# Adaptively Compressed Exchange Operator

Lin Lin*

Department of Mathematics, University of California, Berkeley, California 94720, United States

Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States

ABSTRACT: The Fock exchange operator plays a central role in modern quantum chemistry. The large computational cost associated with the Fock exchange operator hinders Hartree−Fock calculations and Kohn−Sham density functional theory calculations with hybrid exchange-correlation functionals, even for systems consisting of hundreds of atoms. We develop the adaptively compressed exchange operator (ACE) formulation, which greatly reduces the computational cost associated with the Fock exchange operator without loss of accuracy. The ACE formulation is not dependent on the size of the band gap, and thus can be applied to insulating and semiconducting systems, as well as metallic systems. In an iterative framework for solving Hartree−Fock-like systems, such as that observed in planewave-based methods, the ACE formulation only requires moderate modification of the code. The ACE formulation can also be advantageous for other types of basis sets, especially when the storage cost of the exchange operator is

expensive. Numerical results indicate that the ACE formulation can become advantageous, even for small systems with tens of atoms. In particular, the cost of each self-consistent field iteration for the electron density in the ACE formulation is only marginally larger than that of the generalized gradient approximation (GGA) calculation, and thus offers orders-of-magnitude acceleration for Hartree−Fock-like calculations.
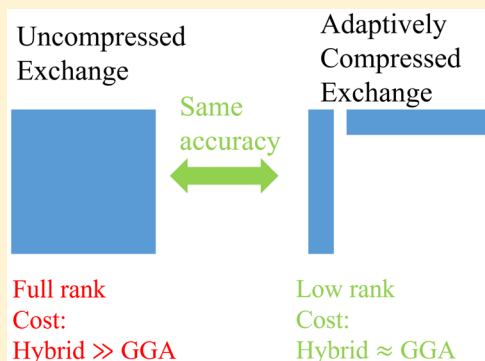
1. INTRODUCTION

The Fock exchange operator, or simply the exchange operator, plays a central role both in wave function theory and in density functional theory (DFT), which are two cornerstones of modern quantum chemistry. Hartree−Fock theory (HF) is the starting point of almost all wave-function-based correlation methods. Kohn−Sham density functional theory (KSDFT)[1,2] is the most widely used electronic structure theory for molecules and systems in condensed phase. The accuracy of KSDFT is ultimately determined by the exchange-correlation (XC) functional employed in the calculation. Despite the great success of relatively simple XC functionals such as local density approximation (LDA),[3,4] generalized gradient approximation (GGA),[5−7] and meta-GGA[8,9] functionals, numerous computational studies in the past two decades suggest that KSDFT calculations with hybrid functionals[10−13] can provide systematically improved description of important physical quantities, such as band gaps, for a vast range of systems. As an example, the B3LYP functional,[10] which is only one specific hybrid functional, has generated more than 55 000 citations (data from ISI Web of Science, January 2016). Hybrid functional calculations are computationally more involved, since it contains a fraction of the Fock exchange term, which is defined using the entire density matrix, rather than the electron density. If the exchange operator is constructed explicitly, the computational cost scales as $O(N_e^4)$, where N_e is the number of electrons of the system. The cost can be reduced to $O(N_e^3)$ by iterative algorithms that avoid the explicit construction of

the exchange operator, but with a large preconstant. Hence, hybrid functional calculations for systems consisting of hundreds of atoms or even less can be a very challenging computational task.

Various numerical methods have been developed to reduce the computational cost of Hartree−Fock-like calculations (i.e., Hartree−Fock calculations and KSDFT calculations with hybrid functionals), most notably methods with asymptotic "linear scaling" complexity.[14,15] The linear scaling methods use the fact that, for an insulating system with a finite HOMO−LUMO gap, the subspace spanned by the occupied orbitals has a compressed representation: it is possible to find a unitary transformation to transform all occupied orbitals into a set of orbitals localized in the real space. This is closely related to the "nearsightedness" of electronic matters.[16,17] Various efforts have been developed to find such localized representation.[18−24] After such localized representation is obtained, the exchange operator also becomes simplified, leading to more efficient numerical schemes for systems of sufficiently large sizes.[24−26] Recent numerical studies indicate that linear scaling methods can be very successful in reducing the cost of the calculation of the exchange term for systems of large sizes with substantial band gaps.[27−29]

In this work, we develop a new method for reducing the computational cost due to the Fock exchange operator. The objective of our method is to find a low rank decomposition of

the exchange operator. However, standard low rank decomposition schemes, such as the singular value decomposition, mandates the low rank operator to yield similar result as the exchange operator does when applied to an arbitrary orbital. This is doomed to fail since the exchange operator is not a low rank operator, and forcefully applied low rank decomposition can lead to unphysical results. The key observation of this work is that, in order to compute physical quantities in Hartree–Fock-like calculations, it is sufficient to construct an operator that yields the same result as the exchange operator does when applied to the *occupied orbitals*. This is possible since the rank of the subspace spanned by the occupied orbitals is known *a priori*. Since occupied orbitals vary in self-consistent field iterations, the compressed representation must be adaptive to the changing orbitals. Hence, our compressed exchange operator is referred to as the adaptively compressed exchange (ACE) operator. We remark that the idea of constructing approximate Fock exchange operator targeting at occupied orbitals has been considered previously. Recent works include real-space-based methods, such as the quantized tensor train (QTT) representation of the exchange operator,[30] and Gaussian orbital-based methods, such as the occ-RI-K method.[31]

The ACE formulation has a few notable advantages:

(1) The ACE is a strictly low rank operator, and there is no loss of accuracy when used to compute physical quantities such as total energies and band gaps.

(2) The effectiveness of the ACE *is not* dependent on the size of the band gap. Hence, the method is applicable to insulators as well as semiconductors or even metals.

(3) The construction cost of the ACE is similar to the one-time application cost of the exchange operator to the set of occupied orbitals. Once constructed, the ACE can be repeatedly used. The cost of applying the ACE is similar to that of applying a nonlocal pseudo-potential operator, thanks to the low rank structure.

(4) In an iterative framework for solving the Hartree–Fock-like equations such as that observed in planewave-based methods, the ACE formulation only requires moderate change of the code.

The ACE formulation can also be advantageous for other types of basis sets, especially when the storage cost of the exchange operator is expensive. This includes real-space methods such as finite-difference methods and finite-element methods. The ACE formulation can be used for small basis sets such as Gaussian orbitals and atomic orbitals as well. However, when it is inexpensive to store the full exchange operator in memory, the advantage of the ACE formulation may only be marginal.

Our numerical results indicate that once the ACE is constructed, the cost of each self-consistent field iteration of the electron density in a hybrid functional calculation is only marginally larger than that of a GGA calculation. The ACE formulation offers significant acceleration, even for small systems with tens of atoms in a serial implementation with a planewave basis set. For moderately larger systems, such as a silicon system with 216 atoms, we observe acceleration by a factor of >50, in terms of the cost of each SCF iteration for the electron density.

The rest of this paper is organized as follows. Section 2 reviews the basic procedure of using iterative methods to solve Hartree–Fock-like equations. Section 3 describes the method of the ACE operator. Section 4 discusses how to apply to the

ACE formulation to Gaussian orbitals and atomic orbitals. The numerical results are presented in section 5, followed by conclusions and future work in section 6.

2. ITERATIVE METHODS FOR SOLVING HARTREE–FOCK-LIKE EQUATIONS

For the sake of simplicity, our discussion below focuses on the Hartree–Fock (HF) equations. The generalization from HF equations to KSDFT equations with hybrid functionals is straightforward and will be mentioned at the end of this section. To simplify the notation, we neglect the spin degeneracy in the discussion below and assume that all orbitals $\{\psi_i(\mathbf{r})\}$ are real. The spin degeneracy is properly included in the numerical results in section 5.

The HF theory requires solving the following set of equations in a self-consistent fashion:

$$H[\{\psi_j\}]\psi_i = \left(-\frac{1}{2}\Delta + V_{\text{ion}} + V_H[\rho] + V_X[\{\psi_j\}]\right)\psi_i = \varepsilon_i\psi_i$$

$$\int \psi_i^*(\mathbf{r})\psi_j(\mathbf{r})\, d\mathbf{r} = \delta_{ij}$$

$$\rho(\mathbf{r}) = \sum_{i=1}^{N_e} |\psi_i(\mathbf{r})|^2 \tag{1}$$

Here, the eigenvalues $\{\varepsilon_i\}$ are ordered nondecreasingly, and N_e is the number of electrons. V_{ion} is a local potential characterizing the electron–ion interaction in all-electron calculations. In pseudo-potential or effective core potential calculations, V_{ion} may contain a low rank and semilocal component as well. V_{ion} is independent of the electronic states $\{\psi_i\}$. $\rho(\mathbf{r})$ is the electron density. The Hartree potential is a local potential, and it is dependent only on the electron density as

$$V_H[\rho](\mathbf{r}, \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}') \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}\, d\mathbf{r}'$$

The exchange operator V_X is a full-rank, nonlocal operator, and it depends on not only the density but also the occupied orbitals $\{\psi_j\}_{j=1}^{N_e}$ as

$$V_X[\{\psi_j\}](\mathbf{r}, \mathbf{r}') = -\sum_{j=1}^{N_e} \frac{\psi_j(\mathbf{r})\psi_j(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \equiv -\frac{\Gamma(\mathbf{r}, \mathbf{r}'; \{\psi_j\})}{|\mathbf{r} - \mathbf{r}'|} \tag{2}$$

Here, $\Gamma(\mathbf{r}, \mathbf{r}'; \{\psi_j\}) = \sum_{j=1}^{N_e} \psi_j(\mathbf{r})\, \psi_j(\mathbf{r}')$ is the single particle density matrix with an exact rank N_e. However, V_X is not a low rank operator, because of the dot product (i.e., the Hadamard product) between Γ and the Coulomb kernel. One common way to solve the HF equations (eqs 1) is to expand the orbitals $\{\psi_j\}_{j=1}^{N_e}$, using a small basis set $\{\chi_\mu\}_{\mu=1}^{N_\mu}$, such as Gaussian-type orbitals, Slater-type orbitals, and numerical atomic orbitals. The basis set is small in the sense that the ratio N_μ/N_e is a small constant (usually in the order of 10). This results in a Hamiltonian matrix H with reduced dimension N_μ. In order to compute the matrix element of H, the four-center integral

$$\iint \frac{\chi_\mu(\mathbf{r})\chi_\alpha(\mathbf{r})\chi_\beta(\mathbf{r}')\chi_\nu(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}\, d\mathbf{r}\, d\mathbf{r}', \quad \alpha, \beta, \mu, \nu = 1, ..., N_\mu$$

must be performed. The cost of the four-center integral is $O(N_\mu^4)$. The quartic scaling becomes very expensive for systems of large sizes. In order to reduce the cost of the

four-center integral, resolution-of-identity (RI)-type techniques (also known as "density fitting")[31−33] are often employed.

For a more complete basis set, such as planewaves and finite elements, the constant N_μ/N_e is much larger (usually 1000 or more), and the cost of forming all four-center integrals is prohibitively expensive, even for very small systems. In such case, it is only viable to use an iterative algorithm, which only requires the application of V_X to several orbitals, rather than the explicit construction of V_X. According to eq 2, V_X applied to any orbital ψ can be computed as

$$(V_X[\{\psi_j\}]\psi)(\mathbf{r}) = -\sum_{j=1}^{N_e} \psi_j(\mathbf{r}) \int \frac{\psi_j(\mathbf{r}')\psi(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,d\mathbf{r}' \tag{3}$$

Equation 3 can be performed by solving N_e Poisson-type problems with an effective charge of the form $\psi_j(\mathbf{r}')\psi(\mathbf{r}')$. For instance, in planewave calculations, if we denote the total number of planewaves by N_g ($N_g \equiv N_\mu$), then the cost for solving each Poisson equation is $O(N_g \log N_g)$ thanks to techniques such as the fast Fourier transform (FFT). Applying V_X to all ψ_i values requires the solution of N_e^2 Poisson problems, and the total cost is $O(N_g \log(N_g)N_e^2)$. The cubic scaling makes iterative algorithms asymptotically less expensive, compared to quartic scaling algorithms associated with the four-center integral calculation. Therefore, for large systems, iterative methods can become attractive, even for calculations with small basis sets such as Gaussian orbitals, as indicated in the recently proposed occ-RI-K method.[31]

The HF equations must be performed self-consistently until the output orbitals $\{\psi_j\}_{j=1}^{N_e}$ from eq 1 agree with those provided as the input to the Hamiltonian operator. However, the Fock exchange energy is only a small fraction (usually <5%) of the total energy, and it is more efficient *not* to update the exchange operator in each self-consistent field iteration. For instance, in planewave-based electronic structure software packages such as Quantum ESPRESSO,[34] the self-consistent field (SCF) iteration of all occupied orbitals can be separated into two sets of SCF iterations. In the inner SCF iteration, the orbitals defining the exchange operator V_X as in eq 2 are fixed, denoted by $\{\varphi_i\}_{i=1}^{N_e}$. The matrix-vector multiplication of V_X and an orbital ψ then is given by

$$(V_X[\{\varphi_j\}]\psi)(\mathbf{r}) = -\sum_{j=1}^{N_e} \varphi_j(\mathbf{r}) \int \frac{\varphi_j(\mathbf{r}')\psi(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,d\mathbf{r}' \tag{4}$$

With V_X fixed, the Hamiltonian operator is only dependent on the electron density ρ, which must be updated in the inner SCF iteration. This allows standard charge mixing schemes, such as Anderson acceleration[35] and Pulay mixing[36] to be used to converge the electron density efficiently. Note that similar techniques to mix the density matrix directly can be prohibitively expensive for large basis sets. Once the inner SCF for the electron density is converged, the output orbitals can simply then be used as the input orbitals to update the exchange operator. The outer SCF iteration continues until convergence is reached. The convergence of the outer iteration can be monitored by the convergence of the Fock exchange energy, which is defined as

$$E_X^{\text{HF}} = -\frac{1}{2} \sum_{i,j=1}^{N_e} \int\int \psi_i(\mathbf{r})\psi_j(\mathbf{r})\psi_j(\mathbf{r}')\psi_i(\mathbf{r}')\frac{1}{|\mathbf{r}-\mathbf{r}'|}\,d\mathbf{r}\,d\mathbf{r}' \tag{5}$$

In each inner SCF iteration, with both the ρ and φ_i parameters fixed, the Hamiltonian operator H becomes a linear operator, and the linear eigenvalue problem

$$\left(-\frac{1}{2}\Delta + V_{\text{ion}} + V_H[\rho] + V_X[\{\varphi_j\}]\right)\psi_i = \varepsilon_i\psi_i \tag{6}$$

must be solved. The linear eigenvalue problem can be solved with iterative algorithms such as the Davidson method[37] and the locally optimal block preconditioned conjugate gradient (LOBPCG) method.[38] Algorithm 1 describes the pseudocode of using iterative methods to solve Hartree−Fock-like equations.

Algorithm 1 Iterative methods for solving Hartree-Fock-like equations.

1: **while** exchange energy is not converged **do**
2:     **while** electron density ρ is not converged **do**
3:         Solve the linear eigenvalue problem (6) with iterative schemes.
4:         Update $\rho^{out}(\mathbf{r}) \leftarrow \sum_{i=1}^{N_e}|\psi_i(\mathbf{r})|^2$.
5:         Update ρ using ρ^{out} and possibly previous history of ρ with charge mixing schemes.
6:     **end while**
7:     Compute the exchange energy E_X.
8:     Update $\{\varphi_j\}_{j=1}^{N_e} \leftarrow \{\psi_j\}_{j=1}^{N_e}$.
9: **end while**

So far, our discussion has been focused on the Hartree−Fock theory. For KSDFT calculations with hybrid functionals, such as the PBE0 functional,[11] the exchange-correlation energy is

$$E_{xc}^{\text{PBE0}} = \frac{1}{4}E_X^{\text{HF}} + \frac{3}{4}E_X^{\text{PBE}} + E_c^{\text{PBE}} \tag{7}$$

Here, E_X^{PBE} and E_c^{PBE} are the exchange and correlation part of the energy from the GGA-type Perdew−Burke−Ernzerhof (PBE) functional,[7] respectively. Hence, the corresponding exchange operator V_X^{PBE0} is simply given by 1/4 of the exchange operator defined in eq 2. For exchange-correlation functionals with screened exchange interactions such as the HSE functional,[12] the exchange-correlation energy is

$$E_{xc}^{\text{HSE}}(\mu) = \frac{1}{4}E_X^{\text{SR}}(\mu) + \frac{3}{4}E_X^{\text{PBE,SR}}(\mu) + E_X^{\text{PBE,LR}} + E_c^{\text{PBE}} \tag{8}$$

Here, $E_X^{\text{PBE,SR}}$ and $E_X^{\text{PBE,LR}}$ refers to short-range and long-range part of the exchange contribution in the PBE functional, respectively. E_X^{SR} is the short-range part of the Fock exchange energy, defined as

$$E_X^{\text{SR}}(\mu) = -\frac{1}{2}\sum_{i,j=1}^{N_e} \int\int \psi_i(\mathbf{r})\psi_j(\mathbf{r})\psi_j(\mathbf{r}')\psi_i(\mathbf{r}')\frac{\text{erfc}(\mu(|\mathbf{r}-\mathbf{r}'|))}{|\mathbf{r}-\mathbf{r}'|}\,d\mathbf{r}\,d\mathbf{r}' \tag{9}$$

Here, erfc is the complementary error function, and μ is an adjustable parameter to control the screening length of the short-range part of the Fock exchange interaction. The only change is to replace the Coulomb kernel by the screened Coulomb kernel, and the screened Coulomb kernel should be used to define the exchange operator V_X^{HSE} accordingly.

3. ADAPTIVELY COMPRESSED EXCHANGE OPERATOR

The most expensive step of Algorithm 1 is the matrix-vector multiplication between the Fock operator V_X and all occupied orbitals. In planewave methods, each set of such matrix-vector multiplication amounts to the solution of N_e^2 Poisson equations. This must be done for each iteration step when

solving the linear eigenvalue problem (eq 6), and in each inner SCF iteration for updating the electron density.

In order to reduce the computational cost, it is desirable to use a low rank decomposition to approximate the Fock exchange operator V_X. However, the exchange operator is a full-rank operator, and a compressed representation, such as the singular value decomposition (SVD), can lead to inaccurate results. However, note that the goal of an SVD is to find an effective operator, denoted by $\tilde{V}_X$, so that the discrepancy measured by $\|V_X\psi - \tilde{V}_X\psi\|_2$ is small for *any orbital ψ*. The key observation of the ACE operator is that the condition above, while desirable, is not necessary to solve Hartree–Fock-like equations. In fact, it is sufficient to construct $\tilde{V}_X$ such that $\|V_X\psi - \tilde{V}_X\psi\|_2$ is small when ψ is *any occupied orbital*, which spans a subspace of strict rank N_e. In this sense, the ACE is designed to be adaptive to the occupied orbitals. When self-consistency of the occupied orbitals is reached, the physical quantities computed in the ACE formulation is exactly the same as that obtained with standard methods for solving Hartree–Fock-like equations. While the idea of constructing approximate Fock exchange operator targeting at occupied orbitals has been considered previously (e.g., see refs 30 and 31), our goal is to construct $\tilde{V}_X$ with exact rank N_e that can be efficiently used without changes in multiple SCF iterations.

More specifically, in each outer iteration, for a given set of orbitals $\{\varphi_i\}_{i=1}^{N_e}$, we first compute

$$W_i(\mathbf{r}) = (V_X[\{\varphi\}]\varphi_i)(\mathbf{r}), \quad i = 1, ..., N_e \tag{10}$$

The ACE operator, denoted by V_X^{ACE}, should satisfy the conditions

$$(V_X^{ACE}\varphi_i)(\mathbf{r}) = W_i(\mathbf{r}) \text{ and } V_X^{ACE}(\mathbf{r},\mathbf{r}') = V_X^{ACE}(\mathbf{r}',\mathbf{r}) \tag{11}$$

One possible choice to satisfy both conditions in eq 11 is

$$V_X^{ACE}(\mathbf{r},\mathbf{r}') = \sum_{i,j=1}^{N_e} W_i(\mathbf{r})B_{ij}W_j(\mathbf{r}') \tag{12}$$

where B is a negative semidefinite matrix to be determined, since V_X is a negative semidefinite operator. In order to determine the matrix B, for any $k, l = 1, ..., N_e$, we require

$$\iint \varphi_k(\mathbf{r})V_X^{ACE}(\mathbf{r},\mathbf{r}')\varphi_l(\mathbf{r}') \, d\mathbf{r} \, d\mathbf{r}' \equiv \int \varphi_k(\mathbf{r})W_l(\mathbf{r}) \, d\mathbf{r}$$
$$= \sum_{i,j=1}^{N_e} \left(\int \varphi_k(\mathbf{r})W_i(\mathbf{r}) \, d\mathbf{r}\right)B_{ij}\left(\int W_j(\mathbf{r}')\varphi_l(\mathbf{r}') \, d\mathbf{r}'\right) \tag{13}$$

Define $M_{kl} = \int \varphi_k(\mathbf{r}) \, W_l(\mathbf{r}) \, d\mathbf{r}$; then, using eq 10, M is a negative semidefinite matrix of size N_e. Equation 13 can be simplified using matrix notation as $M = M B M$, and therefore $B = M^{-1}$. It is straightforward to verify that such choice of B satisfies the requirements in eq 11, despite the fact that eq 13 only projects V_X^{ACE} to the occupied orbitals. Also, because of eq 11, V_X^{ACE} fully agrees with the exchange operator applied to the occupied orbitals when self-consistency is reached, and hence the occupied orbital energies are correct. We will discuss the computation of low-lying unoccupied orbital energies at the end of this section, which is required in the calculation of the HOMO–LUMO gap, as well as excited-state calculations such as GW and Bethe–Salpeter equations.

Perform Cholesky factorization for $-M$, i.e., $M = -LL^T$, where L is a lower triangular matrix, then the solution to eq 13

is $B = -L^{-T}L^{-1}$. Define the projection vector in the ACE formulation as

$$\xi_k(\mathbf{r}) = \sum_{i=1}^{N_e} W_i(\mathbf{r})(L^{-T})_{ik} \tag{14}$$

then the ACE operator is given by

$$V_X^{ACE}(\mathbf{r},\mathbf{r}') = -\sum_{k=1}^{N_e} \xi_k(\mathbf{r})\xi_k(\mathbf{r}') \tag{15}$$

The main advantage of the ACE formulation is the significantly reduced cost of applying V_X^{ACE} to a set of orbitals than that of applying V_X. Once ACE is constructed, the cost of applying V_X^{ACE} to any orbital ψ is similar to the application of a nonlocal pseudopotential, thanks to its low rank structure. ACE only needs to be constructed once when φ_i values are updated in the outer iteration. After construction, the ACE can be reused for all the subsequent inner SCF iterations for the electron density, and each iterative step for solving the linear eigenvalue problem. Since each outer iteration could require 10–100 or more applications of the Hamiltonian matrix H, the cost associated with the solution of the Poisson problem is hence greatly reduced. The pseudocode for iterative methods with the ACE formulation is given in Algorithm 2. Compared to Algorithm 1, the ACE formulation only requires moderate modification of the code.

Algorithm 2 Iterative methods for solving Hartree-Fock-like equations in the ACE formulation.

1: **while** exchange energy is not converged **do**
2:     Compute $\{W_i\}$ according to (10).
3:     Compute $\{\xi_k\}$ according to (14).
4:     **while** electron density ρ is not converged **do**
5:         Solve the linear eigenvalue problem (6) with iterative schemes, with V_X replaced by V_X^{ACE} according to (15).
6:         Update $\rho^{out}(\mathbf{r}) \leftarrow \sum_{i=1}^{N_e} |\psi_i(\mathbf{r})|^2$.
7:         Update ρ using ρ^{out} and possibly previous history of ρ with charge mixing schemes.
8:     **end while**
9:     Compute the exchange energy E_X according to (16).
10:    Update $\{\varphi_j\}_{j=1}^{N_e} \leftarrow \{\psi_j\}_{j=1}^{N_e}$.
11: **end while**

We also remark that ACE can be readily used to reduce the computational cost of the exchange energy, without the need of solving any extra Poisson equation:

$$E_X^{HF} = \frac{1}{2}\sum_{i=1}^{N_e} \iint \psi_i(\mathbf{r})V_X^{ACE}(\mathbf{r},\mathbf{r}')\psi_i(\mathbf{r}') \, d\mathbf{r} \, d\mathbf{r}'$$
$$= -\frac{1}{2}\sum_{i,k=1}^{N_e} \left(\int \psi_i(\mathbf{r})\xi_k(\mathbf{r}) \, d\mathbf{r}\right)^2 \tag{16}$$

So far, we have assumed that the number of $\{\varphi_i\}$ orbitals, denoted by N_φ, is exactly equal to N_e. When unoccupied states are needed, e.g., for the computation of the HOMO–LUMO gap or for excited-state calculations, $N_\varphi > N_e$ should be used. We define the oversampling ratio $r = N_\varphi/N_e$. Similar to the argument that ACE gives the correct occupied orbital energies, when oversampling is used, at self-consistency, we have

$$V_X^{ACE}\psi_i = V_X\psi_i \quad i = 1, ..., N_\varphi$$

Hence, the orbital energies are correct up to the N_φth orbital. Choosing the oversampling ratio $r > 1$ can also be potentially advantageous in the ACE formulation to accelerate the convergence of the outer SCF iteration. This is because when $r > 1$, V_X^{ACE} agrees with the true exchange operator V_X when

applied to orbitals over a larger subspace. Our numerical results, while validating this intuitive understanding, also indicate that the choice $r = 1$ (i.e., $N_\varphi = N_e$) can be good enough for practical hybrid functional calculations when only occupied states are needed.

4. ACE FOR GAUSSIAN ORBITALS AND ATOMIC ORBITALS

The discussion in section 3 focuses on large basis sets such as planewaves, where an iterative method is the only practically viable way for solving the Hartree–Fock-like equations. For Gaussian orbitals and atomic orbitals (referenced hereafter as AO orbitals for simplicity), the computation of the four-center integral is a feasible approach when the total number of basis functions of the system is small. For large systems with a relatively small basis set (e.g., single-ζ or double-ζ basis sets), linear scaling approaches, such as the LinK approach,[39] can become advantageous. On the other hand, for medium-sized systems with larger basis sets (e.g., triple-ζ basis sets or beyond), the recently proposed occ-RI-K method[31] indicates that an iterative framework similar to that used in planewave-based methods can be an attractive alternative.

Assume $\psi_i(\mathbf{r})$ is expanded into the AO basis set $\{\chi_\mu(\mathbf{r})\}$ as $\psi_i(\mathbf{r}) = \sum_\mu \chi_\mu(\mathbf{r}) c_{\mu i}$. The key component of the occ-RI-K method is to compute

$$K_{\mu i} = \int \chi_\mu(\mathbf{r}) V_X(\mathbf{r}, \mathbf{r}') \psi_i(\mathbf{r}') \, d\mathbf{r} \, d\mathbf{r}' \tag{17}$$

Here, i is the index for occupied orbitals only. Hence, the computational cost for constructing $K_{\mu i}$ is less expensive, compared to that for constructing the full exchange operator, $K_{\mu\nu}$:

$$K_{\mu\nu} = \int \chi_\mu(\mathbf{r}) V_X(\mathbf{r}, \mathbf{r}') \chi_\nu(\mathbf{r}') \, d\mathbf{r} \, d\mathbf{r}'$$

$K_{\mu i}$ can be computed using the RI-K technique,[32] instead of Poisson solvers. However, the computation of $K_{\mu i}$ is still the most expensive step, and must be performed for each SCF iteration. One possibility to accelerate the computation is to construct $K_{\mu i}$ less frequently, using a two-level SCF procedure similar to that in Algorithm 1. In the two-level SCF procedure, $K_{\mu i}$ only needs to be computed at the beginning of each outer SCF. If it is inexpensive to store the full exchange operator $K_{\mu\nu}$ in memory, then $K_{\mu\nu}$ in the occ-RI-K formulation can be constructed explicitly, and can be used in multiple inner SCF iterations without changes. If it is not desirable to store the full exchange operator, we demonstrate below that the ACE formulation can be used to provide a compressed representation of the exchange operator as well.

Similar to section 3, we assume the exchange operator in the AO basis set can be expressed using the ACE formulation as

$$K_{\mu\nu}^{\text{ACE}} = \sum_{i,j=1}^{N_e} K_{\mu i} B_{ij} K_{\nu j} \tag{18}$$

Define the matrix

$$M_{ij} = \int \psi_i(\mathbf{r}) V_X(\mathbf{r}, \mathbf{r}') \psi_j(\mathbf{r}') \, d\mathbf{r} \, d\mathbf{r}' = \sum_\mu c_{\mu i} K_{\mu j}$$

and M is a negative semidefinite matrix. Then, similar to eq 13, we have $M = M B M$, which gives $B = M^{-1}$. Perform the Cholesky factorization $M = -LL^T$ and define

$$\xi_{\mu k} = \sum_{i=1}^{N_e} K_{\mu i} (L^{-T})_{ik}$$

we have

$$K_{\mu\nu}^{\text{ACE}} = -\sum_{k=1}^{N_e} \xi_{\mu k} \xi_{\nu k} \tag{19}$$

Hence, we arrive at the ACE formulation for the AO basis set, which is an operator of exact rank N_e, and can be updated only once in a few SCF iterations to reduce the computational cost. When unoccupied orbital energies are needed, we can use the same oversampling strategy as in section 3.

We remark that, to the best of our knowledge, the two-level SCF procedure is not yet standard practice for treating the exchange interaction using AO basis sets. The benefit of the two-level SCF procedure must be studied by future research along this direction.

5. NUMERICAL RESULTS

In this section, we demonstrate the effectiveness of the ACE formulation for accelerating KSDFT calculations with hybrid functionals. The ACE formulation is implemented in the DGDFT software package.[40,41] DGDFT is a massively parallel electronic structure software package for ground-state calculations written in C++. It includes a relatively self-contained module (called PWDFT) for performing standard planewave-based electronic structure calculations. We implement the Heyd–Scuseria–Ernzerhof (HSE06)[12,13] hybrid functional in PWDFT, using periodic boundary conditions with Γ-point Brillouin zone sampling. The screening parameter in the HSE functional μ is set to 0.106 a.u. Our implementation is comparable to that in standard planewave-based software packages such as Quantum ESPRESSO.[34] All results are performed on a single computational core of a 3.4 GHz Intel i-7 processor with 64 GB memory.

We first validate the accuracy of the hybrid functional implementation in PWDFT by benchmarking with Quantum ESPRESSO, and by comparing the converged Fock exchange energy and the HOMO–LUMO gap for a single water molecule (Figure 1) and an 8-atom silicon system (Figure 2). The Hartwigsen–Goedecker–Hutter (HGH) dual-space pseudopotential[42] is used in all calculations. Both Quantum ESPRESSO and PWDFT control the accuracy using a single parameter E_{cut}, which is the kinetic energy cutoff. However,
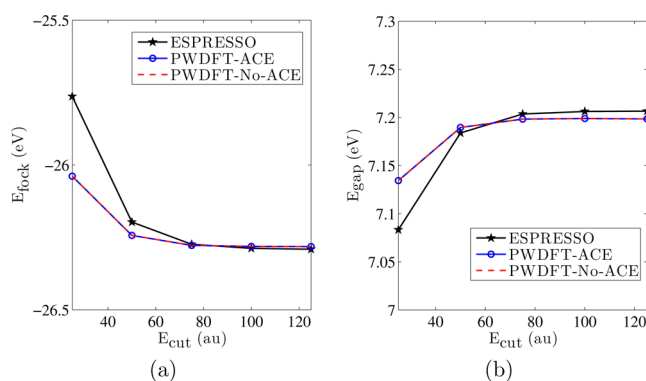

Figure 1. (a) Fock exchange energy and (b) HOMO–LUMO gap obtained from Quantum ESPRESSO and PWDFT with and without the ACE formulation for a water molecule.
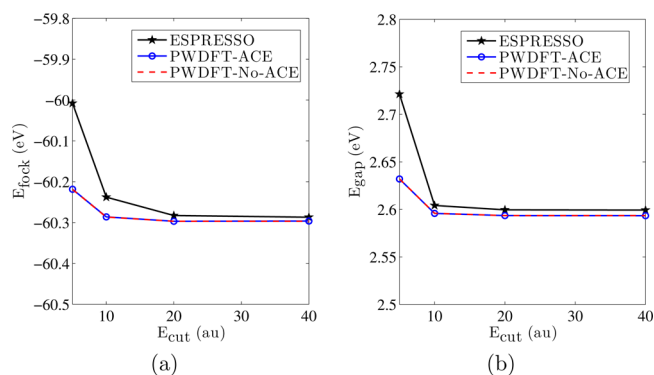
Figure 2. (a) Fock exchange energy and (b) HOMO−LUMO gap obtained from Quantum ESPRESSO and PWDFT with and without the ACE formulation for a silicon system with 8 atoms.

there is a notable difference in the detailed implementation. For instance, PWDFT uses a real-space implementation of the pseudopotential with a pseudocharge formulation,[43] and implements the exchange-correlation functionals via the LibXC[44] library, while Quantum ESPRESSO uses a Fourier space implementation of the HGH pseudopotential converted from the CPMD library,[45] and uses a self-contained implementation of exchange-correlation functionals. Nonetheless, at sufficiently large E_{cut}, the difference of the total Fock exchange energy between Quantum ESPRESSO and PWDFT is only 9 meV for the water system and 11 meV for the silicon system, and the difference of the gap is 8 meV for the water system and 5 meV for the silicon system, respectively. The HOMO−LUMO gap is calculated using the oversampling method discussed at the end of section 3 by setting $r = 2$, which allows the ACE formulation to compute orbital energies correctly up to $N_\varphi = 2N_e$. In both systems, the difference of the results from PWDFT is negligibly small between the standard implementation of hybrid functional (No-ACE), and with the ACE formulation.

In section 3, the oversampling ratio $r = N_\varphi/N_e$ is defined. It is conceivable that as r increases, the convergence of the outer iteration for the orbitals $\{\varphi_i\}$ can accelerate. Figures 3a and
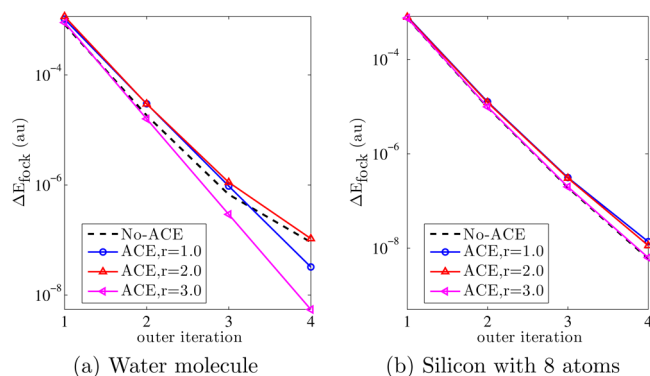

Figure 3. Convergence of the difference of the exchange energy.

Figures 3b report the convergence of the difference of the Fock exchange energy at each outer iteration, with respect to different oversampling ratios r for the water and silicon system, respectively, as a measure of the convergence of the outer iteration. The kinetic energy cutoff for the water and silicon systems is set to 100 and 20 a.u., respectively. The convergence

without the ACE formulation is also included for comparison. We observe that as the oversampling ratio increases, the convergence rate of the outer iteration becomes marginally improved. In fact, the convergence rate using the ACE formulation with $r = 1$ is very close to that without the ACE formulation at all. This indicates that the use of the ACE formulation does not hinder the convergence rate of the hybrid functional calculation.

In order to demonstrate the efficiency of the ACE formulation for hybrid functional calculations, we study three silicon systems with increasing sizes 8, 64, and 216 atoms, respectively. The latter two systems correspond to a silicon unit cell with 8 atoms replicated into a $2 \times 2 \times 2$ and a $3 \times 3 \times 3$ supercell, respectively. Since hybrid functional is implemented in PWDFT so far in the serial mode, we use a relatively small kinetic energy cutoff $E_{cut} = 5$ a.u. in these calculations. Nonetheless, the kinetic energy cutoff mainly affects the cost of the fast Fourier transforms (FFTs), and we expect that the ACE formulation will become more advantageous with a higher E_{cut}, in terms of the reduction of the absolute computational time. Figure 4 shows the time cost of each inner SCF iteration for the electron density, which involves 10 LOBPCG iterations, for the calculation with the HSE functional with and without the ACE formulation. This is the same as the cost per outer SCF iteration in Algorithm 2 when the number of inner SCF iterations is set to be 1. Even in this case, the ACE formulation already exhibits a significant computational advantage, and the reduction of the computational time would be more significant when the same ACE operator is used for multiple inner SCF iterations. For comparison, we also include the time cost of each inner SCF iteration for the electron density in a GGA functional calculation using the Perdew−Burke−Ernzerhof (PBE) functional,[7] of which the cost is much less expensive. The cost of the construction phase of the ACE formulation is marked separately as "ACE,Construct" in Figure 4.

First, we confirm that the cost of each hybrid functional calculations is much higher than that of GGA calculations. The time per inner SCF iteration for the electron density of the HSE calculation without ACE is 42 times higher than that of the PBE calculation for the 64-atom system. This ratio increases by a factor of 58 for the 216-atom system. With the ACE formulation, this ratio is reduced to 1.18 and 1.05, for the 64- and 216-atom systems, respectively (i.e., the cost of each HSE calculation in the ACE formulation is only marginally larger than that of the GGA calculation). Although the construction of the ACE still requires solving a large number of Poisson equations, the overall time is greatly reduced since the ACE, once constructed, can be used for multiple times. Even assuming that the outer SCF iteration only consists of one inner iteration, for the system with 216 atoms, the ACE formulation already achieves an overall acceleration by a factor of 8.8 times, compared to the standard implementation. The ACE formulation becomes orders of magnitude more efficient when multiple inner SCF iterations are required, which is usually the case both in PWDFT and in other software packages, such as Quantum ESPRESSO.

6. CONCLUSION

We have introduced the adaptively compressed exchange (ACE) operator formulation for compressing the Fock exchange operator. The main advantage of the ACE formulation is that there is no loss of accuracy, and its effectiveness is not dependent on the size of the band gap.
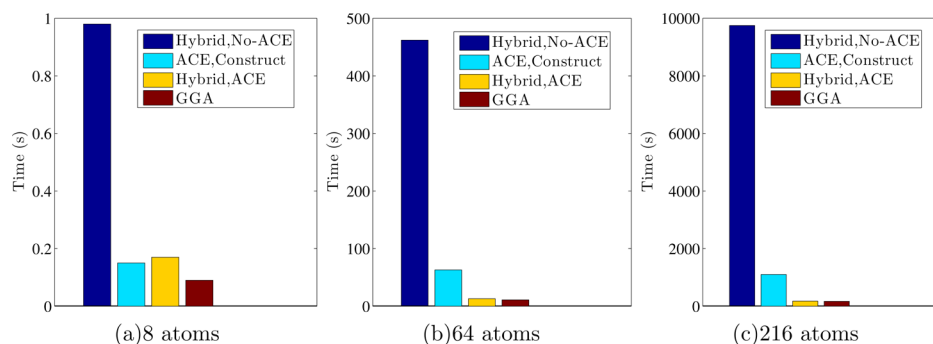
Figure 4. Computational time for silicon systems with increasing system sizes.

Hence, the ACE formulation can be used for insulators and semiconductors, as well as metals. We demonstrated the use of the ACE formulation in an iterative framework for solving Hartree−Fock equations and Kohn−Sham equations with hybrid exchange-correlation functionals. The ACE formulation only requires moderate modification of the code, and it can potentially be applied to all electronic structure software packages for treating the exchange interaction in an iterative framework. The construction cost of the ACE formulation is the same as applying the Fock exchange operator once to the occupied orbitals. Once constructed, the cost of each self-consistent field (SCF) iteration for the electron density in hybrid functional calculations becomes only marginally larger than that of generalized gradient approximation (GGA) calculations. Our numerical results indicate that the computational advantage of the ACE formulation can be clearly observed, even for small systems with tens of atoms.

We have demonstrated the numerical efficiency of the ACE formulation in a planewave basis set, and we have introduced the ACE formulation for Gaussian orbitals and atomic orbitals following the occ-RI-K method. For insulating systems, the cost of the ACE formulation can be further reduced when combined with linear scaling type methods. For range-separated hybrid functionals, it might even be possible to localize the projection vectors ξ_k, because of the screened Coulomb interaction in the real space. This could further reduce the construction as well as the application cost of the ACE, and it opens the door to Hartree−Fock-like calculations for a large range of systems currently beyond reach.

inspires us to extend the ACE formulation to Gaussian orbitals and atomic orbitals.

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: linlin@math.berkeley.edu.

Notes
The authors declare no competing financial interest.

■ REFERENCES

(1) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864−B871.
(2) Kohn, W.; Sham, L. *Phys. Rev.* **1965**, *140*, A1133−A1138.
(3) Ceperley, D. M.; Alder, B. J. *Phys. Rev. Lett.* **1980**, *45*, 566−569.
(4) Perdew, J. P.; Zunger, A. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1981**, *23*, 5048−5079.
(5) Becke, A. D. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098−3100.
(6) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785−789.
(7) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(8) Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. *J. Chem. Phys.* **2003**, *119*, 12129−12137.
(9) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215−241.
(10) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
(11) Perdew, J. P.; Ernzerhof, M.; Burke, K. *J. Chem. Phys.* **1996**, *105*, 9982−9985.
(12) Heyd, J.; Scuseria, G. E.; Ernzerhof, M. *J. Chem. Phys.* **2003**, *118*, 8207−8215.
(13) Heyd, J.; Scuseria, G. E.; Ernzerhof, M. *J. Chem. Phys.* **2006**, *124*, 219906.
(14) Goedecker, S. *Rev. Mod. Phys.* **1999**, *71*, 1085.
(15) Bowler, D. R.; Miyazaki, T. *Rep. Prog. Phys.* **2012**, *75*, 036503.
(16) Kohn, W. *Phys. Rev. Lett.* **1996**, *76*, 3168−3171.
(17) Prodan, E.; Kohn, W. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 11635−11638.
(18) Foster, J. M.; Boys, S. F. *Rev. Mod. Phys.* **1960**, *32*, 300.
(19) Marzari, N.; Vanderbilt, D. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1997**, *56*, 12847.
(20) Marzari, N.; Mostofi, A. A.; Yates, J. R.; Souza, I.; Vanderbilt, D. *Rev. Mod. Phys.* **2012**, *84*, 1419−1475.
(21) Gygi, F. *Phys. Rev. Lett.* **2009**, *102*, 166406.
(22) E, W.; Li, T.; Lu, J. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 1273−1278.
(23) Ozoliņš, V.; Lai, R.; Caflisch, R.; Osher, S. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 18368−18373.
(24) Damle, A.; Lin, L.; Ying, L. *J. Chem. Theory Comput.* **2015**, *11*, 1463−1469.
(25) Wu, X.; Selloni, A.; Car, R. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2009**, *79*, 085102.
(26) Gygi, F.; Duchemin, I. *J. Chem. Theory Comput.* **2013**, *9*, 582−587.
(27) Chen, W.; Wu, X.; Car, R. *Phys. Rev. Lett.* **2010**, *105*, 017802.
(28) DiStasio, R. A.; Santra, B.; Li, Z.; Wu, X.; Car, R. *J. Chem. Phys.* **2014**, *141*, 084502.
(29) Dawson, W.; Gygi, F. *J. Chem. Theory Comput.* **2015**, *11*, 4655−4663.
(30) Khoromskaia, V.; Khoromskij, B.; Schneider, R. *Comput. Methods Appl. Math.* **2011**, *11*, 327−341.

(31) Manzer, S.; Horn, P. R.; Mardirossian, N.; Head-Gordon, M. *J. Chem. Phys.* **2015**, *143*, 024113.

(32) Weigend, F. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285−4291.

(33) Lu, J.; Ying, L. *J. Comput. Phys.* **2015**, *302*, 329.

(34) Giannozzi, P.; Baroni, S.; Bonini, N.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Chiarotti, G. L.; Cococcioni, M.; Dabo, I.; et al. *J. Phys.: Condens. Matter* **2009**, *21*, 395502−395520.

(35) Anderson, D. G. *J. Assoc. Comput. Mach.* **1965**, *12*, 547−560.

(36) Pulay, P. *Chem. Phys. Lett.* **1980**, *73*, 393−398.

(37) Davidson, E. *J. Comput. Phys.* **1975**, *17*, 87−94.

(38) Knyazev, A. V. *SIAM J. Sci. Comput.* **2001**, *23*, 517−541.

(39) Ochsenfeld, C.; White, C. A.; Head-Gordon, M. *J. Chem. Phys.* **1998**, *109*, 1663−1669.

(40) Lin, L.; Lu, J.; Ying, L.; E, W. *J. Comput. Phys.* **2012**, *231*, 2140−2154.

(41) Hu, W.; Lin, L.; Yang, C. *J. Chem. Phys.* **2015**, *143*, 124110.

(42) Hartwigsen, C.; Goedecker, S.; Hutter, J. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1998**, *58*, 3641−3662.

(43) Pask, J. E.; Sterne, P. A. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2005**, *71*, 113101−113104.

(44) Marques, M. A. L.; Oliveira, M. J. T.; Burnus, T. *Comput. Phys. Commun.* **2012**, *183*, 2272−2281.

(45) Hutter, J.; Curioni, A. *Parallel Comput.* **2005**, *31*, 1−17.