is chosen so that periodic boundary conditions can be imposed on region \( D_f \). Note that the above definition of \( \langle r | \hat{\rho}_i | r' \rangle \) is not necessarily hermitian and idempotent. However, this turns out to be insignificant, since the density matrix of the entire system remains hermitian regardless of how \( \langle r | \hat{\rho}_i | r' \rangle \) behaves (see more detail below). In the limit where \( \Delta r = 0 \), o-efsDFT is identical to efsDFT.\(^{28}\)

With the definition of \( \langle r | \hat{\rho}_i | r' \rangle \) in Eq. (6), the o-efsDFT density at position \( r \in C_f \) is given by

\[ \rho(r) = \langle r | \hat{\rho}_i | r \rangle + \langle i | \xi(r) \rangle \langle i | \xi(r) \rangle - \langle r | \hat{\rho}_i | x \rangle \langle x | \hat{\rho}_i | r \rangle \]

\[ = \rho_f(r) + \langle i | \xi(r) \rangle \langle i | \xi(r) \rangle - \langle i | \xi(r) \rangle \langle i | \xi(r) \rangle , \] (7)

where \( \xi(r) = \sum_{i=1}^{N_{\text{occ}}} \varphi_i^\dagger(r) \varphi_i (|x \rangle \langle x|) \) and \( f(y) = f_D \, drf^\dagger (|y \rangle \langle y|) \), i.e., integration over the dressed fragment region. It is easy to verify that the density in Eq. (7) satisfies the following requirements:

(a) In the infinite sampling limit, the properties calculated from o-efsDFT converge to deterministic DFT results.

(b) If \( C_f \) is the system itself, the properties calculated from o-efsDFT are equal to those from the deterministic DFT calculation regardless of the number of stochastic orbitals.

(c) Given any partitioning of the system into core regions, in the limit where the buffered zone grows such that \( D_f \) represents the entire system, the properties calculated from o-efsDFT will again be equivalent to the deterministic results with any number of stochastic orbitals.

We now turn to discuss the calculations of kinetic energy, the nonlocal pseudopotential energy, and the nonlocal nuclear force for each atom with a position \( R_{\text{atom}} \in C_f \) is given by

\[ F_{\text{atom}}^{\text{nl}} = \sum_{i=1}^{N_{\text{occ}}} \langle \varphi_i | p_{\text{atom}} | \varphi_i \rangle + \left\langle \xi | p_{\text{atom}} | \xi \right\rangle , \] (9)

\[ - \sum_{i=1}^{N_{\text{occ}}} \left\langle \langle \xi | p_{\text{atom}} | \xi \rangle \right\rangle , \] (10)

It is straightforward to show that Eqs. (9) and (10) satisfy all three requirements above. Even with overlapped fragmentation, \( F_{\text{atom}}^{\text{nl}} \) is still a random variable with a finite number of stochastic orbitals. Random noise in \( F_{\text{atom}}^{\text{nl}} \) can be designed as a part of the temperature control algorithm in \textit{ab initio} Langevin dynamics, as described in Ref. 29.

IV. IMPLEMENTATION

The o-efsDFT approach can be implemented for any planewave or real-space based approach. The necessary steps to complete the self-consistent iteration to converge the density, energy, and forces can be summarized as follows (for a planewave based approach):

(a) Perform a deterministic DFT calculation in region \( D_f \) for each fragment \( f \) using a real-space grid spacing that equals the grid spacing of the full system.

(b) Using the fragment density as an initial guess, generate the corresponding Kohn–Sham potential.

(c) Generate a set of stochastic orbitals \( |\varphi_i \rangle \cdots |\varphi_i \rangle \) on a real-space grid and then transform to reciprocal space (it requires \( \approx N_g / 2 \) grid points in reciprocal space to store an \( N_g \) real-space orbital due to the isotropic kinetic energy cutoff).

(d) For all stochastic orbitals \( |\chi \rangle \), expand the action of the density matrix in terms of Chebyshev polynomials and tune the chemical potential to satisfy

\[ N_e(\mu) + \int C_f \, drf^\dagger(r) - \int C_f \, dr |\varphi_i(r)\rangle^2 \chi = N_e , \] (11)

where, as before, \( N_e(\mu) = \left\langle \chi | \varphi_i(\mu - f_H) | \chi \right\rangle \). \( \rho_f(r) \) is defined in Eq. (7), \( \rho_f(r) = \sum_{i=1}^{N_{\text{occ}}} \varphi_i^\dagger(r) \varphi_i (|x \rangle \langle x|) \), and \( N_e \) is the number of electrons in the system.
(e) Use the chosen \( \mu \) from the previous step to generate a set of stochastic projected orbitals, \( \langle \xi \rangle = \sqrt{\theta_\mu (\mu - \hat{h}_{KS})} |\psi\rangle \), using a Chebyshev or Newton interpolation polynomial to act with \( \sqrt{\theta_\mu (\mu - \hat{h}_{KS})} \) on \( |\psi\rangle \).

(f) Calculate the density \( \rho(r) \) using Eq. (7) and the corresponding energy with Eqs. (8) and (9). The other energy terms can be obtained directly from the density, similarly to the deterministic DFT.

(g) Use the density as input for the next iteration and go to step (d). Repeat until convergence is achieved.

Throughout the self-consistent iterations, the same random number seed was used to generate the stochastic orbitals. This is necessary to converge the self-consistent procedure to a tolerance similar to deterministic DFT. This also reduces the computational effort associated with projecting the stochastic orbitals onto the fragments so that the first and last terms in Eqs. (7)~(10) were calculated prior to the self-consistent loop.

V. RESULTS AND DISCUSSION

To assess the accuracy of the o-efsDFT method and its limitations, we performed \( I \) point calculations on a set of silicon crystals with varying super-cell size. Bulk silicon is rather challenging for linear-scaling DFT due to its particularly small con crystals with varying super-cell size. Bulk silicon is rather challenging for linear-scaling DFT due to its particularly small con crystals with varying super-cell size.

To assess the accuracy of the o-efsDFT method and its limitations, we performed \( I \) point calculations on a set of silicon crystals with varying super-cell size. Bulk silicon is rather challenging for linear-scaling DFT due to its particularly small con crystals with varying super-cell size.

We studied bulk silicon with 3 super cells of size \( 4 \times 4 \times 4 \) unit cells (\( N_{atom} = 512 \)), \( 6 \times 6 \times 6 \) unit cells (\( N_{atom} = 1728 \)), and \( 8 \times 8 \times 8 \) unit cells (\( N_{atom} = 4096 \)). Core fragments of Si \(_8\) were used to partition all systems, while three different buffer regions were adopted for forming the dressed fragments with 8, 64, and 216 Si atoms. \( N_\beta = 80 \) stochastic orbitals were used in each calculation, but for the non-overlapped Si \(_8\) fragments, we used \( N_\beta = 800 \) stochastic orbitals. It should be noted that calculations with \( N_\beta = 80 \) stochastic orbitals failed to converge without using overlapped fragments. The self-consistent iterations were converged using DIIS\(^{35-37}\) with a convergence criterion set to \( 10^{-8} \) hartree per electron.

Table I summarizes the results for the kinetic, nonlocal, local, and total energies per electron for all three system sizes studied and for all dressed fragments used. The external local potential energy per electron, \( E_{loc}/N_e \), depends linearly on the density and provides a reliable measure of the accuracy of the stochastic density. As observed in Table I, for Si\(_{32}\), \( E_{loc}/N_e \) with dressed fragments of sizes Si\(_{216}\) and Si\(_{326}\), the external potential energy per electron agrees well with the deterministic DFT result, where the differences between the two are well within the error bars of 5 independent o-efsDFT runs.

Although \( E_{loc}/N_e \) of the Si\(_{32}\) system with Si\(_{4096}\) as fragments agrees better with the deterministic result, this is simply a matter of luck and the relevant information regarding the converging with the fragment size is contained in the statistical error. Reductions in the standard deviation of \( E_{loc}/N_e \) with respect to the dressed fragment size were observed in all 3 system sizes.

This is also illustrated in Fig. 3(a), where we plot the density along the x-axis (at fixed y and z) for Si\(_{32}\) for all three dressed fragment sizes. The non-overlapping fragments required 10 times more stochastic orbitals (\( N_\beta = 800 \)) in order to reduce the noise in the density to similar levels as the overlapping Si\(_{4096}\) dressed fragments. For Si\(_{216}\) dressed fragments, the noise in the density dropped by a factor of \( \approx 2 \), suggesting that only \( N_\beta = 20 \) stochastic orbitals are needed to converge the density to the same level of noise as that of Si\(_{4096}\) dressed fragments.

The decrease in the noise with increasing buffer sizes (increase in \( \beta \) region) is not limited to the density itself or to local observables that depends directly on the density, like \( E_{loc}/N_e \). The kinetic energy per electron (\( E_{K}/N_e \)) and non-local pseudopotential energy per electron (\( E_{nl}/N_e \)) also show

| System | Fragment | \( N_\beta \) | \( E_{K}/N_e \) | \( E_{nl}/N_e \) | \( E_{loc}/N_e \) | \( E_{tot}/N_e \) |
|--------|----------|-------------|----------------|----------------|----------------|----------------|
| Si\(_{32}\) | Deterministic | 800 | 6.1274 | 8.6267 | 26.9954 |
| Si\(_{4}\) | 800 | 6.1269(29) | 8.6215(41) | 26.988(77) |
| Si\(_{216}\) | 80 | 6.1277(59) | 8.6223(61) | 26.9955(12) |
| Si\(_{4096}\) | 800 | 6.1284(15) | 8.6265(27) | 26.9959(5) |
| Si\(_{32}\) | 80 | 6.1278(25) | 8.6207(31) | 26.9953(11) |
| Si\(_{4}\) | 80 | 6.1286(7) | 8.6227(6) | 26.9964(2) |
| Si\(_{216}\) | 80 | 6.1286(5) | 8.6209(13) | 26.9944(12) |
| Si\(_{4096}\) | 80 | 6.1296(11) | 8.6243(11) | 26.9957(5) |
a significant reduction in their variance with increasing $B_x$, as can be seen in Table 1. Compared to the deterministic results for Si$_{32x}$, we find that $\alpha$-efsDFT provides energies per electron that are within $10^{-3}$ eV from the deterministic values and that the agreement improves with the size of the buffer regions. This is also illustrated in panels (b) and (c) of Fig. 3, where we plot the total energy per electron (relative to a reference system size for dressed fragment of Si$_{8x}$ [panel (b)] and Si$_{216x}$ [panel (c)] with $N_y = 80$. In panel (b), we also show the result with the Si$_{8x}$ dressed fragment (namely, no overlap between the fragments) with $N_y = 800$, where we also observe a large bias$^{27}$ and find that the energy is larger by $4$ meV per electron compared to the deterministic value.

Our results show that the time used in each self-consistent iteration scales as $O(N_y^{0.77})$. The net CPU time per stochastic orbital over all self-consistent iterations scales as $O(N_y^{0.77})$, which suggests that the number of self-consistent iterations does not significantly increase with increasing system size. We note that the practical scaling is somewhat better than linear ($O(N_y)$) mainly due to improved multithreading timings for larger grids. This however may depend on the computational architecture used and on the range of systems studied. Generally, we expect the approach to scale as $O(N_y)$ for large systems.

In addition to computing the individual contribution to the total energy per electron, $\alpha$-efsDFT also offers an efficient and accurate approach to compute the forces on each atom. Fig. 5(a) shows the average and standard deviation of nuclear forces from 5 independent $\alpha$-efsDFT runs for 50 representative atoms in Si$_{32x}$. The forces are computed for the equilibrium structure and thus should fluctuate about 0. Forces from $\alpha$-efsDFT with non-overlapped fragments and $N_y = 800$ stochastic orbitals fluctuate with comparable standard deviations as those from $\alpha$-efsDFT with Si$_{64x}$ dressed fragments and $N_y = 80$ stochastic orbitals, yet the computational effort is nearly 10 times smaller in the latter. Smaller force fluctuations were observed with growing buffer region, similar to the reduction in the statistical noise in the total energy per electron.

A summary of the statistical error in the forces as a function of the size of the dressed fragment and as a function of the system size are shown in panels (b) and (c) of Fig. 5, respectively. Within the central limit theorem, we assume that the standard deviation of force on the nuclei decays as $\sigma_0/\sqrt{N_y}$, where $\sigma_0$ is the standard deviation with only 1 stochastic orbital. In panel (b), we plot $\sigma_0$ as a function of the dressed fragment size, where a significant reduction is clearly observed. Panel (c) shows that $\sigma_0$ does not change significantly with the system size, suggesting that the number of stochastic orbitals does not increase with the system size for a given statistical error, suggesting that the scaling...
for the force follows that of the energy per electron shown in Fig. 4.

VI. OVERLAPPED VERSUS NON-OVERLAPPED FRAGMENTS

To further understand the o-efsDFT results presented in Sec. V, we examine the role played by the overlap of fragments. In the supplementary material, we derive an expression for the variance of the density in terms of the density itself and the density matrix

$$\text{Var}[\rho(r)] \leq 2V \max_r \left(2|\hat{\rho}^2 + \Delta \hat{\rho}^2| + 4\rho(r)\right) \times \int dr' \left[|\Delta \rho(r')|^2 \right],$$  \hspace{1cm} (12)

where $\Delta \hat{\rho} = \sum_i \hat{\rho}_i - \hat{\rho}$ and $\hat{\rho}$ and $\rho$ are defined in Eq. (6) and Eq. (1), respectively. It is clear that the noise in the density is not just correlated with the density, but also with the density matrix. Thus, fragmentation schemes to reduce the noise in $\rho(r)$ must provide a reasonable approximation for the density matrix across different fragments. This is not the case for the non-overlap fragments, where off-diagonal matrix elements of the full reference density matrix ($\sum_i \hat{\rho}_i$) for different fragments vanish, leading to a significant level of noise. In o-efsDFT with non-vanishing buffer zones, the reference density matrix is allowed to decay at least within distance $\Delta r$ before it is truncated to zero [see Fig. (2)]. This improves the reference density matrix and the resulting noise is significantly reduced as the size of the buffer zone increases.

In Fig. 6, we show the reference density matrix for three different fragmentation schemes. Panels (a) and (b) correspond to non-overlapped fragments with Si$_8$ and Si$_{64}$ fragments, respectively, and panel (c) shows the reference density matrix for Si$_8$ with a dressed fragment of Si$_{64}$. In the eigenfunction representation, the density matrix should equal the identity matrix. This is clearly not the case for panels (a) and (b), but is rather close to the unit matrix for panel (c). This is rather surprising since even with fragments that equal the size of the dressed fragment, the density matrix does not improve. Obviously, the discontinuity in the density matrix at the interfaces between fragments does not depend on the size of the fragment for the non-overlapped case, leading to significant noise in the density even for large fragments. However, it only takes a small fragment, Si$_8$, with a buffer zone of Si$_{64}$ to significantly reduce the noise.

VII. CONCLUSION

Motivated by the need to reduce significantly the noise level in linear-scaling sDFT, we have proposed a new fragmentation scheme that circumvents some of the limitations of our original embedded-fragmented stochastic DFT approach, mainly for covalently bonded systems. While previous attempts to develop a fragmentation scheme were based solely on providing a good reference system for the density alone, here, we showed that this is not a sufficient criterion. Based on the detailed analysis of the variance in the density, we developed a new overlapped fragmentation scheme, which provides an excellent reference for both the density matrix and the density. By dividing the system into non-overlapped core fragments and overlapped dressed fragments, we demonstrated that the existence of a buffer zone is crucial in the reduction of the stochastic noise in the energy per electron and nuclear forces. Moreover, we showed that the stochastic noise does not increase with respect to system size when using a fixed fragment size nor does the number of self-consistent iterations, leading to a practical scaling that is slightly better than the expected linear-scaling. While all applications in this study were limited to bulk silicon with periodic boundary conditions, we wish to iterate that the approach is also valid for open boundary conditions, where we expect it to work equally well.
SUPPLEMENTARY MATERIAL

See supplementary material for derivations of convergence and stochastic fluctuations of density in o-efsDFT.

ACKNOWLEDGMENTS

R.B. gratefully acknowledges support from the Israel Science Foundation (Grant No. 189/14). 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. Goesdecker, 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. R. Baer and M. Head-Gordon, Phys. Rev. Lett. 79, 3962 (1997).
12. A. H. R. Palser and D. E. Manolopoulos, Phys. Rev. B 58, 12704 (1998).
13. S. Goesdecker, Rev. Mod. Phys. 71, 1085 (1999).
14. W. Liang, C. Saravanan, Y. Shao, R. Baer, A. T. Bell, and M. Head-Gordon, J. Chem. Phys. 119, 4117 (2003).
15. W. Yang, Phys. Rev. Lett. 66, 1438 (1991).
16. P. Cortona, Phys. Rev. B 44, 8454 (1993).
17. T. Zhu, W. Pan, and W. Yang, Phys. Rev. B 53, 12713 (1996).
18. F. Shimoji, R. K. Kalia, A. Nakano, and P. Vashishta, Phys. Rev. B 77, 085103 (2008).
19. C. Jacob and J. Neugebauer, Wiley Interdiscip. Rev.: Comput. Mol. Sci. 4, 325 (2014).
20. T. A. Wesolowski, S. Shedge, and X. Zhou, Chem. Rev. 115, 5891 (2015).
21. P. Elliott, M. H. Cohen, A. Wasserman, and K. Burke, J. Chem. Theory Comput. 5, 827 (2009).
22. A. W. Götz, S. M. Beyhan, and L. Visscher, J. Chem. Theory Comput. 5, 3161 (2009).
23. C. Huang and E. A. Carter, J. Chem. Phys. 135, 194104 (2011).
24. J. D. Goodpaster, T. A. Barnes, and T. F. Miller, J. Chem. Phys. 134, 161408 (2011).
25. K. Yu and E. A. Carter, Proc. Natl. Acad. Sci. U. S. A. 114, E10861 (2017).
26. R. Baer, D. Neuhauser, and E. Rabani, Phys. Rev. Lett. 111, 106402 (2013).
27. M. Fabian, B. Shipiro, E. Rabani, D. Neuhauser, and R. Baer, e-print arXiv:1809.08307 (2018).
28. D. Neuhauser, R. Baer, and E. Rabani, J. Chem. Phys. 141, 041102 (2014).
29. E. Arnon, E. Rabani, D. Neuhauser, and R. Baer, J. Chem. Phys. 146, 224111 (2017).
30. R. Kosloff, J. Phys. Chem. 92, 2087 (1988).
31. R. Kosloff, Annu. Rev. Phys. Chem. 45, 145 (1994).
32. C. K. Skylaris and P. D. Haynes, J. Chem. Phys. 127, 164712 (2007).
33. N. Troullier and J. L. Martins, Phys. Rev. B 43, 1993 (1991).
34. L. Kleinman and D. M. Bylander, Phys. Rev. Lett. 48, 1425 (1982).
35. R. D. King-Smith, M. C. Payne, and J. S. Lin, Phys. Rev. B 44, 13063 (1991).
36. P. Pulay, Chem. Phys. Lett. 73, 393 (1980).
37. P. Pulay, J. Comput. Chem. 3, 556 (1982).
38. G. Kresse and J. Furthmuller, Phys. Rev. B 54, 11169 (1996).