THE JOURNAL OF PHYSICAL CHEMISTRY A
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCA Article

Use of Multigrids to Reduce the Cost of Performing Interpolative Separable Density Fitting

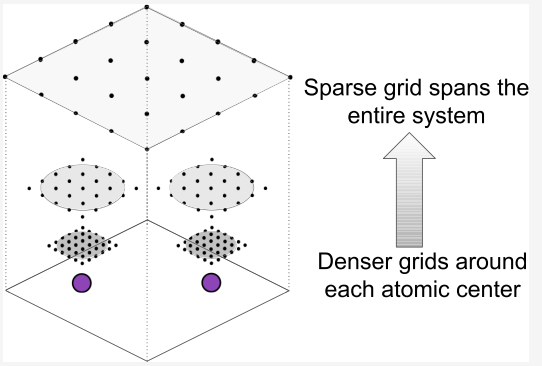
Published as part of The Journal of Physical Chemistry A virtual special issue "Gustavo Scuseria Festschrift".

Kori E. Smyser, Alec White, and Sandeep Sharma*

Cite This: *J. Phys. Chem. A* 2024, 128, 7451−7461 Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: In this article, we present an interpolative separable density fitting (ISDF)-based algorithm to calculate the exact exchange in periodic mean field calculations. In the past, decomposing the two-electron integrals into the tensor hypercontraction (THC) form using ISDF was the most expensive step of the entire mean field calculation. Here, we show that by using a multigrid-ISDF algorithm, both the memory and the CPU cost of this step can be reduced. The CPU cost is brought down from cubic scaling to quadratic scaling with a low computational prefactor which reduces the cost by almost 2 orders of magnitude. Thus, in the new algorithm, the cost of performing ISDF is largely negligible compared to other steps. Along with the CPU cost, the memory cost of storing the factorized two-electron integrals is also reduced by a factor of up to 35. With the current algorithm, we can perform Hartree−Fock calculations on a diamond supercell containing more than 17,000 basis functions and more than 1500 electrons on a single node with no disk usage. For this calculation, the cost of constructing the exchange matrix is only a factor of 4 slower than the cost of diagonalizing the Fock matrix. Augmenting our approach with linear scaling algorithms can further speed up the calculations.

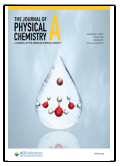


INTRODUCTION

Much of modern ab initio computational chemistry and materials science is based on Kohn−Sham (KS) density functional theory (DFT).[1] The inclusion of exact Hartree−Fock (HF) exchange within this framework[2] has been instrumental to the success of DFT for molecular systems to the point that almost all modern molecular calculations rely on these "hybrid" density functionals.[3] Hybrid functionals can outperform their semilocal counterparts for some properties of periodic solids,[4−9] but the cost of evaluating the nonlocal exchange contribution may be prohibitive.

Methods for the efficient evaluation of exact exchange are well-developed in the context of molecular calculations. Such calculations typically use a relatively small set of local basis functions such that the ratio of basis functions to electrons, N/n, is often less than ten. Computing every element of the fourth-order tensor of electron repulsion integrals, which one might naively expect to be necessary for both the Coulomb and exchange contributions, would scale like $O(N^4)$. However, the locality of the basis functions implies that there are asymptotically only a linear number of *significant* basis function pairs, which means that the Coulomb and exchange contributions can be computed in $O(N^2)$ time.[10] The scaling of the Coulomb contribution can be reduced to linear, $O(N)$, using the multipole expansion and fast multipole method.[11−13] The exchange contribution can also be computed in asymptotically linear time for nonmetallic systems by leveraging locality in the density matrix.[14−17] For most practical molecular calculations, these asymptotically linear methods come with a large prefactor, and it is preferable to reduce the cost of higher-scaling algorithms with tensor factorization. The resolution of the identity (RI) method is one such tensor factorization technique that is commonly applied to both Coulomb and exchange contributions with the RI-J[18,19] and RI-K[20−23] algorithms, respectively. These RI methods are also called "density fitting"; and Dunlap showed how a "robust" fit can be used to make the error in the fitted two-electron integral quadratic in the error for basis function pairs.[24−26] RI approaches usually rely on predetermined, atom-centered basis sets of fitting functions. Circumventing this requirement, local-RI using numerical basis[27,28] and Cholesky decomposition approach yields a factorization of the same form without the need for preoptimized fitting basis sets.[29,30] The

Received: April 14, 2024
Revised: July 3, 2024
Accepted: July 30, 2024
Published: August 26, 2024



https://doi.org/10.1021/acs.jpca.4c02431
J. Phys. Chem. A 2024, 128, 7451−7461

pseudospectral (PS) method is an alternative factorization that uses a partial real-space quadrature to factorize the two-electron integrals,[31] and the chain of spheres algorithm for exchange (COSX)[32] is a commonly used implementation of the PS idea. In recent years, the tensor hypercontraction (THC) method of Martínez and co-workers took the idea of tensor factorization to the logical limit.[33−35] The THC method factorizes the four-index tensor of two-electron integrals into a product of two-index tensors—a drastic factorization. But obtaining an accurate THC factorization is generally difficult so initial applications to correlated methods are limited. However, the "interpolative separable density fitting" (ISDF) method[36] can provide a factorization of THC form with only cubic, $O(N^3)$, scaling, and it has since been used in various algorithms for exact exchange.[37−42]

On the other hand, calculations on periodic solids often use a large basis set of N_g plane waves. In these calculations, the action of the Coulomb operator on just the occupied space is determined by solving n Poisson equations, which leads to quadratic scaling, $O(nN_g \ln N_g)$, or linear scaling when using translational (k-point) symmetry. Unfortunately, the action of the exchange operator on the occupied space is cubic, $O(n^2N_g \ln N_g)$, or quadratic with k-point symmetry. So, in a plane-wave basis set, the exchange contribution is higher scaling than the Coulomb part, and there have been many efforts to reduce this cost. Linear scaling methods have been developed for both the Coulomb and exchange.[16,43] As in the molecular case, traditional linear scaling exchange algorithms rely on locality in the density matrix for insulating systems.[44] An exception is stochastic density functional theory (sDFT), which can reduce the prefactor and scaling of the exchange calculation by using the stochastic resolution of identity method.[45−47] For typical calculations, asymptotically linear scaling methods are not practical, and methods to improve the efficiency without addressing the scaling can result in useful speedups.[48] Additional examples include the adaptively compressed exchange (ACE) method[49] and the auxiliary density matrix method.[50] Methods that use ISDF-THC for exact exchange in solids,[37,38,40−42] including the method presented in this work, fall into this category.

Since its introduction by Lu and Ying in 2015,[36] ISDF has been quickly adopted to speed up the exchange calculations in codes that use Gaussian orbitals,[40,41] numerical atomic orbitals,[51] and plane wave basis sets.[37,38,42] Although the computational scaling of performing ISDF is cubic with the system size (the same as the ultimate computation of the exchange matrix), the computational prefactor is high, making it one of the most expensive steps of the entire calculation. Furthermore, the memory requirements are high, which limits its applicability to small systems unless massively parallel computers are used. Recently, it has been realized that the memory cost can be reduced if one performs interpolative decomposition of the occupied molecular orbitals rather than the entire atomic orbital basis set.[41] This does reduce the memory requirement but this comes at an additional cost of having to perform this decomposition at every self-consistent field (SCF) iteration. There is also additional computational overhead related to constructing the exchange matrix (for details we refer the reader to ref 41).

In this work, we perform the interpolative decomposition on the atomic integrals, which is only done once, and simultaneously reduce the computational and memory cost of performing this step using ideas from so-called "multigrid" approaches.[52−57] As we will show in the results, the computational cost of this step is no longer the leading cost of the algorithm. We can perform calculations on systems with >10,000 basis functions on a single node without running out of memory. Along with reduced memory requirements, we also show that the cost of performing ISDF calculations is nearly eliminated. The key ingredient of the algorithm is to use multiple local grids of varying resolutions, each of which supports only a subset of atomic orbitals (Figure 1). This idea has been used in the past to significantly speed up the calculation of the Coulomb operator in CP2K,[53,57] and here we extend this approach to accelerate exchange evaluation within the ISDF-THC framework.

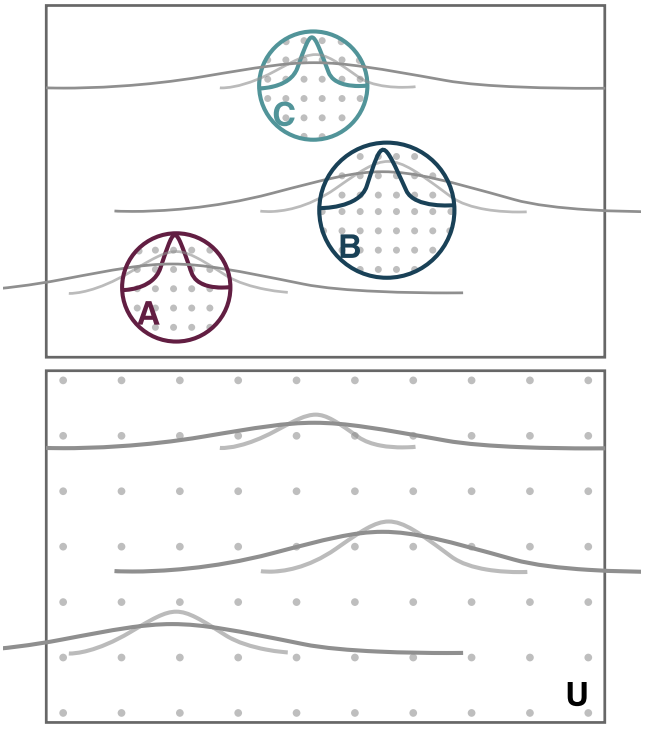


Figure 1. Example of the grid structure used for building the exchange matrix. The most dense grid is only defined on spherical, atom-centered regions (colored circles, A, B, C; top). The volume of each atom-centered grid is determined by a cutoff radius beyond which the local GTOs (L) with exponents larger than α_{min} (colored lines) are expected to go to zero (see eq 5). A sparse universal grid (U) spans the full supercell and supports all global GTOs (G) but is only required by exponents smaller than α_{min}.

In the rest of the paper, we will focus on periodic calculations with Gaussian basis functions in the presence of Pseudopotentials, although the ideas can be extended to mixed Gaussian/plane-wave basis and all-electron calculations. We will begin the paper by recalling how the interpolative decomposition is typically performed to obtain integrals in the THC form. This will be followed by our updated algorithm that shows how the memory and CPU cost of performing this step can be significantly reduced. We end the paper with some results and prospects for future work.

THEORY

A major bottleneck in a simple implementation of hybrid-DFT calculations is the need to evaluate the two-electron integrals,

https://doi.org/10.1021/acs.jpca.4c02431
J. Phys. Chem. A 2024, 128, 7451−7461

$$(\mu\nu|\lambda\sigma) = \int\int \mu(r_1)\nu(r_1)\frac{1}{r_{12}}\lambda(r_2)\sigma(r_2)\mathrm{d}r_1\mathrm{d}r_2$$

Instead of calculating the entire four-index quantity one can decompose it into a product of several two-index quantities. We begin by noting that the product of the orbitals $\mu(\mathbf{R})\nu(\mathbf{R}) = (\mu\nu|\mathbf{R})$ can be viewed as a matrix with two indices, the first index being a composite index consisting of a pair of orbitals, $\mu\nu$, and the second being a set of suitably chosen grid points with a sufficiently high density, $\mathbf{R}$. This matrix $(\mu\nu|\mathbf{R})$ is low-rank and can be decomposed as,

$$(\mu\nu|\mathbf{R}) = \sum_{\xi} (\mu\nu|\xi)\xi(\mathbf{R})$$

where the size of the index ξ is smaller than both the square of the number of basis functions (N^2) and the number of grid points (N_g) (Table 1). An optimal decomposition that minimizes the Frobenius norm of the error is given by SVD. However, with SVD we lose the separability of the original matrix (note that while $(\mu\nu|\mathbf{R}) = \mu(\mathbf{R})\nu(\mathbf{R})$, $(\mu\nu|\xi) \neq \mu(\xi)\nu(\xi)$). One can instead perform interpolative decomposition that ensures that the indices ξ in the matrix $(\mu\nu|\xi)$ are just a subset of grid points $\mathbf{R}$. The disadvantage of interpolative decomposition is that one does not have an optimal algorithm to find it and the Frobenius norm of the error is guaranteed to be greater than or equal to that from SVD, but the separability of the resulting matrix is retained, i.e., $(\mu\nu|\xi) = \mu(\xi)\nu(\xi)$, which more than makes up for the shortcomings.

Table 1. Notation Used in Paper

N_g	Number of grid points
N_ξ	Number of fitting functions
N	Number of basis functions
n	Number of electrons
n_{atom}	Number of atoms
$\mu, \nu, \cdots$	Indices of atomic orbitals
$i, j, \cdots$	Indices of the occupied molecular orbitals
$p, q, \cdots$	Indices of any molecular orbital
ξ	The ISDF fitting functions including ξ_A and R_U shown below
ξ_A	The ISDF fitting functions centered on atom A
R_U	The Sinc functions in the universal grid

After having performed the interpolative decomposition of the orbital products one can then write the two-electron integrals as,

$$(\mu\nu|\lambda\sigma) = \sum_{\xi\xi'} \mu(\xi)\nu(\xi)V(\xi, \xi')\lambda(\xi')\sigma(\xi')$$

where the matrix $V(\xi, \xi')$ can be evaluated numerically as,

$$V(\xi, \xi') = \int\int \xi(r_1)\frac{1}{r_{12}}\xi'(r_2)\mathrm{d}r_1\mathrm{d}r_2 \quad (1)$$

using fast Fourier transform (FFT).

Typically, THC requires more fitting functions $\xi(r)$ than RI—while the error in the two-electron integrals is quadratic in the RI fitting error, it is linear in the THC error. One can use robust tensor hypercontraction, also known as the robust pseudospectral (rPS) method, to make the error from two-electron integrals quadratic in the fitting errors (similar to RI) and use fewer functions, as in RI. rPS is known to produce nonpositive definite two-electron integrals that can cause variational collapse of the SCF cycles.[58,59] But we have never seen this in our previous work because we never use rPS to evaluate the Coulomb matrix—only the exchange matrix, which is itself negative definite. This has also never been observed in the work of Manzer et al.[22] when they use Pair Atomic Resolution of the Identity Approximation for exact exchange (PARI-K).

Interpolative Decomposition. As mentioned in the previous section, one needs to perform an interpolative decomposition of the two-electron integrals. The most common way of doing this is to perform pivoted-QR decomposition of the $(\mu\nu|\mathbf{R})$ matrix. A simple algorithm would lead to a computational cost of $O(N^4)$, making the entire algorithm prohibitively expensive. Lu and Ying[36] in their original paper introduced a randomized algorithm where one first obtains two random matrices G^1 and G^2 of size $N \times p$ each, with $p = \sqrt{N_\xi} + \delta$ orthogonal columns, where N_ξ is the number of THC functions and δ is a small number usually around 5.[60,61] A randomized density matrix is constructed from these matrices according to,

$$\rho_{mn,R} = \left(\sum_{\mu} G^1_{\mu m}\mu(\mathbf{R})\right)\left(\sum_{\nu} G^2_{\nu n}\nu(\mathbf{R})\right)$$

One can then perform a pivoted-QR decomposition on the matrix $\rho_{mn,R}$ to obtain the pivots. The pivots from the randomized matrix will be of similar quality to those obtained from the full matrix $(\mu\nu|\mathbf{R})$ as long as its singular values decay sufficiently quickly, as they do here. The overall cost of the randomized algorithm is $O(N^3)$ which is a significant improvement over the deterministic algorithm.

Matthews suggested[62] that one can improve the efficiency of the algorithm by first forming a matrix

$$M(\mathbf{R}, \mathbf{R}') = \sum_{\mu\nu} (\mu\nu|\mathbf{R})(\mu\nu|\mathbf{R}') \quad (2)$$

and then perform a pivoted-Cholesky decomposition on it to obtain the pivots ξ. The methods give the same pivot points (when randomization is not introduced), and pivoted Cholesky is typically significantly faster than pivoted QR. This algorithm is extremely efficient especially if the matrix M can be stored in memory. Later we will show that for our purposes these matrices are indeed small enough to be stored in memory. It is also worth mentioning that a third approach called centroid-Voronoi-tesselation (CVT), that scales as $O(N^2)$, is also widely used in this context.[37,39]

Having obtained the pivot points ξ, a least-squares algorithm obtains the functions $\xi(\mathbf{R})$ that minimize the error,

$$\min_{\xi(\mathbf{R})} |(\mu\nu|\mathbf{R}) - \sum_{\xi} (\mu\nu|\xi)\xi(\mathbf{R})| \quad (3)$$

This can be done with an $O(NN_\xi N_g + N_\xi^3 + N_\xi^2 N_g)$ cost, dominated by $O(N_\xi^2 N_g)$. As mentioned in the introduction, this algorithm has a rather steep memory requirement because one has to store the fitting functions $\xi(\mathbf{R})$ at the cost of $N_g \times N_\xi$. The value of N_g can become significant even if there is a single sharp function in the basis set.

There are a few ways of overcoming the high cost of ISDF calculation:

1. In a previous publication,[40] we have shown that one can reduce both the memory and CPU cost of ISDF by using a robust fitting procedure, which reduces the number of ISDF functions needed to get an accurate result by about a

https://doi.org/10.1021/acs.jpca.4c02431
J. Phys. Chem. A 2024, 128, 7451−7461

factor of 2 (for instance, compare THC and rPS in Figure 5). Although the cost of ISDF is reduced, it remains the dominant cost of the calculation.

2. The cost of doing ISDF can be eliminated by not doing ISDF but instead by solely relying on FFT and using the occ-RI (occ refers to occupied orbitals) trick of Manzer et al.[23] occ-RI relies on the fact that the value and gradient of the DFT energy can be obtained simply by knowing the occupied-virtual block of the exchange matrix K_{ip}, which is given by

$$K_{ip} = \sum_j \int\int \phi_i(r_1)\phi_j(r_2)\frac{1}{r_{12}}\phi_j(r_1)\phi_p(r_2)\mathrm{d}r_1\mathrm{d}r_2 \quad (4)$$

where we have assumed that the orbitals are real. If all the orbitals are representable on an FFT grid of size N_g then this entire matrix can be evaluated by performing n^2 Poisson solves, and matrix multiplications with the cost equal to $O(n^2N_g\ln(N_g))$ and $O(nNN_g)$, respectively. Out of the two steps, we find that the cost of Poisson solves $O(n^2N_g\ln(N_g))$ dominates. Because the ISDF calculation is not used, the memory requirement for storing the fitting functions $\xi(\mathbf{R})$ $(N_g \times N_\xi)$ is eliminated.

3. Instead of performing ISDF on the products of atomic orbitals once at the start of the calculation, one can perform a new ISDF calculation on the product of molecular orbitals at each SCF iteration. The two dominant costs of this algorithm are the same as that of AO-based ISDF—$O(N_\xi^2N_g)$ for matrix multiplications and $O(N_\xi N_g\ln(N_g))$ for N_ξ Poisson solves. The potential advantage is that the N_ξ required is independent of the basis set, it only depends on the number of electrons, and it is expected to be smaller than the N_ξ from ISDF on atomic orbital pairs. The disadvantage is that one has to perform an ISDF at each SCF iteration and ISDF remains the dominant cost of the calculation. This approach, of performing ISDF on the molecular orbitals, was first pointed out by Hu et al.[37] It was recently extended to use with Gaussian basis sets and k-point sampling by Rettig et al.[41] where they pointed out a few terms that were missing in the gradient of the exchange energy in ref 37.

For Γ-point calculations we expect the CPU cost of the FFT-based approach and rPS to be lower than that of the MO-based ISDF calculations. The memory cost of the FFT-based approach is superior to the other two because one does not have to store the ISDF fitting functions.

In this work, we introduce a third approach that relies on the use of multiple grids of varying resolutions. Our algorithm uses both a single-shot ISDF by using pivoted Cholesky on atomic orbitals and an iterative FFT-based Poisson solution at each SCF cycle. The ISDF is only performed for products of atomic orbitals where at least one of the orbitals is sharp and the FFT is only used to solve the Poisson equation for products where both orbitals are diffuse, as described in more detail below.

Using Multiple Grids for Exchange. In this section we describe the basic idea of our multigrid algorithm for calculating exchange and go into more technical details in the next section. We begin with an uncontracted Gaussian-type orbital (GTO) basis and partition it into two sets. The first set contains sharp Gaussian basis functions with large exponents and the second set contains diffuse basis functions with small exponents. The product of sharp−sharp and sharp-diffuse atomic orbitals are approximated using ISDF on a grid centered around the atom on which the sharp function is centered (see Figure 1). Because the functions are sharp here, they require the grid to have a high point density, though for the same reason they do not span the full unit cell. Notice that the sharp functions with large exponents decay rapidly and their product with any other function is only expected to be nonzero in the local spatial region where it is itself nonzero. The rest of the products between diffuse−diffuse functions are treated using the iterative FFT approach outlined in the second bullet point of previous Section. This does not use ISDF but is solely based on Poisson solves using FFT. Our algorithm can reduce both the memory and CPU cost compared to usual ISDF-based calculations because:

1. We only need to perform ISDF calculations once, before the SCF iterations, instead of at each SCF iteration. We perform n_{atom} (the number of atoms) independent local ISDF calculations and each calculation is cheap. The cost of the local ISDF calculation scales linearly with the size of the system. After the ISDF fitting functions are obtained one has to construct the two-center Coulomb matrix $V(\xi, \xi')$ which scales quadratically with the size of the system, making the entire ISDF calculation quadratic in system size. In this algorithm, we do not need to use the randomized algorithm and the overall cost of ISDF is negligible.
2. The memory cost of our algorithm is largely independent of the size of the basis set and the overall memory requirement is quite low. This is because all of the diffuse−diffuse products, which represent the largest fraction of the non-negligible basis set pairs, are treated using an FFT grid that is independent of the basis set size. A key point is that the number of grid points needed to represent diffuse functions can be fairly small and thus the Poisson solves are cheap. Furthermore, for these pairs of basis functions, we do not store ISDF functions.

Because we use multiple grids of different resolutions, we have called our method "multigrid" inspired by the approach of the same name used to speed up the Coulomb matrix formation.[52−57] Although currently we only employ grids of two resolutions, one for sharp and one for diffuse functions, this approach is readily extended to include a larger number of grids. This can become important when all-electron calculations are performed without the use of Pseudopotentials, where a larger range of resolutions is needed due to the presence of extremely sharp basis functions. The approach here also has similarities to that of Füsti-Molnár and Pulay[63,64] in which the basis functions are partitioned into sharp and diffuse and different approaches are used to evaluate the contributions from various pairs. Recently, a similar approach has been used in the context of stochastic density functional theory where contributions of sharp functions are evaluated exactly while for diffuse functions stochastic resolution of identity is used.[47]

In the next section, we describe in more detail how the various grids are formed and how the entire exchange matrix is calculated.

COMPUTATIONAL DETAILS

To illustrate the algorithm let us imagine we have an atom, A, with uncontracted GTOs with exponents $\alpha_1, \alpha_2, \cdots$ in decreasing order of magnitude. A user-defined value, α_{min}, divides the GTOs into sharp functions that have exponents larger than α_{min} and diffuse functions with exponents smaller than α_{min}. Although the sharp functions require a high-resolution grid, the spatial

https://doi.org/10.1021/acs.jpca.4c02431
J. Phys. Chem. A 2024, 128, 7451−7461

extent of this grid is relatively small because the function decays rapidly in real space. The real space and Fourier space representation of an s-type function of exponent α are $\exp(-\alpha r^2)$ and $\exp(-G^2/(4\alpha))$ up to a multiplicative factor. If we want to represent the functions in real space up to an accuracy of ε_r then we have a local atom-centered grid of radius,

$$r_{\max} = \sqrt{\frac{-\ln(\varepsilon_r)}{\alpha_{\min}}} \tag{5}$$

All exponents greater than $\alpha_{\min}$ are supported by local grids with grid points $\mathbf{R}_A$, $\mathbf{R}_B$, ⋯ respectively, centered on atoms A, B, ⋯, respectively (see the upper panel of Figure 1). The remaining functions are represented on a sparse grid of lower resolution that spans the entire unit cell (U in the lower panel of Figure 1). The sparse universal grid is truncated in the Fourier space with a wavenumber $G_{U,\max}$ such that,

$$G_{U,\max} = \sqrt{-4\alpha \ln(\varepsilon_K)} \tag{6}$$

where ε_K is a threshold that one can decrease continuously to increase the overall accuracy of the calculation.

For all systems considered in this article, we have found that using $\varepsilon_r = 10^{-5}$ gives an overall error that will be below 50 μHa per atom. The optimal value of $\alpha_{\min}$ from a computational cost point of view can be system-dependent. There is a trade-off between the cost of the exchange evaluation and the memory requirement for storing the functions on the dense grid. A higher $\alpha_{\min}$ speeds up the ISDF calculation because fewer functions are considered sharp but it also increases the memory cost for storing $\mu(\mathbf{R}_U)$ because the universal grid U requires more points. We have found that an $\alpha_{\min}$ value of 2.8 Bohr^{-2} is a reasonable choice for the systems studied here.

Now one performs a local interpolative decomposition on each atom-centered grid A such that the equality,

$$\mu_A^L(\mathbf{R}_A)\nu_A^G(\mathbf{R}_A) \approx \sum_{\xi_A} \mu_A^L(\xi_A)\nu_A^G(\xi_A)\xi_A(\mathbf{R}_A) \tag{7}$$

is satisfied up to sufficient accuracy, determined by an ISDF threshold $\varepsilon_{\text{ISDF}}$. In the equation, the subscripts indicate the grid centered on atom A and the superscript L stands for local, indicating all sharp functions that are atom-A centered. ν_A^G are all functions (superscript G stands for global) that have a nonzero value on the grid around atom A, these include both the sharp and diffuse functions on atom A and also functions on other atoms. ξ_A are ISDF fitting points (functions) on the grid around A. At this point, it is also useful to define the set of functions μ_A^N that are present in the global list but are not in the set L. These nonlocal (N) functions have a nonzero value on at least one of the grid points $\mathbf{R}_A$ of local grid centered on A but are not one of the sharp functions centered on atom A.

For the dense atom-centered grids, the ISDF fitting is done using pivoted Cholesky without randomization because, unlike on a full grid, it is inexpensive to build the product density matrix $M(\mathbf{R}_A, \mathbf{R}_A')$ (eq 2) on local grids. After selecting the points using pivoted Cholesky we use a least-squares minimization (eq 3) to obtain the fitting functions $\xi_A(\mathbf{R}_A)$. These functions are fully supported on the local grids and are relatively inexpensive to store in memory. This local ISDF procedure is carried out for each local grid and the number of fitting functions chosen on each grid is controlled by the user-specified tolerance $\varepsilon_{\text{ISDF}}$. The cost of performing the local ISDF calculation for each atom-centered grid is system size-independent.

Once the fitting functions are formed, we calculate the two-centered Coulomb integral $V(\xi_A, \xi_B)$ for all pairs of atom-centered ISDF functions, and $V(\xi_A, \mathbf{R}_U)$ between ISDF functions on atom-centered grids and each grid point on the universal grid. The cost of constructing these matrices is equal to $O(N_{\xi,L}N_g \ln(N_g)) + O(N_{\xi,L}^2)$, where $N_{\xi,L}$ are the total number of ISDF functions from all atom-centered grids. The first term comes from having to perform an FFT for each local ISDF fitting function and the second term comes from evaluating the matrix $V(\xi_A, \xi_B)$. Notice that because each ISDF function is local the cost of this matrix evaluation is only quadratic as opposed to cubic with fully nonlocal ISDF functions. Finally, the calculation of the matrix $V(\xi_A, \mathbf{R}_U)$ only requires the potential due to the atom-centered fitting functions on the universal grids and does not require any matrix multiplications. One can readily evaluate the entire matrix $V(\mathbf{R}_U, \mathbf{R}_U')$ using a single FFT, using the fact that this is a circulant matrix. In our algorithm, we avoid storing this matrix and use FFT to calculate the exchange matrix using the iterative FFT algorithm described previously.

The analysis here shows that the entire cost of the ISDF calculation is quadratic in the system size which is lower than the cubic cost of forming the exchange matrix. Therefore, the overhead of performing these ISDF calculations is negligible. Table 2 summarizes the various thresholds used to construct the exchange matrix.

Table 2. Accuracy of the Exchange Calculation is Determined by Four Different Thresholds As Outlined Above[a]

symbol	meaning
$\alpha_{\min}$	Gaussians with exponents greater than $\alpha_{\min}$ are sharp and others are diffuse
ε_r	Threshold that determines the extent of the local grid (see eq 5 and upper panel of Figure 1)
$\varepsilon_{\text{ISDF}}$	Threshold used during pivoted-Cholesky decomposition that determines the number of ISDF functions
ε_K	Determines the number of grid points that make up the universal sparse grid (see eq 6)

[a]For all the calculations in the paper we fix ε_r to be 10^{-5} and $\alpha_{\min}$ is chosen to be 2.8 Bohr^{-2}. The parameters ε_K and $\varepsilon_{\text{ISDF}}$ are varied to obtain the desired accuracy.

Building the Exchange Matrix. In the multigrid approach, the expressions of the two-electron integrals have the THC form, however, there are now four terms,

$$\begin{aligned}(\mu\nu|\lambda\sigma) =& \sum_{\xi\xi'} (\mu|\xi)^L(\nu|\xi)^G V(\xi,\xi')(\lambda|\xi')^G(\sigma|\xi')^L \\ &+ \sum_{\xi\xi'} (\mu|\xi)^N(\nu|\xi)^L V(\xi,\xi')(\lambda|\xi')^L(\sigma|\xi')^N \\ &\times \sum_{\xi\xi'} (\mu|\xi)^L(\nu|\xi)^G V(\xi,\xi')(\lambda|\xi')^L(\sigma|\xi')^N \\ &+ \sum_{\xi\xi'} (\mu|\xi)^N(\nu|\xi)^L V(\xi,\xi')(\lambda|\xi')^G(\sigma|\xi')^L\end{aligned} \tag{8}$$

The various matrices with superscripts L, N, and G are described in Figure 2. The four terms arise to ensure that products of basis functions are only evaluated on the appropriate grid. For example, there are diffuse functions that are nonzero on local grids of various atoms and also on the diffuse grid. The equation ensures that the product of diffuse functions is only evaluated on the most sparse grid; on the sharp grids, only products of sharp−sharp and sharp-diffuse are evaluated.

https://doi.org/10.1021/acs.jpca.4c02431
J. Phys. Chem. A 2024, 128, 7451−7461

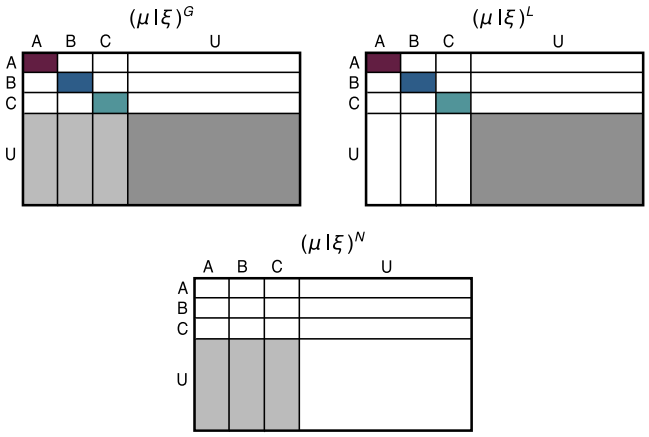


Figure 2. It is useful to define the set of functions μ_A^N that are present in the global list but are not in the local set L. These nonlocal functions (N) have a nonzero value on at least one of the grid points $\mathbf{R}_A$ of local grid centered on A but are not one of the sharp functions centered on atom A. Graphical representations of the matrices $(\mu|\xi)^G$, $(\mu|\xi)^L$, and $(\mu|\xi)^N$, where the superscripts refer to the sets of global (G), local (L), and nonlocal (N) functions are shown. The colors here follow from Figure 1 and represent sets of atom-centered GTOs that are local to atom-centered grids. The white blocks correspond to zeros in the matrix and the shaded blocks represent nonzero entries. The columns of the matrices correspond to ISDF functions for atom-centered grids (A, B, C) or all the grid points on the universal grid U.

We use occ-RI and only the occupied-virtual part of the exchange matrix is constructed using the two-electron integrals in eq 8. We would like to reemphasize that in eq 8 one needs access to the two-electron integrals $V(\mathbf{R}_U, \mathbf{R}_{U'})$ because all points on the sparse universal grids are included as ISDF points. However, this matrix is never stored and we rely on the fact that this matrix is diagonal in the Fourier space and the action of this matrix on any function is evaluated using FFTs.

It is worth pointing out that one can potentially speed up the calculations by utilizing the block structure of the various matrices. However, in our current work, the entire code is implemented in Python and we find that implementing block multiplication by using for loops incurs an overhead that nullifies any benefit of reducing the computations. In a future publication, this can be remedied by implementing some of these matrix multiplications in an optimized C-code.

RESULTS AND DISCUSSION

The algorithm described above was implemented in Python, with a small portion that allows for the direct calculation of atomic orbitals in C. The one-electron integrals comprising the core Hamiltonian are calculated using a multigrid branch of PySCF.[65] We benchmark the performance of the multigrid ISDF method with occ-RI using two systems: diamond, with a conventional unit cell containing eight carbon atoms, and lithium hydride, with a conventional unit cell containing four lithium and four hydrogen atoms. For all calculations the Goedecker-Teter-Hutter (GTH) pseudopotentials[66−68] and uncontracted GTH−CC-XZVP basis sets[69] of Ye are used throughout. We use a Kinetic energy cutoff of the plane waves in FFT of 70 E_h for diamond and 130 E_h for lithium hydride to ensure that the error from the finite plane-wave cutoff is less than 5 μHa per atom. For all calculations shown in this section, we fix $\epsilon_r = 10^{-5}$ and $\alpha_{min} = 2.8$ Bohr^{-2}.

Accuracy of Multigrid ISDF. In multigrid ISDF there are two types of fitting functions, the first set is local and is supported on a dense grid (their number is denoted by $N_{\xi,1}$). These functions are obtained through ISDF. The memory and CPU cost of using these functions increases quadratically with $N_{\xi,1}$. The second type of ISDF functions are Sinc functions uniformly placed throughout the unit cell (their number is denoted by $N_{\xi,2}$). The memory and CPU cost of the calculation increases only linearly with the number of these functions.

In Figure 3 we show how the number of local ISDF functions changes as one reduces the ϵ_{ISDF} threshold and the accompanying reduction in the error of the calculation. As shown, the number of local ISDF functions needed is relatively small and the number of these functions does not change significantly with the size of the basis set. The error stops decreasing exponentially because we have kept ϵ_K fixed which fixes the number of ISDF functions on the diffuse grid.

The number of uniform ISDF functions ($N_{\xi,2}$) depends on two settings: the largest GTO exponent that is less than the threshold α_{min} and the threshold ϵ_K. These functions constitute the majority of the ISDF fitting functions and often far exceed the number of ISDF functions on the dense grid. $N_{\xi,2}$ can vary with the system and basis set. For example, in the TZ basis set of lithium hydride there are no exponents between 3.1 and 1.4 (see Table 3), thus with the α_{min} of 2.8 we find that the $N_{\xi,2}$ for the TZ basis set is smaller than that for the DZ basis set. Table 3 gives detailed information on the number of basis functions and ISDF

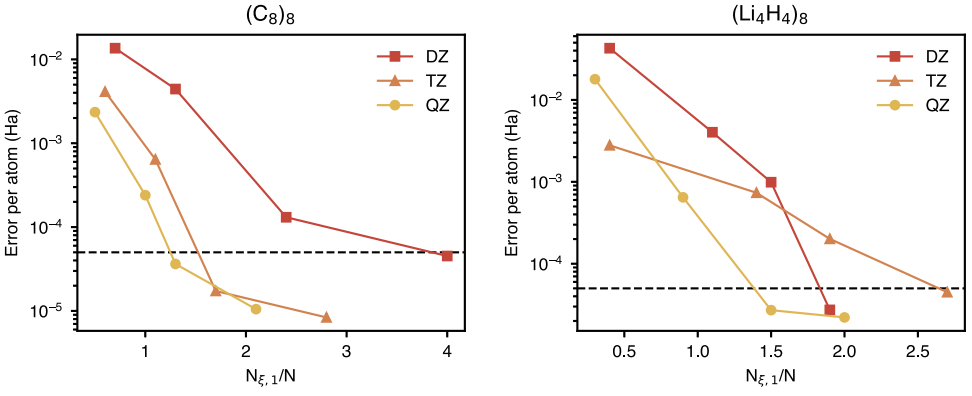


Figure 3. Error per atom incurred from the ISDF approximation for diamond (C_8) and lithium hydride (Li_4H_4) supercells with an increasing number of local fitting functions for fixed ϵ_K of 10^{-2} and 10^{-3} for diamond and lithium hydride, respectively. The four points on each curve are obtained by using $\epsilon_{ISDF} = 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}$.

https://doi.org/10.1021/acs.jpca.4c02431
J. Phys. Chem. A 2024, 128, 7451−7461

Table 3. Basis Functions and ISDF Fitting Functions for 2 × 2 × 2 Supercell of Diamond $(C_8)_8$ and Lithium Hydride $(Li_4H_4)_8$[a]

basis	$N_{\xi,1}$	$N_{\xi,2}$	α_1	α_2	N	N_{sharp}
Diamond $(C_8)_8$						
DZ	3240	17,576	4.3−4.3	1.3−0.1	1344	256
TZ	4128	27,000	5.4−5.4	2.0−0.1	2368	256
QZ	5256	39,304	6.2−6.2	2.6−0.1	3968	256
Lithium hydride $(Li_4H_4)_8$						
DZ	1728	74,088	8.4−7.3	2.1−0.1	896	160
TZ	4568	39,304	10.9−3.1	1.4−0.1	1696	320
QZ	6120	74,088	12.5−4.5	2.3−0.1	3136	320

[a] $N_{\xi,1}$, the number of ISDF functions on the dense grid; $N_{\xi,2}$, the number of uniformly placed Sinc functions on the diffuse grid; α_1, the range of exponents local to the dense grid; α_2, the range of exponents supported by the diffuse grid; N, the total number of basis functions; and N_{sharp}, the number of sharp basis functions.

functions supported on the dense and sparse grids, along with the range of exponents for the basis functions on each grid.

Next, we show that once the various thresholds are selected for a unit cell, the error per atom does not increase with the size of the system. Figure 4 shows the error per atom with increasing N_ξ/N for the diamond conventional unit cell C_8 and supercells of increasing sizes: $(C_8)_2$, $(C_8)_4$, and $(C_8)_8$; it also shows the lithium hydride conventional unit cell Li_4H_4 and supercells of increasing sizes: $(Li_4H_4)_2$, $(Li_4H_4)_4$, and $(Li_4H_4)_8$. Here the uncontracted GTH−CC-TZVP basis is used with $\varepsilon_K = 10^{-2}$ and 10^{-3} for diamond and lithium hydride, respectively, which determine the number of grid points in the sparse universal grid. Three different ISDF thresholds are used, $\varepsilon_{ISDF} = 10^{-2}$, 10^{-3}, 10^{-4}, which increase the number of ISDF fitting functions used on the dense grids. One can see that with decreasing thresholds the errors in the calculations exponentially decrease. The number of fitting functions N_ξ includes the ISDF functions from the atom-centered grids and also all the grid points in the universal sparse grid.

The N_ξ/N in these calculations is larger than in conventional AO-based ISDF by almost a factor of 2. It is worth remembering that the ISDF functions on the sparse grid most greatly contribute to N_ξ/N. For these functions, we do not store or calculate the two-center Coulomb integrals and thus, the overall CPU and memory cost is significantly lower than AO-based ISDF as we show in the next subsection.

Cost of the Multigrid Calculations. Figure 5 compares the wall times of single Coulomb and exchange builds and the ISDF wall time that includes the time to find ISDF functions ξ and build the matrix $V(\xi, \xi')$(eq 1). The ISDF times here are weighted by a factor of 1/7. A major hurdle overcome by the multigrid method is the cost of ISDF. The wall times from the multigrid ISDF are one to 2 orders of magnitude faster than the single-grid THC and rPS methods (with occ-RI) from ref 40. In the multigrid ISDF method, this is reduced because only small subsets of product densities are fit on regions of the most dense grids (eq 2). The most significant efficiency gain comes from not fitting the ISDF functions on universal sparse grid and calculating the matrix $V(\xi, \xi')$ (eq 1) in a direct fashion during the exchange build.

The memory requirement for the largest arrays is significantly reduced in the multigrid builds. For the multigrid method, the cost of 2-center integrals $V(\xi_A, \xi_B)$, $V(\xi_A, R_U)$ for the 2 × 2 × 2 diamond TZ data shown here is about 1 GB (note that we do not store the 2-center integrals $V(R_U, R_U')$ in memory), and the cost of storing the orbital matrices $(\mu|\xi)$ require less than 1 GB. For the LiH data shown here, the total memory required by the MG method is less than 2 GB. For diamond, the THC method requires about 53 GB and rPS 28 GB. For 1 × 2 × 2 LiH supercell, THC and rPS require about 54 and 36 GB, respectively. Compared to the single grid methods, the MG method requires up to about 35 times less memory for the data here. The smaller system is used for LiH as the 2 × 2 × 2 supercell required more than 128 GB of memory when using the THC method. The scaling and memory requirements for TZ basis set is shown in Figure 6.

Next, we calculate the scaling of the various steps with increasing system size. Figure 7 shows the cost of single Coulomb and exchange builds, the total ISDF times (as in Figure 5 are weighted by 1/7), and the wall time per SCF iteration for diagonalizing the Fock matrix for diamond supercells of up to 80 unit cell copies, depending on the basis, and lithium hydride supercells of up to 48 copies. Calculations using GTH−CC-DZVP (DZ, left), GTH−CC-TZVP (TZ, middle), and GTH−CC-QZVP (QZ, right) are shown. The thresholds used are the same as reported in Figure 5. The following conclusions can be drawn from the graphs:

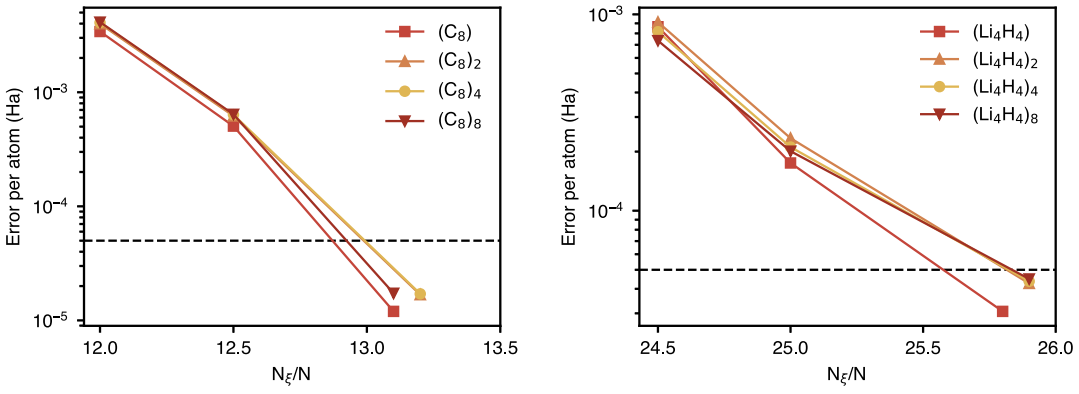


Figure 4. Results in the two graphs are obtained using the TZ basis set. N_ξ in the graph includes ISDF basis functions from both the dense and sparse grids. The number of ISDF functions on the sparse grid was $N_{\xi,2}/N \approx 11$ for diamond, and 23 for lithium hydride. The accuracy is independent of system size so smaller systems may be used when choosing an ISDF threshold. The data here can be reproduced with the following settings: $\varepsilon_K = 10^{-2}$ and $\varepsilon_{ISDF} = 10^{-2}$, 10^{-3}, and 10^{-4} for diamond, and $\varepsilon_K = 10^{-3}$ and $\varepsilon_{ISDF} = 10^{-3}$, 10^{-4}, and 10^{-5} for lithium hydride.

https://doi.org/10.1021/acs.jpca.4c02431
J. Phys. Chem. A 2024, 128, 7451−7461

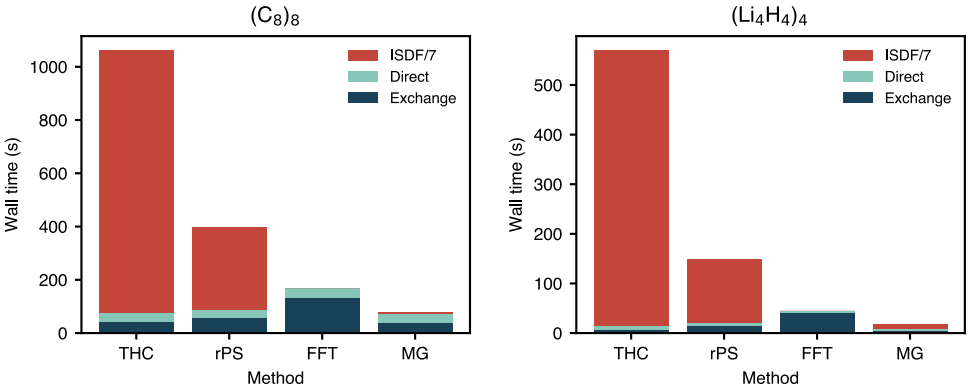


Figure 5. Wall times for a single exchange build using the AO-based ISDF (THC), robust pseudospectral method (rPS), single-grid with no ISDF (FFT), and multigrid ISDF (MG) methods using TZ basis set on a *single core* of Intel(R) Xeon(R) CPU E5−2680 v3 @ 2.50 GHz processor. The wall time for the full ISDF procedure (red) is divided by the number of iterations to evaluate the per-iteration cost. The systems reported are a $2 \times 2 \times 2$ supercell of C_8 (diamond) and a $1 \times 2 \times 2$ supercell of Li_4H_4 (lithium hydride). For an accuracy of ≈50 μHa/atom, we used $\varepsilon_K = 10^{-2}$ and $\varepsilon_{ISDF} = 10^{-4}$ for diamond and $\varepsilon_K = 10^{-3}$ and $\varepsilon_{ISDF} = 10^{-5}$ for Li_4H_4 in the multigrid calculations. The ISDF parameters $N_\xi/N = 13$ and 6 for Li_4H_4 and 7 and 4 for diamond were used for the THC and rPS methods, respectively, to get a similar accuracy. For the MG method, the wall times for building the exchange (dark blue) and Coulomb (Direct, light blue) matrices are nearly equivalent. They are also comparable to the times for the THC and rPS methods. The ISDF wall time, however, is one to 2 orders of magnitude faster using the MG method.

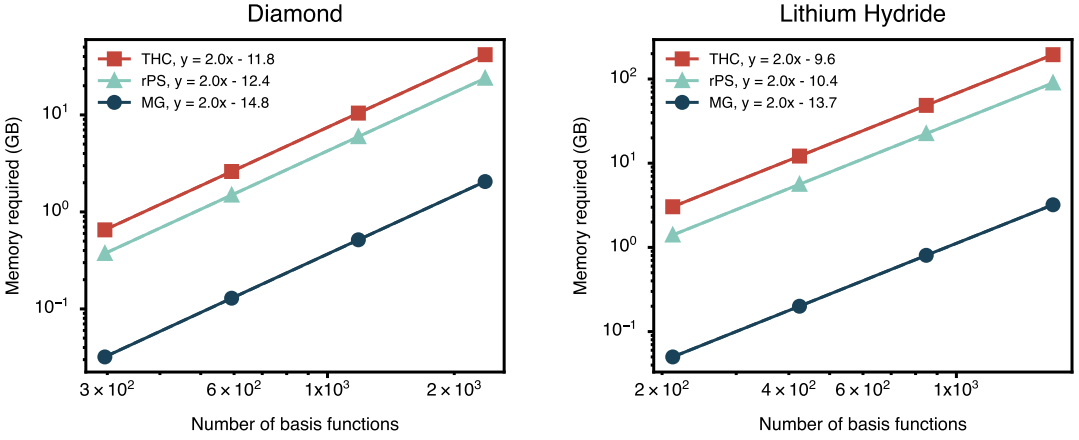


Figure 6. Estimated memory required for THC (red), rPS (light blue) and MG (dark blue) are shown in the figure for the Diamond and LiH systems with a TZ basis set with increasing supercell size ($1 \times 1 \times 1$, $1 \times 1 \times 2$, $1 \times 2 \times 2$, and $2 \times 2 \times 2$). All three methods show that the memory requirements scale quadratically with the size of the system, however, MG requires between one to 2 orders of magnitude smaller memory. We refer the reader to the main text for more detailed discussion of the dominant memory cost in these calculations.

1. The multigrid ISDF scales quadratically N^2, which is due to the local nature of the fitting. For errors under 50 μHa, we used an ISDF threshold ε_{ISDF} of 10^{-4} for diamond and 10^{-5} for lithium hydride which corresponds to N_ξ/N of 9 to 84, depending on the system and basis used. The majority of these fitting functions are Sinc functions on the sparse grid. A large value of 84 is sometimes needed for smaller basis set such as DZ because the number of Sinc functions can be quite large relative to the size of the basis set.
2. The scaling of the exchange matrix is cubic with the system size. The cost of the calculation for the largest basis set is only about a factor of 4 more expensive than the cost of performing diagonalization.
3. The cost of Coulomb is nearly linear with the size of the system. There is some deviation from linearity because in these calculations we only use 2−4 Coulomb grids. By adding more grids, linear scaling is possible; doing so comes with additional CPU cost but no additional memory requirement.
4. The cost of the exchange is cheaper than the cost of Coulomb calculations for fairly large systems containing up to 3000 basis functions in the case of diamond with the QZ basis set and for around 1000 basis functions in the case of lithium hydride with the QZ basis set.
5. The largest calculation, in terms of the number of basis functions, was a $3 \times 3 \times 4$ diamond supercell with the QZ basis set containing 17,856 basis functions and 1,152 electrons. Because we use tighter thresholds, both ε_{ISDF} and ε_K, for lithium hydride the largest calculation for it was a $3 \times 3 \times 3$ supercell containing 431 electrons and 10,584 basis functions; this was also with the QZ basis. The number of electrons per unit cell are also smaller in LiH than in Diamond.

CONCLUSIONS

In this paper, we have shown that efficient calculations of exchange matrices can be performed using our multigrid ISDF algorithm. The multigrid ISDF is significantly more efficient than the usual ISDF algorithm both in terms of memory and

https://doi.org/10.1021/acs.jpca.4c02431
J. Phys. Chem. A 2024, 128, 7451−7461

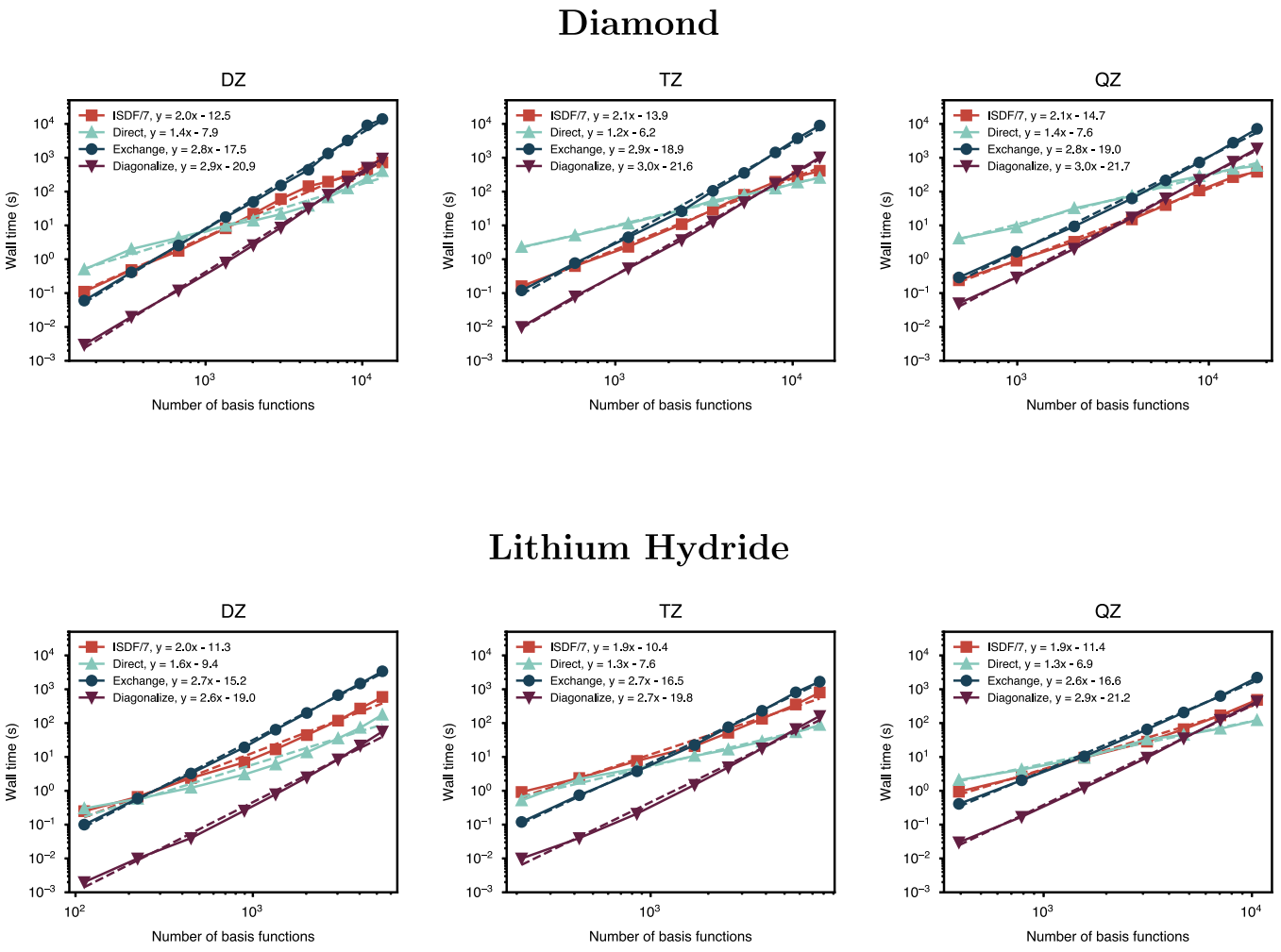


Figure 7. Wall times for performing ISDF (weighted by 1/7, dark blue), a single build of the Coulomb (Direct, light blue) and exchange (purple) matrices, and diagonalizing the Fock matrix (red) on a *single core* of Intel(R) Xeon(R) CPU E5−2680 v3 @ 2.50 GHz processor. The calculations were performed using the uncontracted GTH−CC-DZVP (DZ, left), GTH−CC-TZVP (TZ, middle), and GTH−CC-QZVP (QZ, right) basis sets. The systems used are diamond supercells (top) of 1, 4, 8, 12, 18, 27, 36, 48, 64, and 80 unit cell copies (DZ), 1, 4, 8, 12, 18, 27, 36, and 48 copies (TZ), and 1, 4, 8, 12, 18, 27, and 36 copies (QZ), and lithium hydride supercells (bottom) of 1, 4, 8, 12, 18, 27, 36, and 48 unit cell copies (DZ), 1, 4, 8, 12, 18, 27, and 36 copies (TZ), and 1, 4, 8, 12, 18, and 27 copies (QZ). The settings used here are the same as in Figure 5.

CPU cost. With this algorithm, relatively large calculations (containing >17,000 basis functions) can be performed without running out of memory on a single node. With this technique, the exchange calculation is more efficient than Coulomb calculations for systems with up to about 1000 basis functions for TZ and QZ basis sets in diamond and lithium hydride systems. For large basis sets, such as QZ, the cost of exchange evaluation is only a factor of 4 more expensive than the cost of diagonalization and because the scaling of the two steps is similar we expect that this result will hold for even larger systems.

In our current implementation, we have not parallelized the calculations, although it should be possible to do so with high efficiency because most of the operations involve a series of FFTs that can be embarrassingly parallelized, or matrix multiplications that are also amenable to parallelization. We have also not made use of linear scaling approaches to leverage the fact that the exchange matrix is near-sighted, which can be used to further improve the efficiency of the calculations, particularly for systems with large band gaps. The approach we have outlined here has many extensions that we are actively exploring, including the ability to perform all-electron calculations, use of mixed Gaussian−plane-wave basis set, use for molecular calculations, the calculation of nuclear gradients, and also use of k-point symmetry. It should be pointed out that the use of ISDF allows one to obtain a nearly linear cost ($O(N_k \log(N_k))$) with the number of k-points N_k, however the order of contraction used to obtain this scaling is different than the one we have used in our current paper. We are currently working toward addressing this issue in a future publication. Given that the cost of the exchange evaluation is cheaper than Coulomb for systems containing up to 1000 basis functions and around 100 electrons (these numbers are larger for diamond with the QZ basis set), one can likely perform hybrid DFT calculations with a similar cost as pure DFT calculations if the unit cells contain 100 or fewer electrons (because the scaling with N_k is similar for both Coulomb and exchange).

AUTHOR INFORMATION

Corresponding Author

Sandeep Sharma − *Department of Chemistry, University of Colorado, Boulder, Colorado 80302, United States;* orcid.org/0000-0002-6598-8887; Email: sanshar@gmail.com

https://doi.org/10.1021/acs.jpca.4c02431
J. Phys. Chem. A 2024, 128, 7451−7461

Authors

Kori E. Smyser − *Department of Chemistry, University of Colorado, Boulder, Colorado 80302, United States;* orcid.org/0000-0002-3697-0717

Alec White − *Quantum Simulation Technologies, Inc., Boston, Massachusetts 02135, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpca.4c02431

Author Contributions

S.S. and K.S. performed research, implemented code, and ran calculations. All authors contributed to data analysis and writing the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

K.E.S. was supported through the National Science Foundation grant CHE-2145209 and S.S. through a grant from the Camille and Henry Dreyfus Foundation. This work utilized resources from the University of Colorado Boulder Research Computing Group, which is supported by the National Science Foundation(awards ACI1532235 and ACI-1532236), the University of Colorado Boulder, and Colorado State University.

■ REFERENCES

(1) Kohn, W.; Sham, L. J. Self-consistent equations including exchange and correlation effects. *Physical review* **1965**, *140*, A1133.

(2) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(3) Mardirossian, N.; Head-Gordon, M. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. *Molecular physics* **2017**, *115*, 2315−2372.

(4) Heyd, J.; Peralta, J. E.; Scuseria, G. E.; Martin, R. L. Energy band gaps and lattice parameters evaluated with the Heyd-Scuseria-Ernzerhof screened hybrid functional. *J. Chem. Phys.* **2005**, *123*, 174101.

(5) Paier, J.; Marsman, M.; Hummer, K.; Kresse, G.; Gerber, I. C.; Ángyán, J. G. Screened hybrid density functionals applied to solids. *J. Chem. Phys.* **2006**, *124*, 154709.

(6) Finazzi, E.; Di Valentin, C.; Pacchioni, G.; Selloni, A. Excess electron states in reduced bulk anatase TiO2: Comparison of standard GGA, GGA+U, and hybrid DFT calculations. *J. Chem. Phys.* **2008**, *129*, 154113.

(7) Hai, X.; Tahir-Kheli, J.; Goddard, W. A. Accurate band gaps for semiconductors from density functional theory. *J. Phys. Chem. Lett.* **2011**, *2*, 212−217.

(8) Basera, P.; Saini, S.; Arora, E.; Singh, A.; Kumar, M.; Bhattacharya, S. Stability of non-metal dopants to tune the photo-absorption of TiO2 at realistic temperatures and oxygen partial pressures: A hybrid DFT study. *Sci. Rep.* **2019**, *9*, 11427.

(9) Kovacic, Z.; Likozar, B.; Hus, M. Photocatalytic CO2 reduction: A review of ab initio mechanism, kinetics, and multiscale modeling simulations. *ACS catalysis* **2020**, *10*, 14984−15007.

(10) Almlöf, J.; Faegri, K.; Korsell, K. Principles for a direct SCF approach to LICAO-MO ab-initio calculations. *J. Comput. Chem.* **1982**, *3*, 385−399.

(11) White, C. A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M. The continuous fast multipole method. *Chem. Phys. Lett.* **1994**, *230*, 8−16.

(12) Challacombe, M.; Schwegler, E.; Almlöf, J. Fast assembly of the Coulomb matrix: A quantum chemical tree code. *J. Chem. Phys.* **1996**, *104*, 4685−4698.

(13) Kudin, K. N.; Scuseria, G. E. A fast multipole method for periodic systems with arbitrary unit cell geometries. *Chem. Phys. Lett.* **1998**, *283*, 61−68.

(14) Challacombe, M.; Schwegler, E. Linear scaling computation of the Fock matrix. *J. Chem. Phys.* **1997**, *106*, 5526.

(15) Ochsenfeld, C.; White, C. a.; Head-Gordon, M. Linear and sublinear scaling formation of Hartree-Fock-type exchange matrices. *J. Chem. Phys.* **1998**, *109*, 1663.

(16) Goedecker, S. Linear scaling electronic structure methods. *Rev. Mod. Phys.* **1999**, *71*, 1085−1123.

(17) Ko, H.-Y.; Jia, J.; Santra, B.; Wu, X.; Car, R.; DiStasio, R. A., Jr. Enabling Large-Scale Condensed-Phase Hybrid Density Functional Theory Based Ab Initio Molecular Dynamics. 1. Theory, Algorithm, and Performance. *J. Chem. Theory Comput.* **2020**, *16*, 3757−3785.

(18) Sierka, M.; Hogekamp, A.; Ahlrichs, R. Fast evaluation of the Coulomb potential for electron densities using multipole accelerated resolution of identity approximation. *J. Chem. Phys.* **2003**, *118*, 9136.

(19) Sodt, A.; Subotnik, J. E.; Head-Gordon, M. Linear scaling density fitting. *J. Chem. Phys.* **2006**, *125*, 194109.

(20) Polly, R.; Werner, H. J.; Manby, F. R.; Knowles, P. J. Fast Hartree-Fock theory using local density fitting approximations. *Mol. Phys.* **2004**, *102*, 2311−2321.

(21) Sodt, A.; Head-Gordon, M. Hartree-Fock exchange computed using the atomic resolution of the identity approximation. *J. Chem. Phys.* **2008**, *128*, 104106.

(22) Manzer, S. F.; Epifanovsky, E.; Head-Gordon, M. Efficient implementation of the pair atomic resolution of the identity approximation for exact exchange for hybrid and range-separated density functionals. *J. Chem. Theory Comput.* **2015**, *11*, 518−527.

(23) Manzer, S.; Horn, P. R.; Mardirossian, N.; Head-Gordon, M. Fast, accurate evaluation of exact exchange: The occ-RI-K algorithm. *J. Chem. Phys.* **2015**, *143*, No. 024113.

(24) Dunlap, B. I. Robust variational fitting: Gaspar's variational exchange can accurately be treated analytically. *Journal of Molecular Structure: THEOCHEM* **2000**, *501*−*502*, 221−228.

(25) Dunlap, B. I. Robust and variational fitting: Removing the four-center integrals from center stage in quantum chemistry. *Journal of Molecular Structure: THEOCHEM* **2000**, *529*, 37−40.

(26) Hollman, D. S.; Schaefer, H. F.; Valeev, E. F. Fast construction of the exchange operator in an atom-centred basis with concentric atomic density fitting. *Mol. Phys.* **2017**, *115*, 2065−2076.

(27) Ihrig, A. C.; Wieferink, J.; Zhang, I. Y.; Ropo, M.; Ren, X.; Rinke, P.; Scheffler, M.; Blum, V. Accurate localized resolution of identity approach for linear-scaling hybrid density functionals and for many-body perturbation theory. *New J. Phys.* **2015**, *17*, No. 093020.

(28) Kokott, S.; Merz, F.; Yao, Y.; Carbogno, C.; Rossi, M.; Havu, V.; Rampp, M.; Scheffler, M.; Blum, V. Efficient All-electron Hybrid Density Functionals for Atomistic Simulations Beyond 10,000 Atoms. *arXiv preprint arXiv:2403.10343* **2024**.

(29) Beebe, N.; Linderberg, J. Simplifications in the generation and transformation of two-electron integrals in molecular calculations. *Int. J. Quantum Chem.* **1977**, *12*, 683−705.

(30) Aquilante, F.; Lindh, R.; Bondo Pedersen, T. Unbiased auxiliary basis sets for accurate two-electron integral approximations. *J. Chem. Phys.* **2007**, *127*, 114107.

(31) Friesner, R. A. Solution of self-consistent field electronic structure equations by a pseudospectral method. *Chem. Phys. Lett.* **1985**, *116*, 39−43.

(32) Neese, F.; Wennmohs, F.; Hansen, A.; Becker, U. Efficient, approximate and parallel Hartree-Fock and hybrid DFT calculations. A 'chain-of-spheres' algorithm for the Hartree-Fock exchange. *Chem. Phys.* **2009**, *356*, 98−109.

(33) Hohenstein, E. G.; Parrish, R. M.; Sherrill, C. D.; Martínez, T. J. Communication: Tensor hypercontraction. III. Least-squares tensor hypercontraction for the determination of correlated wavefunctions. *J. Chem. Phys.* **2012**, *137*, 221101.

(34) Parrish, R. M.; Hohenstein, E. G.; Martínez, T. J.; Sherrill, C. D. Tensor hypercontraction. II. Least-squares renormalization. *J. Chem. Phys.* **2012**, *137*, 224106.

(35) Hohenstein, E. G.; Parrish, R. M.; Martínez, T. J. Tensor hypercontraction density fitting. I. Quartic scaling second- and third-

order Møller-Plesset perturbation theory. *J. Chem. Phys.* **2012**, *137*, No. 044103.
(36) Lu, J.; Ying, L. Compression of the electron repulsion integral tensor in tensor hypercontraction format with cubic scaling cost. *J. Comput. Phys.* **2015**, *302*, 329−335.
(37) Hu, W.; Lin, L.; Yang, C. Interpolative Separable Density Fitting Decomposition for Accelerating Hybrid Density Functional Calculations with Applications to Defects in Silicon. *J. Chem. Theory Comput.* **2017**, *13*, 5420−5431.
(38) Dong, K.; Hu, W.; Lin, L. Interpolative Separable Density Fitting through Centroidal Voronoi Tessellation with Applications to Hybrid Functional Electronic Structure Calculations. *J. Chem. Theory Comput.* **2018**, *14*, 1311−1320.
(39) Lee, J.; Lin, L.; Head-Gordon, M. Systematically Improvable Tensor Hypercontraction: Interpolative Separable Density-Fitting for Molecules Applied to Exact Exchange, Second- A nd Third-Order Møller-Plesset Perturbation Theory. *J. Chem. Theory Comput.* **2020**, *16*, 243−263.
(40) Sharma, S.; White, A. F.; Beylkin, G. Fast Exchange with Gaussian Basis Set Using Robust Pseudospectral Method. *J. Chem. Theory Comput.* **2022**, *18*, 7306−7320.
(41) Rettig, A.; Lee, J.; Head-Gordon, M. Even Faster Exact Exchange for Solids via Tensor Hypercontraction. *J. Chem. Theory Comput.* **2023**, *19*, 5773−5784.
(42) Zhang, Z.; Yin, X.; Hu, W.; Yang, J. Machine Learning K-Means Clustering of Interpolative Separable Density Fitting Algorithm for Accurate and Efficient Cubic-Scaling Exact Exchange Plus Random Phase Approximation within Plane Waves. *J. Chem. Theory Comput.* **2024**, *20*, 1944−1961.
(43) Bowler, D. R.; Miyazaki, T. O(N) methods in electronic structure calculations. *Rep. Prog. Phys.* **2012**, *75*, No. 036503.
(44) Wu, X.; Selloni, A.; Car, R. Order- N implementation of exact exchange in extended insulating systems. *Physical Review B - Condensed Matter and Materials Physics* **2009**, *79*, No. 085102.
(45) Baer, R.; Neuhauser, D.; Rabani, E. Self-averaging stochastic kohn-sham density-functional theory. *Phys. Rev. Lett.* **2013**, *111*, No. 106402.
(46) Neuhauser, D.; Rabani, E.; Cytter, Y.; Baer, R. Stochastic Optimally Tuned Range-Separated Hybrid Density Functional Theory. *J. Phys. Chem. A* **2016**, *120*, 3071−3078.
(47) Bradbury, N. C.; Allen, T.; Nguyen, M.; Neuhauser, D. Deterministic/Fragmented-StocLhastic Exchange for Large-Scale Hybrid DFT Calculations. *J. Chem. Theory Comput.* **2023**, *19*, 9239−9247.
(48) Todorova, T.; Seitsonen, A. P.; Hutter, J.; Kuo, I.-F. W.; Mundy, C. J. Molecular dynamics simulation of liquid water: Hybrid density functionals. *J. Phys. Chem. B* **2006**, *110*, 3685−3691.
(49) Lin, L. Adaptively Compressed Exchange Operator. *J. Chem. Theory Comput.* **2016**, *12*, 2242−2249.
(50) Guidon, M.; Hutter, J.; VandeVondele, J. Auxiliary Density Matrix Methods for Hartree-Fock Exchange Calculations. *J. Chem. Theory Comput.* **2010**, *6*, 2348−2364.
(51) Qin, X.; Liu, J.; Hu, W.; Yang, J. Interpolative separable density fitting decomposition for accelerating Hartree−Fock exchange calculations within numerical atomic orbitals. *J. Phys. Chem. A* **2020**, *124*, 5664−5674.
(52) Lippert, G.; Hutter, J.; Parrinello, M. The Gaussian and augmented-plane-wave density functional method for ab initio molecular dynamics simulations. *Theor. Chem. Acc.* **1999**, *103*, 124−140.
(53) Vandevondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Commun.* **2005**, *167*, 103−128.
(54) Laino, T.; Mohamed, F.; Laio, A.; Parrinello, M. An efficient real space multigrid QM/MM electrostatic coupling. *J. Chem. Theory Comput.* **2005**, *1*, 1176−1184.
(55) Beck, T. L. 5 Real-Space and Multigrid Methods in Computational Chemistry. *Reviews in Computational Chemistry* **2008**, *26*, 223.
(56) Del Ben, M.; Hutter, J.; VandeVondele, J. Second-order Møller−Plesset perturbation theory in the condensed phase: An efficient and massively parallel Gaussian and plane waves approach. *J. Chem. Theory Comput.* **2012**, *8*, 4177−4188.
(57) Kühne, T. D.; Iannuzzi, M.; Del Ben, M.; Rybkin, V. V.; Seewald, P.; Stein, F.; Laino, T.; Khaliullin, R. Z.; Schütt, O.; Schiffmann, F.; Golze, D.; Wilhelm, J.; Chulkov, S.; Bani-Hashemian, M. H.; Weber, V.; Borštnik, U.; Taillefumier, M.; Jakobovits, A. S.; Lazzaro, A.; Pabst, H.; Müller, T.; Schade, R.; Guidon, M.; Andermatt, S.; Holmberg, N.; Schenter, G. K.; Hehn, A.; Bussy, A.; Belleflamme, F.; Tabacchi, G.; Glöß, A.; Lass, M.; Bethune, I.; Mundy, C. J.; Plessl, C.; Watkins, M.; VandeVondele, J.; Krack, M.; Hutter, J.; et al. CP2K: An electronic structure and molecular dynamics software package -Quickstep: Efficient and accurate electronic structure calculations. *J. Chem. Phys.* **2020**, *152*, 194103.
(58) Merlot, P.; Kjærgaard, T.; Helgaker, T.; Lindh, R.; Aquilante, F.; Reine, S.; Pedersen, T. B. Attractive electron−electron interactions within robust local fitting approximations. *J. Comput. Chem.* **2013**, *34*, 1486−1496.
(59) Wirz, L. N.; Reine, S. S.; Pedersen, T. B. On Resolution-of-the-Identity Electron Repulsion Integral Approximations and Variational Stability. *J. Chem. Theory Comput.* **2017**, *13*, 4897−4906.
(60) Halko, N.; Martinsson, P. G.; Tropp, J. A. Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions. *SIAM Review* **2011**, *53*, 217−288.
(61) Liberty, E.; Woolfe, F.; Martinsson, P.-G.; Rokhlin, V.; Tygert, M. Randomized algorithms for the low-rank approximation of matrices. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 20167−20172.
(62) Matthews, D. A. Improved Grid Optimization and Fitting in Least Squares Tensor Hypercontraction. *J. Chem. Theory Comput.* **2020**, *16*, 1382−1385.
(63) Füsti-Molnár, L.; Pulay, P. Accurate molecular integrals and energies using combined plane wave and Gaussian basis sets in molecular electronic structure theory. *J. Chem. Phys.* **2002**, *116*, 7795−7805.
(64) Füsti-Molnár, L.; Pulay, P. The Fourier transform Coulomb method: Efficient and accurate calculation of the Coulomb operator in a Gaussian basis. *J. Chem. Phys.* **2002**, *117*, 7827−7835.
(65) Sun, Q.; Berkelbach, T. C.; Blunt, N. S.; Booth, G. H.; Guo, S.; Li, Z.; Liu, J.; McClain, J. D.; Sayfutyarova, E. R.; Sharma, S.; Wouters, S.; Chan, G. K.; et al. PySCF: the Python-based simulations of chemistry framework. *WIREs Comput. Mol. Sci.* **2018**, *8*, No. e1340.
(66) Goedecker, S.; Teter, M.; Hutter, J. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B* **1996**, *54*, 1703−1710.
(67) Hartwigsen, C.; Goedecker, S.; Hutter, J. Relativistic separable dual-space Gaussian pseudopotentials from H to Rn. *Phys. Rev. B* **1998**, *58*, 3641−3662.
(68) Hutter, J. *New optimization of GTH pseudopotentials for PBE, SCAN, PBE0 functionals. GTH pseudopotentials for Hartree-Fock. NLCC pseudopotentials for PBE*. 2019; https://github.com/juerghutter/GTH, Accessed: 2024−05−23.
(69) Ye, H.-Z.; Berkelbach, T. C. Correlation-Consistent Gaussian Basis Sets for Solids Made Simple. *J. Chem. Theory Comput.* **2022**, *18*, 1595−1606.

https://doi.org/10.1021/acs.jpca.4c02431
J. Phys. Chem. A 2024, 128, 7451−7461