Tempering stochastic density functional theory

Minh Nguyen, Wenfei Li, Yangtao Li
Department of Chemistry and Biochemistry, University of California at Los Angeles, Los Angeles, California 90095, USA

Roi Baer
Fritz Haber Center of Molecular Dynamics and Institute of Chemistry, The Hebrew University of Jerusalem, Jerusalem, 91904 Israel

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

Daniel Neuhauser
Department of Chemistry and Biochemistry, University of California at Los Angeles, and California Nanoscience Institute, Los Angeles, California 90095, USA
(Dated:)

We introduce a tempering approach with stochastic density functional theory (sDFT), labeled t-sDFT, which reduces the statistical errors in the estimates of observable expectation values. This is achieved by rewriting the electronic density as a sum of a "warm" component complemented by "colder" correction(s). Since the warm component is larger in magnitude but faster to evaluate, we use many more stochastic orbitals for its evaluation than for the smaller-sized colder correction(s). This results in a significant reduction of the statistical fluctuations and the bias compared to sDFT for the same computational effort. We demonstrate the method's performance on large hydrogen-passivated silicon nanocrystals (NCs), finding a reduction in the systematic error in the energy by more than an order of magnitude, while the systematic errors in the forces are also quenched. Similarly, the statistical fluctuations are reduced by factors of ≈4-5 for the total energy and ≈1.5-2 for the forces on the atoms. Since the embedding in t-sDFT is fully stochastic, it is possible to combine t-sDFT with other variants of sDFT such as energy-widnw sDFT and embedded-fragmented sDFT.

I. INTRODUCTION

Kohn-Sham density functional theory (KS-DFT) is widely used for calculating properties of molecular and extended systems [1]. In particular, the method is useful for determining the structure based on the estimates it provides for the forces on the corresponding nuclei [2–4]. However, applying KS-DFT for systems with hundreds or thousands of atoms is challenging due to a high scaling (potentially quadratic but eventually cubic for large systems). Lower scaling implementations of the theory have been developed, but their use is often limited to low-dimensional structures [5, 6] or systems with strictly localized electrons [7, 8].

In previous work we introduced stochastic density functional theory (sDFT) [9] which avoids the costly diagonalization step in KS-DFT at the cost of introducing statistical uncertainties in the density and other observables. The statistical errors can be reduced by using an embedded-fragmented (ef-sDFT) technique [10–12] which is based on dividing the system into fixed-size fragments and expressing the total electron density, $n(\boldsymbol{r})$, as the sum of fragment densities plus a correction term which is evaluated stochastically. This technique reduces the statistical fluctuations in the estimates of the atomic forces and the energies, [10, 11] and the magnitude of this reduction is controlled by varying the size of the fragments and the number of stochastic realizations. An additional approach for mitigating the statistical errors is the energy-window sDFT (ew-sDFT) scheme [13] and its combination with the embedded-fragmented technique [10, 14] to further reduce the statistical errors.

Here we propose a tempering method, referred to as t-sDFT, as a complementary technique for reducing the statistical noise, where the density for the desired temperature is calculated using a higher-temperature reference density and smaller corrections. This idea has been implemented before within the energy renormalization group in the context of telescopically expanding the Hamiltonian matrix in a series [15]. In Sec. II, we describe the t-sDFT method and in Sec. III we benchmark and analyze its efficacy using large hydrogenated silicon clusters.

Sec. IV we discuss the conclusions and summarize.

II. METHODOLOGY

A. Stochastic Density Functional Theory

Our starting point is the following expression for the electron density $n(\boldsymbol{r})$ (assuming a spin-unpolarized system) [9]:

$$n(\boldsymbol{r}) = 2 \times \mathrm{Tr}\left[\sqrt{\hat{\rho}_\beta}\,|\boldsymbol{r}\rangle\langle\boldsymbol{r}|\,\sqrt{\hat{\rho}_\beta}\right], \tag{1}$$

where $\boldsymbol{r}$ is a point on a 3D grid that spans the space containing the electron density of the system and has a volume element dV. The operator

$$\hat{\rho}_\beta = f_{\beta\,\mu}\left(\hat{h}\right) \tag{2}$$

is a smooth low band-pass Fermi-Dirac (FD) filter, $f_{\beta\,\mu}(\varepsilon) = \left(1+e^{\beta(\varepsilon-\mu)}\right)^{-1}$, which blocks high energies ($\varepsilon > \mu+\beta$). Here, the chemical potential μ must be adjusted such that the integrated density equals the number of electrons, N_e,

$$\sum_{\boldsymbol{r}} n(\boldsymbol{r})\,dV = N_e, \tag{3}$$

while β is the inverse temperature. In the low-temperature limit ($\beta\varepsilon_{\mathrm{gap}} \gg 1$, where $\varepsilon_{\mathrm{gap}}$ is the fundamental KS gap) $n(\boldsymbol{r})$ converges to the ground state KS-density. Note that the filter $\hat{\rho}_\beta$ depends on the chemical potential but to avoid a plethora of indices we do not explicitly note this dependence below.

In Eq. (2), the Kohn-Sham Hamiltonian is

$$\hat{h} = \hat{t} + \hat{v}[n](\boldsymbol{r}) \tag{4}$$

where $\hat{t}$ is the kinetic energy operator on the grid and $\hat{v}[n](\boldsymbol{r})$ is the density-dependent KS potential, composed of the electron-nuclei, Hartree, and exchange-correlation components. Eqs. (1)-(4) must be solved simultaneously to yield the self-consistent electron density.

In sDFT, we introduce a stochastic resolution of the identity [16] which transforms the trace in Eq. (1) to an expectation value [9]

$$n(\boldsymbol{r}) = 2\left\langle\left|\left\langle\boldsymbol{r}\left|\sqrt{\hat{\rho}_\beta}\right|\chi\right\rangle\right|^2\right\rangle_\chi, \tag{5}$$

where $|\chi\rangle$ is a stochastic orbital taking the randomly signed values $\langle\boldsymbol{r}\,|\chi\rangle = \pm(dV)^{-1/2}$.

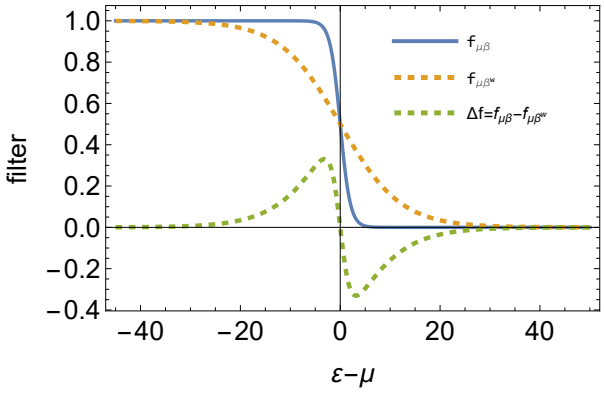

Figure 1. The desired low temperature filter, $f_{\beta\,\mu}(\varepsilon)$, the high temperature filter $f_{\beta^w\,\mu}(\varepsilon)$ and the correction for $\beta = 6\beta^w$.

To apply the filter, $\sqrt{\hat{\rho}_\beta}$, we use a Chebyshev expansion of length K [17]

$$\sqrt{\hat{\rho}_\beta}\,|\chi\rangle = \sum_{k=0}^{K} c_k(\beta,\mu)\left|\zeta^{(k)}\right\rangle, \tag{6}$$

where $\left|\zeta^{(k)}\right\rangle$ are defined by the Chebyshev polynomial recursion relations: $\left|\zeta^{(0)}\right\rangle = |\chi\rangle$, $\left|\zeta^{(1)}\right\rangle = \hat{h}_s\,|\chi\rangle$ and $\left|\zeta^{(k+1)}\right\rangle = 2\hat{h}_s\left|\zeta^{(k)}\right\rangle - \left|\zeta^{(k-1)}\right\rangle$. Here, $\hat{h}_s = \left(\hat{h}-\bar{\varepsilon}\right)/\Delta\varepsilon$ is a normalized KS Hamiltonian where $\bar{\varepsilon}$ and $\Delta\varepsilon$ are chosen such that the spectrum of $\hat{h}_s$ lies within the interval $[-1,1]$; $c_k(\beta,\mu)$ are then the Chebyshev expansion coefficient of the filter $\sqrt{f_\beta(\varepsilon)}$[17]. The expansion length K terminates the series when $|c_{k>K}|$ is smaller than a predetermined cutoff value.

In practice, the expected value appearing in Eq. (5) is evaluated approximately by taking a finite sample of N_s stochastic orbitals:

$$n(\boldsymbol{r}) \approx \frac{2}{N_s}\sum_{s=1}^{N_s}\left|\left\langle\boldsymbol{r}\left|\sqrt{\hat{\rho}_\beta}\right|\chi_s\right\rangle\right|^2. \tag{7}$$

Furthermore, to ensure $N_e = \sum_{\boldsymbol{r}} n(\boldsymbol{r})\,dV$ we tune the chemical potential μ to satisfy the relation

$$N_e = 2\sum_{k=0}^{K} b_k(\beta,\mu)\,M_k, \tag{8}$$

where

$$M_k = \frac{1}{N_s}\sum_{s=1}^{N_s}\left\langle\chi_s\left|\zeta^{(k)}\right.\right\rangle \tag{9}$$

are the stochastic estimates of the Chebyshev moments [18] and $b_k(\beta,\mu)$ are the Chebyshev expansion coefficients corresponding to the function $f_{\beta\,\mu}(\varepsilon)$, to differentiate from $c_k(\beta,\mu)$ which are the expansions coefficients of $\sqrt{f_{\beta\,\mu}(\hat{h})}$.

As a result of using stochastic orbitals, the sDFT density and associated observables have two contributions to the errors. First, the usual stochastic fluctuations that scale as $\mathrm{O}(N_s^{-\frac{1}{2}})$. But in addition, there is also a systematic deviation which scales as $\mathrm{O}(N_s^{-1})$ that appears due to the nonlinear SCF procedure (essentially, because the filtering operator applied on each orbital depends on the density, which itself depends on the set of filtered orbitals).

Increasing the number of sampling, N_s, will decrease both type of errors, at the cost of additional work.

We quantify the numerical work for each system by the total number of Hamiltonian operations needed (i.e., action by the Hamiltonian on a function), which is approximately

$$W \simeq KN_s.$$

In practice the work needs to be multiplied by an overall factor of about 1.7 due to the need to determine μ based on Eq. (8) but since this factor is common to all our methods here we do not include it.

B. Tempering Stochastic Density Functional Theory

We now describe the tempering, designed to reduce the statistical errors in sDFT without increasing the computational effort. Consider, in addition to the desired filter $\hat{\rho}_\beta$ (see Eq. (2)) a higher temperature filter $\hat{\rho}_{\beta^w}$ ($\beta^w < \beta$) and the correction term:

$$\Delta\hat{\rho} = \hat{\rho}_\beta - \hat{\rho}_{\beta^w}, \tag{10}$$

shown in Fig. (1) for a typical case of $\beta/\beta^w = 6$. Note that (a) all values of $\Delta\hat{\rho}$ are much smaller than unity and (b) the high temperature filter $\hat{\rho}_{\beta^w}$ is smoother than the low temperature one, so its Chebyshev expansion is shorter (the Chebyshev expansion length of $\hat{\rho}_\beta$ is proportional to β [9]).

The electron density in Eq. (5) can therefore be written as $n(\boldsymbol{r}) = n_{\beta^w}(\boldsymbol{r}) + \Delta n(\boldsymbol{r})$, where the two terms are evaluated separately using distinct independent sets of stochastic orbitals, χ_s^w, $s = 1,\ldots,N_s^w$ for the warmer density

$$n_{\beta^w}(\boldsymbol{r}) = \frac{2}{N_s^w}\sum_{s=1}^{N_s^w}\left|\left\langle \boldsymbol{r}\left|\sqrt{\hat{\rho}_{\beta^w}}\right|\chi_s^w\right\rangle\right|^2 \tag{11}$$

and χ_s^Δ, $s = 1,\ldots,N_s^\Delta$ for the correction term:

$$\Delta n(\boldsymbol{r}) = \frac{2}{N_s^\Delta}\sum_{s=1}^{N_s^\Delta}\left(\left|\left\langle \boldsymbol{r}\left|\sqrt{\hat{\rho}_\beta}\right|\chi_s^\Delta\right\rangle\right|^2 - \left|\left\langle \boldsymbol{r}\left|\sqrt{\hat{\rho}_{\beta^w}}\right|\chi_s^\Delta\right\rangle\right|^2\right) \tag{12}$$

The chemical potential in t-sDFT is adjusted to fulfill Eq. (3) which here can be rewritten as:

$$\begin{aligned} N_e = {} & 2\sum_{k=0}^{K^w} b_k(\beta^w,\mu)\, M_k^w \\ & + 2\sum_{k=0}^{K}\left[b_k(\beta,\mu) - b_k(\beta^w,\mu)\right] M_k^\Delta. \end{aligned} \tag{13}$$

Here, K^w and K are respectively, the Chebyshev expansion lengths for the warm reference density and the correction terms, with $\frac{K^w}{K} \sim \frac{\beta^w}{\beta} \ll 1$ (note that since the correction term involves the original low-temperature density, the number of Chebyshev terms it requires, K, is the same as in the original sDFT). The corresponding Chebyshev moments, defined in analogy to Eq. (9), are:

$$\begin{aligned} M_k^w &= \frac{1}{N_s^w}\sum_{s=1}^{N_s^w}\left\langle \chi_s^w\left|\zeta_w^{(k)}\right.\right\rangle, \\ M_k^\Delta &= \frac{1}{N_s^\Delta}\sum_{s=1}^{N_s^\Delta}\left\langle \chi_s^\Delta\left|\zeta_\Delta^{(k)}\right.\right\rangle \end{aligned} \tag{14}$$

As demonstrated in Fig. 1, the correction density $\Delta n(\boldsymbol{r})$ is much smaller than the warm density $n_{\beta^w}(\boldsymbol{r})$ which is similar in overall magnitude to the total density. This gives the key for the efficiency of the approach compared with the original sDFT calculation. Specifically, compared to an sDFT calculation with polynomial expansion length K and N_s stochastic orbitals, and aiming for the same overall work as in sDFT, we get that:

- The computational work required to calculate the warm density is $W^w = K^w N_s^w$. Since the Chebyshev expansion lengths are proportional to β, the warmer temperature density $n_{\beta^w}(\boldsymbol{r})$ requires a much *shorter* Chebyshev expansion length than the original sDFT density ($K^w = \frac{\beta^w}{\beta}K$), so many more stochastic orbitals can be used to evaluate it for the same overall computational cost (i.e., $N_s^\Delta \gg N_s$).
- The required computational work for correction term $\Delta n(\boldsymbol{r})$ is $W^\Delta = (K + K_w)N_s^\Delta$ as both terms in the RHS of Eq. (14) use the same $\zeta_\Delta^{(k)}$ orbitals. Since the numerical magnitude of the correction term is much smaller

System	Band-gap	β	Correction filter			Warm filter			W^{tot}
	(eV)	(eV^{-1})	K	N_s^Δ	$W^\Delta = KN_s^\Delta$	K^w	N_s^w	$W^w = K^w N_s^w$	
$Si_{35}H_{36}$	3.4	1.83	2000	6	12,000	$K \times \beta^w/\beta$	$24 \times \beta/\beta^w$	48,000	60,000
$Si_{87}H_{76}$	2.5	2.94	3200	6	19,200	$K \times \beta^w/\beta$	$24 \times \beta/\beta^w$	76,800	96,000
$Si_{353}H_{196}$	1.6	4.60	5000	6	30,000	$K \times \beta^w/\beta$	$24 \times \beta/\beta^w$	120,000	150,000

Table I. The Chebyshev expansion lengths K, K^w, and the number of stochastic orbitals N_s^Δ, N_s^w used in our simulations. We also show the required numerical work W, defined as the number of Hamiltonian operations. Note that for each system, we increase the number of high-temperature orbitals (N_s^w) with temperature (i.e., with increasing β/β^w) such that the total work W^{tot} is independent of β/β^w.

than the overall density, it is sufficient to use fewer stochastic orbitals ($N_s^\Delta \ll N_s$) to achieve a similar statistical error. Also note that since $K_w \ll K$ (as $\beta_w \ll \beta$), we can safely approximate $W^\Delta \simeq KN_s^\Delta$.

Overall, the partitioning of the filter into a larger component at a higher temperature with a shorter Chebyshev expansion, and a smaller correction term, offers an additional knob to control the statistical error by using $N_s^w \gg N_s$ without increasing the overall computational effort.

III. RESULTS

We studied three hydrogen-terminated silicon nanocrystals of different size, $Si_{35}H_{36}$, $Si_{87}H_{76}$ and $Si_{353}H_{196}$. An LDA functional [19] was applied with norm-conserving pseudopotentials [20] using the Kleinman-Bylander form [21], and we used the Martyna-Tuckerman reciprocal-space method for treating long-range interactions [22]. The grid spacing was $0.55a_0$, and the energy cutoff was 15 Hartree for all systems. To gather sufficient statistics, $N_{\rm ind} = 10$ independent runs with different stochastic numbers were used for each of the calculations below.

For each system, we performed calculations for several β/β^w ratios. As these systems are semiconductors we are simply interested in the limit where the Fermi-Dirac distribution is effectively a step function. We therefore replace the Fermi-Dirac distribution by the complementary error function, $f_{\beta\,\mu}(\hat{h}) = \frac{1}{2}\mathrm{erfc}(\beta(\hat{h}-\mu))$,which looks similar to the Fermi-Dirac distribution but does not require a very small β to be effectively a step function.

The numerical parameters for the runs are summarized in Table I. There are several points to note.

First, since the three systems have progressively smaller band gap E_g, the larger system requires the largest value of β and a correspondingly larger Chebyshev expansion length K.

Furthermore, we compare t-sDFT to sDFT, where in the latter we use (for all systems) $N_s = 30$ orbitals. For each system, the value of β (associated with the low temperature) was the same in the sDFT and t-sDFT, so the number of polynomials terms (K), was the same for sDFT and for the correction ($\Delta n(\boldsymbol{r})$) in t-sDFT.

Finally, note that the warm temperature calculations which require most of the numerical work have their expansion length K^w and their number of stochastic orbitals N_s^w chosen in a manner that renders the total work W^{tot} independent of the ratio β/β^w. This allows us to compare the efficacy of the method in terms of the reduction of fluctuations as a function of β/β^w.

Energies

Fig. 2 shows, for the three different systems, the averaged total energies per particle based on the $N_{\rm ind} = 10$ runs and the associated error bars for different β^w values. For simplicity, the results are depicted as a function of β/β^w. In addition, we compare to the deterministic values.

Consider first the starting point for each figure, $\beta/\beta^w = 1$, which is simply sDFT (i.e., with no correction terms). For that case we use, as mentioned, only $N_s = 30$ stochastic orbitals, and since this is a very small number, the sDFT runs show a significant systematic deviation, i.e., the averaged energy-per-particle is several standard deviations away from the deterministic value.

The origin of the systematic deviation is in the SCF iterations, i.e., when the density is prepared from a finite number of stochastic orbitals, a self-consistent solution would give a systematic deviation which scales as N_s^{-1} (for a fuller discussion of these deviations see Ref. [12]). Interestingly (see also the SI), for a fixed N_s the systematic deviation decreases by a factor of 2 when the system size increases by a factor of 10, while the stochastic error decreases faster with system size, scaling as $N_e^{-1/2}$ due to self-averaging.

Turning to t-sDFT (i.e., $\beta^w < \beta$) we see that both

the systematic deviation and the stochastic error decrease as β/β^w increases.

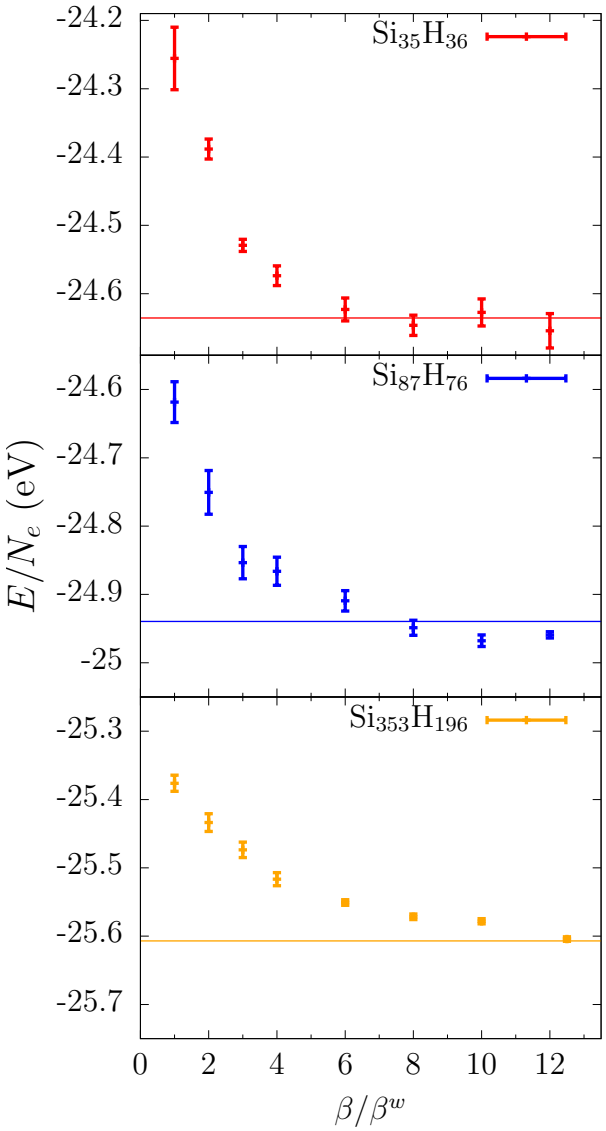

Figure 2. The energy per electron (in eV) as a function of β/β^w for three cluster sizes, based on $N_{\rm ind} = 10$ independent runs. Also included is the deterministic value in each system (horizontal line). The numerical work W, i.e., the number of Hamiltonian operations, is constructed to be independent of β/β^w (see Table I for details). The first point in each graph, $\beta/\beta^w = 1$, corresponds to sDFT (no tempering). Since the number of orbitals used is very small ($N_s = 30$ for sDFT), these sDFT results show marked systematic deviation (i.e., deviation of the average energy from the deterministic value) and fluctuation errors. Both type of statistical errors decrease markedly in t-sDFT, especially when $\beta/\beta^w \sim 7-10$, due to the much larger number of stochastic orbitals used in the main (warm) density part.

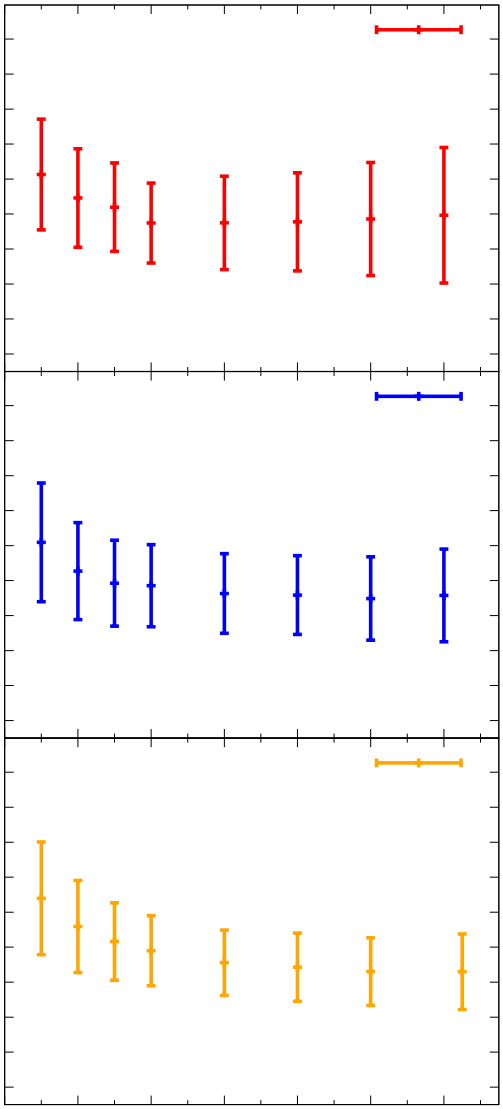

Figure 3. Analogous to Fig. 2, but shows δF,the error in the averaged force relative to the deterministic value, normalized over all silicon atoms and over the $N_{\rm ind} = 10$ runs, and the associated standard deviation σ_F (Eqs. (15,16)). In sDFT δF is significantly larger than σ_F, indicating some amount of systematic deviation. In t-sDFT, around $\beta/\beta^w \sim 7-10$, both the stochastic and especially the systematic errors decrease, i.e., σ_F decreases and $\delta F \sim \sigma_F$.

As evident from Fig. 2 (and verified by a second-order polynomial fit of the error in the energy as a function of β/β^w in the SI) once $\beta/\beta^w \sim 7-12$ there is essentially no bias while the stochastic fluctuations decrease by a factor of around 4-5.

The reduction in the statistical error and the bias as β/β^w increases for a fixed W results from the fact the significant contribution to the density comes from the warmer temperature, for which we us a much larger number of stochastic orbitals, thereby diminishing both the systematic and statistical errors relative to sDFT. The correction term requires a much smaller number of stochastic orbitals despite the need to use longer Chebyshev series.

Finally note that with increasing system size, β/β^w values which are optimal (i.e., lead to small statistical fluctuations in the energy per electron) shift to larger ratios. This is a result of quantum confinement, leading to a decrease in the KS gap with increasing systems size.

Forces and Density

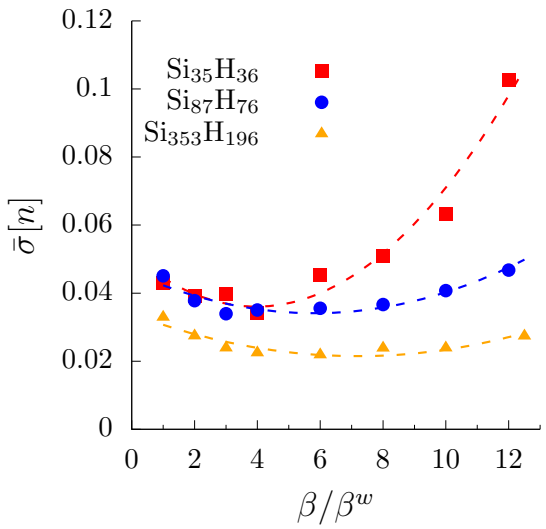

Figure 4. The normalized integral of the standard deviation of the density per electron, together with a parabolic fit. For the smaller system the density deviation decreases between $\beta/\beta^w = 2-4$, and for the two larger systems the stochastic errors decrease around a larger range $\beta/\beta^w = 2-10$, by up to 30%-40%.

We next show how t-sDFT reduces the errors in the atomic forces compared to sDFT (we only analyzed the forces on the silicon atoms). Fig. 3 is similar to the energy plot in Fig. 2, but here we plot the normalized deviation of the averaged stochastic forces from the deterministic forces, δF:

$$(\delta F)^2 \equiv \frac{1}{N_{\rm Si}} \sum_{i=1}^{N_{\rm Si}} |\bar{\boldsymbol{F}}^i - \boldsymbol{F}^{d,i}|^2, \qquad (15)$$

where a bar indicates averaging over the $N_{\rm ind} = 10$ runs and "d" stands for deterministic; i is an index over the silicon atoms. The error bars in Figure 3, σ_F, indicate the standard deviation of the normalized averaged force of the silicon atoms, i.e.,

$$\sigma_F^2 = \frac{1}{(N_{\rm ind}-1)N_{\rm ind}N_{\rm Si}} \sum_{j=1}^{N_{\rm ind}} \sum_{i=1}^{N_{\rm Si}} |\boldsymbol{F}^{i,j} - \bar{\boldsymbol{F}}^i|^2, \quad (16)$$

where $\boldsymbol{F}^{i,j}$ is the force over atom i in the j'th independent run.

Note that the magnitude of the errors in the forces is large but this could be reduced by increasing the number of independent runs or stochastic orbitals. However, since the purpose of the study is to uncover the behavior with β^w, we use a small number of stochastic orbitals to reduce the computational effort and thus, apply the approach for many values of β/β^w and for different system sizes.

As we previously discussed, overall the sDFT forces are similar in the three systems, since the local environment is similar and therefore the errors are primarily a function of the number of stochastic orbitals, N_s. The reduction in the errors in the forces is appreciable but less significant than for the energy. Using again a 2nd order fit (see the SI) we get that the reduction in error in the forces is about 30% for the smallest system and becomes 50% for the largest cluster.

To compare the deviation in the density using t-sDFT to sDFT, we use the integral of the standard deviation of the averaged density per electron, defined as $\bar{\sigma}[n] \equiv (N_{\rm ind}-1)^{-1/2} N_e^{-1} \int \sigma(n(\boldsymbol{r}))d\boldsymbol{r}$, where $\sigma(n(\boldsymbol{r}))$ is the standard deviation in the local density. Fig. 4 shows that tempering again reduces the stochastic error as long as β/β^w is again 7-10. The reduction in the deviation of the density is similar in magnitude to that of the total atomic forces, up to 30%-40%, and is much less dramatic than the error reduction in the total energies.

Finally, note that Fig. 4 shows that when the value of β/β^w is very large, the density fluctuations start rising with the β/β^w ratio; for large ratios, the warm density deviates significantly from the low-temperature density, so the difference between the two density is significant which causes large statistical fluctuations.

IV. CONCLUSIONS

We presented here a method for stochastic density functional theory that reduces the statistical error in the total energy. Our scheme (t-sDFT) relies on decomposing the density into a large high-temperature term with one or more correction densities. The new method expands the density in terms of the inverse temperature parameter, β, to take advantage of the fact that with lower β (i.e., a higher effective temperature) fewer Chebyshev polynomials are needed, thus enabling the use of more stochastic orbitals without increasing the computational cost.

Our method reduces the standard deviation in the total energy by a factor of around 4-5, which corresponds to reducing the total number of required stochastic orbitals by more than an order of magnitude. This is only for the total energies, while the error in the forces and density is reduced by a smaller amount only 30%-50% and 30%-40% respectively. Interestingly, this is the opposite behavior relative to energy-window sDFT where the error in the forces is improved significantly but not the error in the total energies.

A natural extension of this work is the implementation of multiple-β tempering with more than two values of β, as done earlier for deterministic renormalization-group studies; the formalism is presented in Appendix A. In addition, in this work we have not implemented embedded fragments, an approach that independently reduces the standard deviation in the energy and forces. In future work, the two methods would be combined, to hopefully reduce the stochastic error even further. Further work would also explore how to optimize the choice of β values and number of stochastic orbitals used to reduce the stochastic deviations.

ACKNOWLEDGMENTS

This paper was supported by the Center for Computational Study of Excited State Phenomena in Energy Materials (C2SEPEM), which is funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division via Contract No. DE-AC02- 05CH11231, as part of the Computational Materials Sciences Program. In addition, RB gratefully acknowledges the support from the US-Israel Binational Science Foundation (BSF) under Grant No. 2018368. Computational resources were supplied through the XSEDE allocation TG-CHE170058.

DATA AVAILABLITY STATEMENT

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

APPENDIX A: MULTIPLE β

The general expansion of the filter $\hat{\rho}_\beta$ for L temperatures, ordered so $\beta \equiv \beta_1 > \beta_2 > ... > \beta_L$, is

$$\hat{\rho}_\beta = \hat{\rho}_{\beta_L} - \sum_{\ell=1}^{L-1} \Delta\hat{\rho}_\ell, \tag{17}$$

where

$$\Delta\hat{\rho}_\ell = \hat{\rho}_{\beta_{\ell+1}} - \hat{\rho}_{\beta_\ell}. \tag{18}$$

This expansion leads to an expression for the density similar to Eqs. 11-12. The case we studied in the paper is simply $L = 2$, with $\beta_1 \equiv \beta$ and $\beta_2 \equiv \beta^w$.

[1] R. O. Jones. Density functional theory: Its origins, rise to prominence, and future. *Reviews of Modern Physics*, 87(3):897–923, August 2015.

[2] Brian Kolb and T Thonhauser. Molecular Biology at the Quantum Level: Can Modern Density Functional Theory Forge the Path? *Nano LIFE*, 2(02), 2012.

[3] Daniele Selli, Gianluca Fazio, and Cristiana Di Valentin. Modelling realistic TiO_2 nanospheres: A benchmark study of SCC-DFTB against hybrid DFT. *The Journal of Chemical Physics*, 147(16):164701, 2017. Publisher: AIP Publishing LLC.

[4] Christoph Freysoldt, Blazej Grabowski, Tilmann Hickel, Jörg Neugebauer, Georg Kresse, Anderson Janotti, and Chris G. Van de Walle. First-principles calculations for point defects in solids. *Rev. Mod. Phys.*, 86(1):253–305, March 2014. Publisher: American Physical Society.

[5] Roi Baer and Martin Head-Gordon. Sparsity of the Density Matrix in Kohn-Sham Density Functional Theory and an Assessment of Linear System-Size Scaling Methods. *Phys. Rev. Lett.*, 79(20):3962–3965, November 1997.

[6] G. E. Scuseria. Linear scaling density functional calculations with Gaussian orbitals. *J. Phys. Chem. A*, 103(25):4782–4790, 1999.

[7] T. Ozaki. Efficient low-order scaling method for large-scale electronic structure calculations with localized basis functions. *Phys. Rev. B*, 82(7):075131,

2010.

[8] Carlos Romero-Muniz, Ayako Nakata, Pablo Pou, David R Bowler, Tsuyoshi Miyazaki, and Rubén Pérez. High-accuracy large-scale DFT calculations using localized orbitals in complex electronic systems: The case of graphene–metal interfaces. *Journal of Physics: Condensed Matter*, 30(50):505901, 2018. Publisher: IOP Publishing.

[9] Roi Baer, Daniel Neuhauser, and Eran Rabani. Self-Averaging Stochastic Kohn-Sham Density-Functional Theory. *Phys. Rev. Lett.*, 111(10):106402, September 2013.

[10] Daniel Neuhauser, Roi Baer, and Eran Rabani. Communication: Embedded fragment stochastic density functional theory. *J. Chem. Phys.*, 141(4):041102, 2014.

[11] Ming Chen, Roi Baer, Daniel Neuhauser, and Eran Rabani. Overlapped embedded fragment stochastic density functional theory for covalently-bonded materials. *J. Chem. Phys.*, 150(3):034106, January 2019.

[12] Marcel D. Fabian, Ben Shpiro, Eran Rabani, Daniel Neuhauser, and Roi Baer. Stochastic density functional theory. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 10.1002/wcms.1412(0):e1412, 2019.

[13] Ming Chen, Roi Baer, Daniel Neuhauser, and Eran Rabani. Energy window stochastic density functional theory. *J. Chem. Phys.*, 151(11):114116, September 2019.

[14] Ming Chen, Roi Baer, Daniel Neuhauser, and Eran Rabani. Stochastic density functional theory: Real- and energy-space fragmentation for noise reduction. *J. Chem. Phys.*, 154(20):204108, May 2021.

[15] Roi Baer and Martin Head-Gordon. Energy renormalization-group method for electronic structure of large systems. *Physical Review B-Condensed Matter*, 58(23):15296–15299, 1998.

[16] Michael F Hutchinson. A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines. *Commun Stat Simul Comput.*, 19(2):433–450, 1990.

[17] R. Kosloff. Time-Dependent Quantum-Mechanical Methods for Molecular- Dynamics. *J. Phys. Chem.*, 92(8):2087–2100, 1988.

[18] Otto F Sankey, David A Drabold, and Andrew Gibson. Projected random vectors and the recursion method in the electronic-structure problem. *Phys. Rev. B*, 50(3):1376, 1994.

[19] 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.

[20] N. Troullier and J. L. Martins. Efficient Pseudopotentials for Plane-Wave Calculations. *Phys. Rev. B*, 43(3):1993–2006, 1991.

[21] Leonard Kleinman and D. M. Bylander. Efficacious Form for Model Pseudopotentials. *Phys. Rev. Lett.*, 48(20):1425–1428, May 1982.

[22] 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.