New-Generation Nonlocal Orbital-free DFT Reaches Beyond the Nanoscale

Xuecheng Shao,1,2,∗ Wenhui Mi,1,2,† and Michele Pavanello1,2,‡

1Department of Physics, Rutgers University, Newark, NJ 07102, USA
2Department of Chemistry, Rutgers University, Newark, NJ 07102, USA
(Dated: October 16, 2020)

When nonlocal kinetic energy functionals are employed in orbital-free density functional theory (OF-DFT), the simulations approach the accuracy of conventional DFT provided that the kernel of the nonlocal functional is density dependent. Unfortunately, this leads to slow OF-DFT algorithms with limited applicability. In this work, we solve this issue by developing a density optimization method that is only slightly more expensive than a local or semilocal OF-DFT optimization. This signifies a paradigm shift for OF-DFT which can now be employed in large-scale simulations delivering an accuracy much beyond parametrized force fields and on par with Kohn-Sham DFT. We showcase the method computing with merely a single CPU the electronic structure of silicon nanoparticles up to 15 nm in size (or 100k atoms) and analyzing the polarization induced by the surface truncations. These constitute the largest ab initio simulations for silicon based materials to date.

In recent years, there have been many important advances bringing OF-DFT to tackle semiconductors[24, 25] and even clusters[26, 27] at a few hundred meV/atom accuracy compared to KS-DFT.

The main advance has been the development of new-generation, nonlocal kinetic energy density functionals (KEDFs) with density-dependent kernels[25–27]. These rely on splines to evaluate a kernel at each real-space point[26]. The spline procedure coupled with the need to evaluate several tens or hundreds of convolution integrals for each energy and potential evaluation makes these KEDFs accurate but computationally expensive[25, 28–30]. The end result is that new-generation, accurate KEDFs (such as HC[25], LMGP[26], and others[27]) have never been applied to large-scale simulations. And with that, the promise to realize the futuristic vision discussed above has so far been unfulfilled.

In this work, we change this state of affairs by developing a completely new density optimization procedure for OF-DFT. With our development, even the most expensive nonlocal KEDF with density-dependent kernel can be employed in OF-DFT simulations at virtually the same computational cost as semilocal KEDFs. This is a paradigm shift for OF-DFT as new-generation KEDFs can now be employed in such difficult tasks as computing the electronic structure of semiconducting nanoparticles with hundreds of thousands of atoms and in general in any large-scale simulation.

Since early on, it was thought that the OF-DFT problem “can be solved iteratively to self-consistency by any Kohn-Sham computer program: just select the lowest eigenvalue. The solution is very simple and quick, for there is only one orbital.”[31] Such a Self-Consistent Field (SCF) technique has been attempted by several before with poor outcomes[32–34]. Naturally, the technique used nowadays is the Direct Energy Minimization (DEM) of the total energy functional[6, 28, 29, 35]. With DEM, the high cost of new-generation KEDFs creeps in due to the need to evaluate the KEDF potential many times during an optimization.

Since the mid-sixties, scientists have hoped that one day first-principles device-level and large-scale materials engineering would be feasible and widely available such that computational models could in part or totally replace experiments[1]. This futuristic vision can only be accomplished if accurate ab initio quantum mechanical electronic structure methods are computationally cheap and can model system sizes beyond the nanoscale. As opposed to the widely employed Kohn-Sham density functional theory (KS-DFT)[2], orbital-free DFT (OF-DFT)[3, 4] is easily formulated in linear scaling algorithms and realized in fast computer codes that have been applied to systems containing all the way up to hundreds of millions of atoms[5–7].

Thus, the question is: can we begin implementing the wanted futuristic vision? Unfortunately, the answer is a resounding “no”. The main reason for the negative outlook is that OF-DFT has only reached beyond the nanoscale when computationally cheap energy functionals with limited applicability, such as Thomas-Fermi[8, 9] or other semilocal functionals[10] and simple nonlocal functionals[11, 12] are employed.

Conversely, KS-DFT would deliver the needed accuracy, but the O(N^3) polynomial scaling, where N is a measure of system size, cripples its ability to reach beyond the nanoscale. Linear scaling KS-DFT methods[13–16] and embedding methods[17–21] exist and are successful in extending the applicability of KS-DFT. However, the length scales sought after in this work are still beyond their reach.

It is clear that OF-DFT is a powerful alternative to KS-DFT as it typically involves linear scaling algorithms with a small prefactor[3, 4] and delivers an accuracy essentially equivalent to KS-DFT for simple metals[22, 23].
The SCF technique is realized considering the Lagrangian, \( L[n] = E[n] - \mu \left( \int n(r) dr - N_e \right) \), which is differentiated to find density functions, \( n(r) \), that minimize the total energy, \( E[n] = T_s[n] + E_{Hxc}[n] + \int v_{\text{loc}}(r)n(r)dr \) \( (T_s \text{ and } E_{Hxc} \text{ are the noninteracting kinetic energy and the Hartree-exchange-correlation energies and } v_{\text{loc}} \text{ is the local external potential}) \), subject to the constraint that the density always integrates to a preset number of electrons, \( N_e \). Functional differentiation of the Lagrangian leads to

\[
\hat{h}_{OF} \phi(r) = \mu \phi(r). \tag{1}
\]

Where the pseudo-wavefunction \( \phi(r) = \sqrt{n(r)} \), \( \mu \) is the chemical potential, and the OF-DFT Hamiltonian is given by

\[
\hat{h}_{OF} = -\frac{1}{2} \nabla^2 + v_{\text{Hxc}}[n](r) + v_{\text{Pauli}}[n](r) + v_{\text{loc}}(r), \tag{2}
\]

where the Pauli potential is the difference between the total KEDF potential and the potential of the von Weizsäcker functional, \( v_{\text{Pauli}}[n](r) = v_T[n](r) - v_{\text{W}}[n](r) \). \([32]\)

In this work, we solve the long-standing question of why the SCF procedure delivers poor results for OF-DFT being only capable to reach convergence for very small model systems\([10, 32]\). We found the issue to not lie in numerical instabilities as previously thought \([33, 34]\), but instead it is due to the so-called level-swapping problem, whereby whenever the energy ordering of occupied and virtual states are swapped from one SCF cycle to the next, the SCF procedure cannot converge \([36]\). We analyze this issue in Figure 1. All computations presented in this work are carried out with DFTpy \([30]\). For additional details, we refer to the supplementary materials.\([37]\)

Figure 1 clearly shows that the HOMO-LUMO gap decreases exponentially with system size. This explains why for small model systems (e.g., number of atoms less than 32) the gap is still large enough for the SCF procedure to converge without problems. For large system sizes, the gap is too small and the level-swapping problem appears impeding SCF convergence. A typical way around is to access an ensemble state by slightly smearing the occupations across states within a small window of energy that is still big enough to not fluctuate too much from one SCF cycle to the next \([36]\). This, however, is not an option for OF-DFT because smearing would require computing by diagonalization a number of states that grows like the system size thereby defeating the purpose of using OF-DFT (i.e., to avoid the \( O(N^3) \) complexity of diagonalizations).

We solve the problem by changing the functional to be minimized. In Table I, we summarize three ways of carrying out a density optimization. SCF is equivalent to finding the lowest eigenstate of the Hamiltonian as in Eq. (1) locking the minimizer to only consider pure states. DEM and the new method (mSCF, defined in Table I) are different because they allow the minimizer to probe nonpure states (i.e., ensemble states).

In mSCF, the external potential includes the local potential and the Pauli potential which contains the most expensive energy term in OF-DFT when new-generation KEDFs are employed (i.e., the nonlocal part of the KEDF). In this way, the mSCF optimizer needs to evaluate the nonlocal KEDF potential only once every SCF cycle. In Table II, we show that the number of potential calls for the commonly adopted DEM method grows with system size, while for mSCF not only the number of potential calls is much reduced, but is also insensitive to system size. In mSCF, the von Weizsäcker and the Hartree-xc potentials are invoked as often as in the DEM method. However, they are computationally affordable if (semi)local xc functionals are adopted (as is most often the case). From the timings reported in Figure 2 is clear that the mSCF method is much superior to the commonly adopted DEM, cutting the timing down by orders of magnitude compared to the current state of the art and maintaining linear scalability with system size. These results show that the developed method opens the door to using new-generation nonlocal KEDFs for large-scale simulations.

We should remark that the mSCF method delivers densities and energies identical to the common DEM method which for new-generation KEDFs were shown to be essentially on top of KS-DFT results \([25, 26]\). In Figure 3, we compare our implementation of the new-generation LMGP KEDF against KS-DFT and one of the most accurate force field available to date (Tersoff)\([38]\).

As expected, Figure 3 as well as Table S2 show that OF-DFT energies computed with the LMGP KEDF lie essentially on top of the KS-DFT (LPP) results and very close to the KS-DFT (USPP). To better understand how
TABLE I. The three density optimization methods considered in this work. The external potential is to be considered constant during the minimization. After the minimization is completed, the density-dependent parts of the external potential are updated much like in a typical SCF procedure. The acronyms are as follows. DEM: direct energy minimization; SCF: self-consistent field; mSCF: the newly proposed modified SCF method that combines SCF with DEM. mSCF probes ensemble states without the need to compute extra eigenstates through diagonalization.

| Method | Functional to minimize | External potential \(v_{ext}\) | Ensemble | SCF |
|--------|------------------------|-------------------------------|----------|-----|
| DEM    | \(T_{vW}[n] + E_{Hxc}[n] + T_{\text{Pauli}}[n] + \int v_{ext}(r)n(r)dr\) | \(v_{\text{loc}}(r)\) | Yes       | No  |
| SCF    | \(T_{vW}[n] + \int v_{ext}(r)n(r)dr\) | \(v_{Hxc}(r) + v_{\text{Pauli}}(r) + v_{\text{loc}}(r)\) | No       | Yes |
| mSCF   | \(T_{vW}[n] + E_{Hxc}[n] + \int v_{ext}(r)n(r)dr\) | \(v_{\text{Pauli}}(r) + v_{\text{loc}}(r)\) | Yes       | Yes |

TABLE II. Total number of nonlocal KEDF potential and energy calls during an OF-DFT density optimization (single-point calculation) of bulk Aluminum and Silicon. For mSCF, the number of calls is equal to the number of SCF cycles needed to reach self-consistency.

| # Atoms | Aluminum | Silicon |
|---------|----------|---------|
|         | DEM      | mSCF    | DEM      | mSCF    |
| 8       | 41       | 10      | 67       | 20      |
| 32      | 41       | 10      | 67       | 20      |
| 128     | 59       | 13      | 92       | 13      |
| 512     | 200      | 12      | 211      | 13      |
| 2048    | 595      | 8       | 531      | 10      |

FIG. 2. Wall times for the density optimization of (a) Si clusters and (b) Si bulk supercells. DEM is the commonly adopted direct energy minimization method. mSCF is the newly proposed method. Timings for Al clusters and bulk supercells are available in the supplementary materials.[37] LMGP nonlocal KEDF with density-dependent kernel was employed in both cases.

well OF-DFT ranks the 100 random Si cluster structures, we report in Table S2 the ranking residual standard error (RSE). This is a measure of the error from the trendlines plotted in Figure 3 (i.e., perfect ranking score). We find that OF-DFT’s RSE is 95/75/66 meV which is a decisive improvement compared to the Tersoff force field’s RSE value of 417/307/318 meV for Si\(_{100}/200/300\). We remark that the slopes of the ranking trendline for KS-DFT (LPP) is exactly the same as OF-DFT, and its RSE is 68/54/54 meV. This shows that the majority of OF-DFT’s RSE value is not due to the KEDF employed but rather to the choice of local pseudopotential.

To showcase the ability of our method to approach nanoscale systems, we employ new-generation KEDFs to compute the electronic structure of hydrogen-passivated Si nanoparticles (Si-NPs) of three sizes (about 3, 7, and 15 nm) and two shapes (polyhedral and spherical), see Figure S2 for a depiction of their structures. We focus on their bulk polarization resulting from the surface truncations depicted in Figure 4. As opposed to force fields, OF-DFT in principle can predict any electronic property of materials because the electrons are fully taken into account. Predicting the decay of surface effects (fields, charges, and other electronic properties) is a valuable tool for quantum material science [41]. Bulk polarization of semiconducting materials eludes first-principles simulation techniques because it extends over large distances. For example, the space charge layer width of silicon sur-
FIG. 4. Plane cut views of the bulk polarization of H-passivated Si-NPs. Polyhedral shape: (a) Si_{10117}H_{396}, (b) Si_{10109}H_{1764}, (c) Si_{101199}H_{8460}. Spherical shape: (d) Si_{1118}H_{432}, (e) Si_{10773}H_{1924}, (f) Si_{100788}H_{8704}. The polarization is computed by taking the difference of the electrostatic potentials of Si-NPs and bulk Si. The estimated diameter of the highlighted unpolarized bulk is shown in the figure insets. (a) and (d) are too small to display any appreciable polarization.

faces can be as large as 100 nm and as “small” as 3 nm depending on doping [42] and surface passivation [43]. These length scales are unfortunately still outside the realm of applicability of conventional \textit{ab initio} electronic structure methods while here we approach a 100k-atom Si-NP with a single CPU.

Figure 4 focuses on a 6 nm wide cut centered at the center of the Si-NPs and shows that the polarization of Si-NPs due to the surface truncation extends deeply into the nanoparticle’s bulk leaving only a small core of unpolarized bulk. The estimated widths of the unpolarized bulk regions are indicated in the figure. For the small Si-NPs (insets a and d), there are not enough bulk electrons in them to display appreciable polarization. Conversely, the larger Si-NPs show unpolarized bulks of about 3.3 (4.3) nm and 4.3 (5.4) nm for the 7 and 15 nm polyhedral (spherical) NPs. This converts to a polarization layer width of 2.0 (1.5) nm and 5.7 (5.1) nm for the polyhedral (spherical) NPs. These length scales are compatible with the mentioned sizes of the space charge layer of doped Si surfaces [43], showing that OF-DFT with new-generation KEDFs can be employed as a predictive tool for materials whose intrinsic dimensionality goes well into the tens of nm length scale.

In conclusion, we solve a long-standing problem and are able to apply an SCF-like solver to OF-DFT which we dub mSCF. By evaluating the nonlocal part of the kinetic potential only once per SCF cycle, mSCF allows us to employ otherwise computationally expensive, new-generation kinetic energy functionals almost as efficiently as semilocal functionals. The resulting method opens the door to large-scale OF-DFT simulations at an accuracy that we estimate to be on average below 100 meV/atom compared to KS-DFT. We showcase the method’s computational linear scalability with system size well into the hundreds of thousands of atoms for clusters and bulk systems. And finally, the method’s predictivity is tested computing for the first time with an \textit{ab initio} method the bulk polarization in Si nanoparticles of realistic sizes (up to 15 nm) uncovering a polarization layer that is dependent on the nanoparticle’s size and shape.

**ACKNOWLEDGMENTS**

This material is based upon work supported by the National Science Foundation under Grant No. CHE-1553993. We thank the Office of Advanced Research Computing at Rutgers for providing access to the Amarel cluster. We also thank the Molecular Sciences Software Institute for support through a Software Investment Fellowship.
[1] T. Gould, The Conversation (2016), Welcome to Lab 2.0 where computers replace experimental science.
[2] W. Kohn and L. J. Sham, Phys. Rev. 140, 1133 (1965).
[3] Y. A. Wang and E. A. Carter, in Theoretical Methods in Condensed Phase Chemistry, edited by S. D. Schwartz (Kluwer, Dordrecht, 2000) pp. 117–184.
[4] W. C. Witt, G. Beatriz, J. M. Dieterich, and E. A. Carter, J. M. Res. 33, 777 (2018).
[5] “Atlas is used to calculate 100 million atoms statics energy,” (2019).
[6] X. Shao, Q. Xu, S. Wang, J. Lv, Y. Wang, and Y. Ma, Comput. Phys. Commun. 233, 78 (2018).
[7] L. Hung and E. A. Carter, Chem. Phys. Lett. 475, 163 (2009).
[8] E. Fermi, Rend. Accad. Naz. Lincei 6, 602 (1927).
[9] L. A. Thomas, Proc. Camb. Phil. Soc. 23, 542 (1927).
[10] D. J. González, L. E. González, J. M. López, and M. J. Stott, Phys. Rev. B 65, 184201 (2002).
[11] L.-W. Wang and M. P. Teter, Phys. Rev. B 45, 13196 (1992).
[12] S. Goedecker, Rev. Mod. Phys. 71, 1085 (1999).
[13] D. R. Bowler and T. Miyazaki, Rep. Prog. Phys. 75, 036503 (2012).
[14] J. Lehtomäki, I. Makkonen, M. A. Caro, A. Harju, and O. Lopez-Acevedo, J. Chem. Phys. 141, 234102 (2014).
[15] L. L. Espinosa, A. Karpenko, M. Caro, and O. Lopez-Acevedo, Phys. Chem. Chem. Phys. 17, 31463 (2015).
[16] M. G. Reuter, N. M. Boffi, M. A. Ratner, and T. Seideman, J. Chem. Phys. 138, 084707 (2013).
[17] R. Schlaf, R. Hinogami, M. Fujitani, S. Yae, and Y. Nakato, J. Vac. Sci. Technol., A 17, 164 (1999).
[18] D. C. Gleason-Rohrer, B. S. Brunschwig, and N. S. Lewis, J. Phys. Chem. C 117, 18031 (2013).