

# Interpolative Separable Density Fitting through Centroidal Voronoi Tessellation with Applications to Hybrid Functional Electronic Structure Calculations

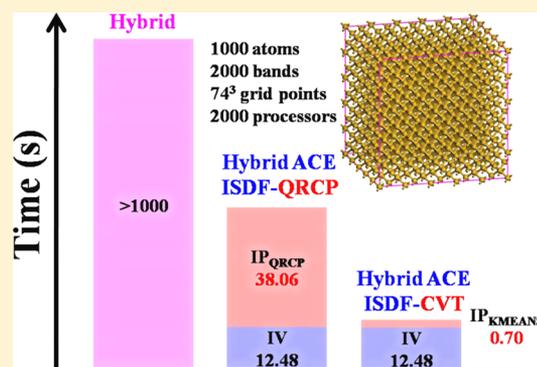
Kun Dong,<sup>\*,†</sup> Wei Hu,<sup>\*,‡,§</sup> and Lin Lin<sup>\*,§,‡,†</sup>

<sup>†</sup>Center for Applied Mathematics, Cornell University, Ithaca, New York 14853, United States

<sup>‡</sup>Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States

<sup>§</sup>Department of Mathematics, University of California, Berkeley, California 94720, United States

**ABSTRACT:** The recently developed interpolative separable density fitting (ISDF) decomposition is a powerful way for compressing the redundant information in the set of orbital pairs and has been used to accelerate quantum chemistry calculations in a number of contexts. The key ingredient of the ISDF decomposition is to select a set of nonuniform grid points, so that the values of the orbital pairs evaluated at such grid points can be used to accurately interpolate those evaluated at all grid points. The set of nonuniform grid points, called the interpolation points, can be automatically selected by a QR factorization with column pivoting (QRCP) procedure. This is the computationally most expensive step in the construction of the ISDF decomposition. In this work, we propose a new approach to find the interpolation points based on the centroidal Voronoi tessellation (CVT) method, which offers a much less expensive alternative to the QRCP procedure when ISDF is used in the context of hybrid functional electronic structure calculations. The CVT method only uses information from the electron density and can be efficiently implemented using a K-Means algorithm. We find that this new method achieves comparable accuracy to the ISDF-QRCP method, at a cost that is negligible in the overall hybrid functional calculations. For instance, for a system containing 1000 silicon atoms simulated using the HSE06 hybrid functional on 2000 computational cores, the cost of the QRCP-based method for finding the interpolation points is 38.1 s, while the CVT procedure only takes 0.7 s. We also find that the ISDF-CVT method enhances the smoothness of the potential energy surface in the context of ab initio molecular dynamics (AIMD) simulations with hybrid functionals.



## 1. INTRODUCTION

Orbital pairs of the form  $\{\varphi_i(\mathbf{r})\psi_j(\mathbf{r})\}_{ij=1}^N$ , where  $\varphi_i, \psi_j$  are single particle orbitals, appear ubiquitously in quantum chemistry. A few examples include the Fock exchange operator, the MP2 amplitude, and the polarizability operator.<sup>1,2</sup> When  $N$  is proportional to the number of electrons  $N_e$  in the system, the total number of orbital pairs is  $N^2 \sim O(N_e^2)$ . On the other hand, the number of degrees of freedom needed to resolve all orbital pairs on a dense grid is only  $O(N_e)$ . Hence as  $N_e$  becomes large, the set of all orbital pairs contains apparent redundant information. In order to compress the redundant information and to design more efficient numerical algorithms, many algorithms in the past few decades have been developed. Pseudospectral decomposition,<sup>3,4</sup> Cholesky decomposition,<sup>5–8</sup> density fitting (DF) or resolution of identity (RI),<sup>9,10</sup> and tensor hypercontraction (THC)<sup>11,12</sup> are only a few examples toward this goal. When the single particle orbitals  $\varphi_i, \psi_j$  are already localized functions, “local methods” or “linear scaling methods”<sup>13–16</sup> can be applied to construct such decomposition with cost that scales linearly with respect to  $N_e$ . Otherwise, the storage cost of the matrix to represent all orbital pairs on a grid

is already  $O(N_e^3)$ , and the computational cost of compressing the orbital pairs is then typically  $O(N_e^4)$ .

Recently, Lu and Ying developed a new decomposition called the interpolative separable density fitting (ISDF),<sup>17</sup> which takes the following form

$$\varphi_i(\mathbf{r})\psi_j(\mathbf{r}) \approx \sum_{\mu=1}^{N_\mu} \zeta_\mu(\mathbf{r})(\varphi_i(\hat{\mathbf{r}}_\mu)\psi_j(\hat{\mathbf{r}}_\mu)) \quad (1)$$

For a given  $\mathbf{r}$ , if we view  $\varphi_i(\mathbf{r})\psi_j(\mathbf{r})$  as a row of the matrix  $\{\varphi_i\psi_j\}$  discretized on a dense grid, then the ISDF decomposition states that all such matrix rows can be approximately expanded using a linear combination of matrix rows with respect to a selected set of interpolation points  $\{\hat{\mathbf{r}}_\mu\}_{\mu=1}^{N_\mu}$ . The coefficients of such linear combination, or interpolating vectors, are denoted by  $\{\zeta_\mu(\mathbf{r})\}_{\mu=1}^{N_\mu}$ . Here  $N_\mu$  can be interpreted as the numerical rank of the ISDF decomposition. Compared to the standard density fitting method, the three-tensor  $(\varphi_i(\hat{\mathbf{r}}_\mu)\psi_j(\hat{\mathbf{r}}_\mu))$  with three indices  $i, j$ ,

Received: November 5, 2017

Published: January 25, 2018

$\mu$  takes a separable form. This reduces the storage cost of the decomposed tensor from  $O(N_e^3)$  to  $O(N_e^2)$  and the computational cost from  $O(N_e^4)$  to  $O(N_e^3)$ . Note that if the interpolation points  $\{\hat{\mathbf{r}}_\mu\}_{\mu=1}^{N_\mu}$  are chosen to be on a uniform grid, then the ISDF decomposition reduces to the pseudospectral decomposition, where  $N_\mu \sim O(N_e)$  but with a large preconstant. For instance, the pseudospectral decomposition can be highly inefficient for molecular systems, where the grid points in the vacuum contribute nearly negligibly to the orbital pairs. On the other hand, by selecting the interpolation points carefully, e.g., through a randomized QR factorization with column pivoting (QRCP) procedure,<sup>18</sup> the number of interpolation points can be significantly reduced. The QRCP based ISDF decomposition has been applied to accelerate a number of applications, at least in the context of pseudopotential approximation where the wave functions are smooth, including two-electron integral computation,<sup>17</sup> correlation energy in the random phase approximation,<sup>19</sup> density functional perturbation theory,<sup>20</sup> and hybrid density functional calculations.<sup>21</sup> For example, when iterative solvers are used for hybrid density functional calculations, the Fock exchange operator  $V_X$  defined in terms of a set of orbitals  $\{\varphi_i\}$  needs to be repeatedly applied to another set of Kohn–Sham orbitals  $\{\psi_j\}$

$$(V_X[\{\varphi_i\}]\psi_j)(\mathbf{r}) = -\sum_{i=1}^{N_e} \varphi_i(\mathbf{r}) \int K(\mathbf{r}, \mathbf{r}') \varphi_i(\mathbf{r}') \psi_j(\mathbf{r}') d\mathbf{r}' \quad (2)$$

where  $K(\mathbf{r}, \mathbf{r}')$  is the kernel for the Coulomb or the screened Coulomb operator. The integration in eq 2 is often carried out by solving Poisson-like equations, using, e.g., a fast Fourier transform (FFT) method, and the computational cost is  $O(N_e^3)$  with a large preconstant. This is typically the most time-consuming component in hybrid functional calculations and can be accelerated by the ISDF decomposition for the orbital pairs  $\{\varphi_i\psi_j\}$ .

In ref 17, the interpolation points and the interpolation vectors are determined simultaneously through a randomized QR factorization with column pivoting (QRCP) applied to  $\{\varphi_i(\mathbf{r})\psi_j(\mathbf{r})\}$  directly. We recently found that the randomized QRCP procedure has  $O(N_e^3)$  complexity but with a relatively large preconstant and may not be competitive enough when used repeatedly. In order to overcome such difficulty, we proposed a different approach in ref 21 that determines the two parts separately and reduces the computational cost. We use the relatively expensive randomized QRCP procedure to find the interpolation points in advance, and only recompute the interpolation vectors whenever  $\{\varphi_i(\mathbf{r})\psi_j(\mathbf{r})\}$  has been updated using an efficient least-squares procedure that exploits the separable nature of the matrix to be approximated. As a result, we can significantly accelerate hybrid functional calculations using the ISDF decomposition in all but the first SCF iteration.

In this work, we further remove the need of performing the QRCP decomposition completely and, hence, significantly reduce the computational cost. Note that an effective choice of the set of interpolation points should satisfy the following two conditions. (1) The distribution of the interpolation points should roughly follow the distribution of the electron density. In particular, there should be more points when the electron density is high and less or even zero points if the electron density is very low. (2) The interpolation points should not be

very close to each other. Otherwise, matrix rows represented by the interpolation points are nearly linearly dependent, and the matrix formed by the interpolation vectors will be highly ill-conditioned. The QRCP procedure satisfies both (1) and (2) simultaneously and, thus, is an effective way for selecting the interpolation points. Here we demonstrate that (1) and (2) can also be satisfied through a much simpler centroidal Voronoi tessellation (CVT) procedure applied to a weight vector such as the electron density.

The Voronoi tessellation technique has been widely used in computer science<sup>22</sup> and scientific and engineering applications such as image processing,<sup>23</sup> pattern recognition,<sup>24</sup> and numerical integration.<sup>25</sup> The concept of Voronoi tessellation can be simply understood as follows. Given a discrete set of weighted points, the CVT procedure divides a domain into a number of regions, each consisting of a collection of points that are closest to its weighted centroid. Here we choose the electron density as the weight and the centroids as the interpolation points. The centroids must be located where the electron density is significant and, hence, satisfy requirement 1. The centroids are also mutually separated from each other by a finite distance due to the nearest neighbor principle and, hence, satisfy requirement 2. Although detailed analysis of the error stemming from such a choice of interpolation points is very difficult for general nonlinear functions, we find that the CVT procedure approximately minimizes the residual of the ISDF decomposition (1). In practice, the CVT procedure only applies to one vector (the electron density) instead of  $O(N_e^2)$  vectors and hence is very efficient.

We apply the ISDF-CVT method to accelerate hybrid functional calculations in a planewave basis set. We perform such calculations for different systems with insulating (liquid water), semiconducting (bulk silicon), and metallic (disordered silicon aluminum alloy) characters, as well as ab initio molecular dynamics (AIMD) simulations. We find that the ISDF-CVT method achieves similar accuracy to that obtained from the ISDF-QRCP method, with significantly improved efficiency. For instance, for a bulk silicon system containing 1000 silicon atoms computed on 2000 computational cores with kinetic energy cutoff being 10 Ha, the QRCP procedure finds the interpolation points with 38.1 s, while the CVT procedure only takes 0.7 s. Since the solution of the CVT procedure is continuous with respect to changes in the electron density, we also find that the CVT procedure produces a smoother potential energy surface than that by the QRCP procedure in the context of ab initio molecular dynamics (AIMD) simulations.

The remainder of the paper is organized as follows. We briefly introduce the ISDF decomposition in Section 2. In Section 3 we describe the ISDF-CVT procedure and its implementation for hybrid functional calculations. We present numerical results of the ISDF-CVT method in Section 4 and conclude in Section 5. We also provide the theoretical justification of the CVT method in Appendix A.

## 2. INTERPOLATIVE SEPARABLE DENSITY FITTING (ISDF) DECOMPOSITION

In this section, we briefly introduce the ISDF decomposition<sup>17</sup> evaluated using the method developed in ref 21, which employs a separate treatment of the interpolation points and interpolation vectors.

First, assume the interpolation points  $\{\hat{\mathbf{r}}_{\mu}\}_{\mu=1}^{N_{\mu}}$  are known, and then the interpolation vectors can be efficiently evaluated using a least-squares method as follows. Using a linear algebra notation, eq 1 can be written as

$$Z \approx \Theta C \quad (3)$$

where each column of  $Z$  is given by  $Z_{ij}(\mathbf{r}) = \varphi_i(\mathbf{r})\psi_j(\mathbf{r})$  sampled on a dense real space grids  $\{\mathbf{r}_i\}_{i=1}^{N_g}$ , and  $\Theta = [\zeta_1, \zeta_2, \dots, \zeta_{N_{\mu}}]$  contains the interpolating vectors. Each column of  $C$  indexed by  $(i, j)$  is given by

$$[\varphi_i(\hat{\mathbf{r}}_1)\psi_j(\hat{\mathbf{r}}_1), \dots, \varphi_i(\hat{\mathbf{r}}_{\mu})\psi_j(\hat{\mathbf{r}}_{\mu}), \dots, \varphi_i(\hat{\mathbf{r}}_{N_{\mu}})\psi_j(\hat{\mathbf{r}}_{N_{\mu}})]^T$$

Equation 3 is an overdetermined linear system with respect to the interpolation vectors  $\Theta$ . The least-squares approximation to the solution is given by

$$\Theta = ZC^T(CC^T)^{-1} \quad (4)$$

It may appear that the matrix–matrix multiplications  $ZC^T$  and  $CC^T$  take  $O(N_g^4)$  operations because the size of  $Z$  is  $N_g \times N^2$  and the size of  $C$  is  $N_{\mu} \times N^2$ . However, both multiplications can be carried out with fewer operations due to the separable structure of  $Z$  and  $C$ . The computational complexity for computing the interpolation vectors is  $O(N_g^3)$ , and numerical results indicate that the preconstant is also much smaller than that involved in hybrid functional calculations.<sup>21</sup> Hence the interpolation vectors can be obtained efficiently using the least-squares procedure.

The problem for finding a suitable set of interpolation points  $\{\hat{\mathbf{r}}_{\mu}\}_{\mu=1}^{N_{\mu}}$  can be formulated as the following linear algebra problem. Consider the discretized matrix  $Z$  of size  $N_g \times N^2$  and find  $N_{\mu}$  rows of  $Z$  so that the rest of the rows of  $Z$  can be approximated by the linear combination of the selected  $N_{\mu}$  rows. This is called an interpolative decomposition,<sup>26</sup> and a standard method to achieve such a decomposition is the QR factorization with column pivoting (QRCP) procedure<sup>26</sup> as

$$Z^T \Pi = QR \quad (5)$$

Here  $Z^T$  is the transpose of  $Z$ ,  $Q$  is an  $N^2 \times N_g$  matrix that has orthonormal columns,  $R$  is an upper triangular matrix, and  $\Pi$  is a permutation matrix chosen so that the magnitude of the diagonal elements of  $R$  form a nonincreasing sequence. The magnitude of each diagonal element  $R$  indicates how important the corresponding column of the permuted  $Z^T$  is and whether the corresponding grid point should be chosen as an interpolation point. The QRCP factorization can be terminated when the  $(N_{\mu} + 1)$ th diagonal element of  $R$  becomes less than a predetermined threshold. The leading  $N_{\mu}$  columns of the permuted  $Z^T$  are considered to be linearly independent numerically. The corresponding grid points are chosen as the interpolation points. The indices for the chosen interpolation points  $\{\hat{\mathbf{r}}_{\mu}\}$  can be obtained from indices of the nonzero entries of the first  $N_{\mu}$  columns of the permutation matrix  $\Pi$ .

The QRCP decomposition satisfies requirements 1 and 2 discussed in the Introduction. First, QRCP permutes matrix columns of  $Z^T$  with large norms to the front and pushes matrix columns of  $Z^T$  with small norms to the back. Note that the square of the vector 2-norm of the column of  $Z^T$  labeled by  $\mathbf{r}$  is just

$$\sum_{i,j=1}^N \varphi_i^2(\mathbf{r})\psi_j^2(\mathbf{r}) = \left( \sum_{i=1}^N \varphi_i^2(\mathbf{r}) \right) \left( \sum_{j=1}^N \psi_j^2(\mathbf{r}) \right) \quad (6)$$

In the case when  $\varphi_i, \psi_j$  are the set of occupied orbitals, the norm of each column of  $Z^T$  is simply the electron density. Hence the interpolation points chosen by QRCP will occur where the electron density is significant. Second, once a column is selected, all other columns are immediately orthogonalized with respect to the chosen column. Hence nearly linearly dependent matrix columns will not be selected repeatedly. As a result, the interpolation points chosen by QRCP are well separated spatially.

It turns out that the direct application of the QRCP procedure (eq 5) still requires  $O(N_g^4)$  computational complexity. The key idea used in ref 17 to lower the cost is to randomly subsample columns of the matrix  $Z$  to form a smaller matrix  $\tilde{Z}$  of size  $N_g \times \tilde{N}_{\mu}$  where  $\tilde{N}_{\mu}$  is only slightly larger than  $N_{\mu}$ . Applying the QRCP procedure to this subsampled matrix  $\tilde{Z}$  approximately yields the choice of interpolation points, but the computational complexity is reduced to  $O(N_g^3)$ . In the context of hybrid density functional calculations, we demonstrated that the cost of the randomized QRCP method can be comparable to that of applying the exchange operator in the planewave basis set.<sup>21</sup> However, the ISDF decomposition can still significantly reduce the computational cost, since the interpolation points only need to be performed once for a fixed geometric configuration.

### 3. CENTROIDAL VORONOI TESSELLATION BASED ISDF DECOMPOSITION

In this section, we demonstrate that the interpolation points can also be selected from a Voronoi tessellation procedure. For a  $d$ -dimensional space, the Voronoi tessellation partitions a set of points  $\{\mathbf{r}_i\}_{i=1}^{N_g}$  in  $\mathbb{R}^d$  into a number of disjoint cells. The partition is based on the distance of each point to a finite set of points, called its generators. In our context, let  $\{\hat{\mathbf{r}}_{\mu}\}_{\mu=1}^{N_{\mu}}$  denote such a set of generators, and the corresponding cell of a given generator  $\hat{\mathbf{r}}_{\mu}$  is defined through a cluster of points  $C_{\mu}$  is

$$C_{\mu} = \{\mathbf{r}_i | \text{dist}(\mathbf{r}_i, \hat{\mathbf{r}}_{\mu}) \leq \text{dist}(\mathbf{r}_i, \hat{\mathbf{r}}_{\nu}) \text{ for all } \mu \neq \nu\} \quad (7)$$

The distance can be chosen to be any metric, e.g., the  $L^2$  distance as  $\text{dist}(\mathbf{r}, \mathbf{r}') = \|\mathbf{r} - \mathbf{r}'\|$ . In the case when the distances of a point  $\mathbf{r}$  to  $\hat{\mathbf{r}}_{\mu}, \hat{\mathbf{r}}_{\nu}$  are exactly the same, we may arbitrarily assign  $\mathbf{r}$  to one of the clusters.

The Centroidal Voronoi tessellation (CVT) is a specific type of Voronoi tessellation in which the generator  $\hat{\mathbf{r}}_{\mu}$  is chosen to be the centroid of its cell. Given a weight function  $\rho(\mathbf{r})$  (such as the electron density), the centroid of a cluster  $C_{\mu}$  is defined as

$$\mathbf{c}(C_{\mu}) = \frac{\sum_{\mathbf{r}_i \in C_{\mu}} \mathbf{r}_i \rho(\mathbf{r}_i)}{\sum_{\mathbf{r}_i \in C_{\mu}} \rho(\mathbf{r}_i)} \quad (8)$$

Combined with the  $L^2$  distance, CVT can be viewed as a minimization problem over both all possible partition of the cells and the centroids as<sup>27</sup>

$$\{C_{\mu}^*, \mathbf{c}_{\mu}^*\} = \underset{\{C_{\mu}, \mathbf{c}_{\mu}\}}{\text{argmin}} \sum_{\mu=1}^{N_{\mu}} \sum_{\mathbf{r}_k \in C_{\mu}} \rho(\mathbf{r}_k) \left\| \mathbf{r}_k - \mathbf{c}_{\mu} \right\|^2 \quad (9)$$

and the interpolation points are then chosen to be the minimizers  $\hat{\mathbf{r}}_\mu = \mathbf{c}_\mu(C_\mu^*) = \mathbf{c}_\mu^*$ . Following the discussion in the Introduction, the electron density as the weight function  $\rho$  enforces that the interpolation points should locate at points where the electron density is significant and hence satisfies requirement 1. Since the cells  $C_\mu^*$  are disjoint, the centroids  $\mathbf{c}_\mu^*$  are also separated by a finite distance away from each other, which hence satisfies requirement 2. In Appendix A we provide another theoretical justification in the sense that the CVT method approximately minimizes the residual error of the ISDF decomposition.

Many algorithms have been developed to efficiently compute the Voronoi tessellation.<sup>28</sup> One most widely used method is Lloyd's algorithm,<sup>29</sup> which in the discrete case is equivalent to the K-Means algorithm.<sup>27</sup> The K-Means algorithm is an iterative method that greedily minimizes the objective by taking alternating steps between  $\{C_\mu\}$  and  $\{\mathbf{c}_\mu\}$ . In this work, we adopt a weighted version of the K-Means algorithm, which is demonstrated in Algorithm 1. Note that the K-Means algorithm

**Algorithm 1:** Weighted K-Means Algorithm to Find Interpolation Points for Density Fitting

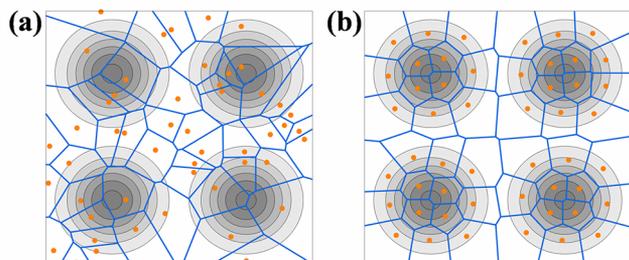
```

Input : Grid points  $\{\mathbf{r}_i\}_{i=1}^{N_g}$ , Weight function  $\rho(\mathbf{r})$ , Initial centroids  $\{\mathbf{c}_\mu^{(0)}\}$ 
Output: Interpolation points  $\{\hat{\mathbf{r}}_\mu\}_{\mu=1}^{N_\mu}$ 
1 Set  $t \leftarrow 0$ 
2 do
3   Classification step: for  $i = 1$  to  $N_g$  do
4     Assign point  $\mathbf{r}_i$  to the cluster  $C_\mu^{(t)}$  if  $\mathbf{c}_\mu^{(t)}$  is the closest centroid to  $\mathbf{r}_i$ 
5   end
6   Update step: for  $\mu = 1$  to  $N_\mu$  do
7      $\mathbf{c}_\mu^{(t+1)} \leftarrow \sum_{\mathbf{r}_j \in C_\mu^{(t)}} \mathbf{r}_j \rho(\mathbf{r}_j) / \sum_{\mathbf{r}_j \in C_\mu^{(t)}} \rho(\mathbf{r}_j)$ 
8   end
9   Set  $t \leftarrow t + 1$ 
10 while  $\{\mathbf{c}_\mu^{(t)}\}$  not converged and maximum steps not reached;
11 for  $\mu = 1$  to  $N_\mu$  do
12   Set  $\hat{\mathbf{r}}_\mu \leftarrow \mathbf{c}_\mu^{(t)}$ 
13 end

```

can be straightforwardly parallelized. We distribute the grid points evenly at the beginning. The classification step is the most time-consuming step and can be locally computed for each group of grid points. After this step, the weighted sum and total weight of all clusters can be reduced from and broadcast to all processors for the next iteration.

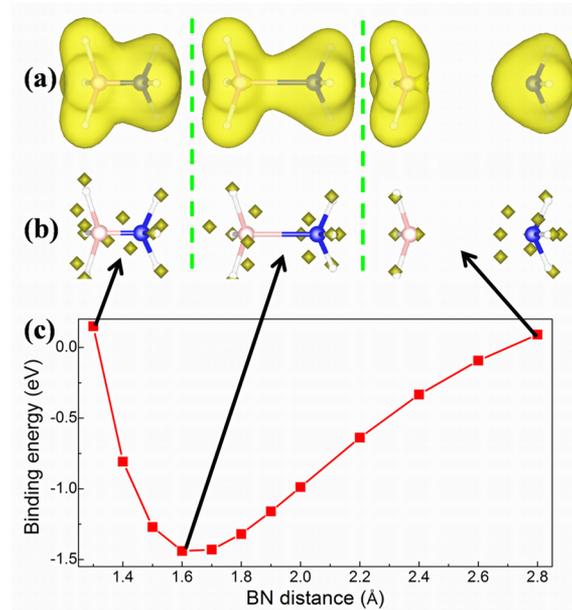
In order to demonstrate the CVT procedure, we consider the weight function  $\rho(\mathbf{r})$  given by the summation of four Gaussian functions in a 2D domain. The initial choice of centroids, given by 40 uniformly distributed random points, together with its associated Voronoi tessellations are plotted in Figure 1a. Figure 1b demonstrates the converged centroids and the associated



**Figure 1.** Schematic illustration of the CVT procedure in a 2D domain, including (a) initial random choice of centroids and Voronoi tessellation and centroidal Voronoi tessellation generated by the weighted K-Means algorithm. The weight function is given by the linear superposition of four Gaussian functions.

Voronoi tessellation using the weighted K-Means algorithm. We observe that the centroids concentrate on where the weight function is significant and are well-separated.

We also show how the interpolation points are placed and moved in real chemical systems, i.e., the ammonia–borane ( $\text{BH}_3\text{NH}_3$ ) decomposition reaction process. Figure 2a shows



**Figure 2.** Decomposition reaction process of  $\text{BH}_3\text{NH}_3$  computed with hybrid functional (HSE06) calculations by using the CVT procedure to select interpolation points, including (a) the electron density (yellow isosurfaces), (b) the interpolation points (yellow squares)  $\{\hat{\mathbf{r}}_\mu\}_{\mu=1}^{N_\mu}$  ( $N_\mu = 8$ ) selected from the real space grid points  $\{\mathbf{r}_j\}_{j=1}^{N_g}$  ( $N_g = 100^3$  and  $E_{\text{cut}} = 60$  Ha) when the BN distance respectively is 1.3, 1.7, and 2.8 Å, and (c) the binding energy as a function of BN distance for  $\text{BH}_3\text{NH}_3$  in a  $10 \text{ \AA} \times 10 \text{ \AA} \times 10 \text{ \AA}$  box. The white, pink, and blue pink balls denote hydrogen, boron, and nitrogen atoms, respectively.

the electron density of the molecule at the compressed, equilibrium, and dissociated configurations, respectively, according to the energy landscape in Figure 2c. We plot the interpolation points found by the weighted K-Means algorithm in Figure 2b. At the compressed configuration, all the interpolation points are distributed evenly around the molecule. As the bond length increases, some interpolation points are transferred from  $\text{BH}_3$  to  $\text{NH}_3$ . Finally at the dissociated configuration,  $\text{NH}_3$  has more interpolation points around the molecule, since there are more electrons in  $\text{NH}_3$  than  $\text{BH}_3$ . Along the decomposition reaction process, both the transfer of the interpolation points and the potential energy landscape are smooth with respect to the change of the bond length.

#### 4. NUMERICAL RESULTS

We demonstrate the accuracy and efficiency of the ISDF-CVT method for hybrid functional calculations by using the DGDFT (Discontinuous Galerkin Density Functional Theory) software package.<sup>30–34</sup> DGDFT is a massively parallel electronic structure software package designed for large scale DFT calculations involving up to tens of thousands of atoms. It includes a self-contained module called PWDFT for performing plane-wave based electronic structure calculations (mostly for benchmark and validation purposes). We implemented the ISDF-CVT method in PWDFT. We use the Message Passing

**Table 1. Accuracy of ACE-ISDF Based Hybrid Functional Calculations (HSE06) Obtained by Using the CVT Method To Select Interpolation Points, with Varying Rank Parameter  $c$  for Semiconducting  $\text{Si}_{216}$  and Metallic  $\text{Al}_{176}\text{Si}_{24}$  Systems<sup>a</sup>**

$c$	$E_{\text{VBM}}$	$E_{\text{CBM}}$	$E_{\text{gap}}$	$\Delta E_{\text{HF}}$	$\Delta E$	$\Delta F$	$T_{\text{KMEANS}}$
ACE-ISDF: Semiconducting $\text{Si}_{216}$ ( $N_{\text{band}} = 432$ )							
4.0	6.7467	8.3433	-1.5967	$2.69 \times 10^{-03}$	$3.08 \times 10^{-03}$	$5.04 \times 10^{-03}$	0.228
5.0	6.6852	8.2231	-1.5379	$9.46 \times 10^{-04}$	$1.12 \times 10^{-03}$	$2.29 \times 10^{-03}$	0.248
6.0	6.6640	8.1522	-1.4882	$3.76 \times 10^{-04}$	$4.62 \times 10^{-04}$	$1.05 \times 10^{-03}$	0.301
7.0	6.6550	8.1163	-1.4613	$1.55 \times 10^{-04}$	$1.98 \times 10^{-04}$	$6.49 \times 10^{-04}$	0.312
8.0	6.6510	8.1030	-1.4520	$7.33 \times 10^{-05}$	$9.55 \times 10^{-05}$	$3.07 \times 10^{-04}$	0.349
9.0	6.6490	8.0980	-1.4490	$3.60 \times 10^{-05}$	$4.96 \times 10^{-05}$	$2.30 \times 10^{-04}$	0.398
10.0	6.6479	8.0959	-1.4480	$1.78 \times 10^{-05}$	$2.64 \times 10^{-05}$	$1.30 \times 10^{-04}$	0.477
12.0	6.6472	8.0945	-1.4473	$4.46 \times 10^{-06}$	$8.91 \times 10^{-06}$	$8.37 \times 10^{-05}$	0.530
16.0	6.6469	8.0937	-1.4468	$1.51 \times 10^{-07}$	$1.41 \times 10^{-06}$	$3.20 \times 10^{-05}$	0.773
20.0	6.6468	8.0935	-1.4467	$4.06 \times 10^{-07}$	$3.33 \times 10^{-07}$	$1.20 \times 10^{-05}$	0.830
24.0	6.6468	8.0935	-1.4467	$2.99 \times 10^{-07}$	$1.06 \times 10^{-07}$	$5.18 \times 10^{-06}$	0.931
ACE	6.6468	8.0934	-1.4466	$0.00 \times 10^{00}$	$0.00 \times 10^{00}$	$0.00 \times 10^{00}$	-
ACE-ISDF: Metallic $\text{Al}_{176}\text{Si}_{24}$ ( $N_{\text{band}} = 312$ )							
4.0	7.9258	8.0335	-0.1076	$3.80 \times 10^{-03}$	$4.03 \times 10^{-03}$	$8.01 \times 10^{-03}$	0.430
5.0	7.8537	7.9596	-0.1059	$1.60 \times 10^{-03}$	$1.69 \times 10^{-03}$	$3.18 \times 10^{-03}$	0.535
6.0	7.8071	7.9127	-0.1056	$6.07 \times 10^{-04}$	$6.39 \times 10^{-04}$	$1.48 \times 10^{-03}$	0.611
7.0	7.7843	7.8860	-0.1017	$2.07 \times 10^{-04}$	$2.17 \times 10^{-04}$	$1.03 \times 10^{-03}$	0.731
8.0	7.7749	7.8749	-0.1000	$7.43 \times 10^{-05}$	$7.77 \times 10^{-05}$	$4.40 \times 10^{-04}$	0.948
9.0	7.7718	7.8710	-0.0992	$3.02 \times 10^{-05}$	$3.20 \times 10^{-05}$	$1.98 \times 10^{-04}$	0.947
10.0	7.7709	7.8697	-0.0989	$1.48 \times 10^{-05}$	$1.60 \times 10^{-05}$	$1.80 \times 10^{-04}$	1.096
12.0	7.7703	7.8690	-0.0987	$4.64 \times 10^{-06}$	$5.60 \times 10^{-06}$	$8.51 \times 10^{-05}$	1.305
16.0	7.7702	7.8688	-0.0986	$6.35 \times 10^{-07}$	$1.41 \times 10^{-06}$	$3.24 \times 10^{-05}$	1.646
20.0	7.7701	7.8687	-0.0986	$1.70 \times 10^{-08}$	$5.30 \times 10^{-07}$	$1.91 \times 10^{-05}$	2.037
ACE	7.7701	7.8687	-0.0986	$0.00 \times 10^{00}$	$0.00 \times 10^{00}$	$0.00 \times 10^{00}$	-

<sup>a</sup>The unit for VBM ( $E_{\text{VBM}}$ ), CBM ( $E_{\text{CBM}}$ ), and the energy gap  $E_{\text{gap}}$  is eV. The unit for the error in the Hartree-Fock exchange energy  $\Delta E_{\text{HF}}$  and the total energy  $\Delta E$  is Ha/atom, and the unit for the error in atomic forces  $\Delta F$  is Ha/Bohr. We use the results from the ACE-enabled hybrid functional calculations as the reference. The last column shows the time (in seconds) for K-Means with different  $c$  values, with 434 cores for  $\text{Si}_{216}$  and 314 cores for  $\text{Al}_{176}\text{Si}_{24}$  on Edison.

Interface (MPI) to handle data communication. We use the Hartwigsen–Goedecker–Hutter (HGH) norm-conserving pseudopotential.<sup>35</sup> The atomic valence electron configuration is  $1s^1$  for the H atom,  $2s^2 2p^1$  for the B atom,  $2s^2 2p^3$  for the N atom,  $2s^2 2p^4$  for the O atom, and  $3s^2 3p^2$  for the Si atom in our DFT calculations, respectively. All calculations use the HSE06 functional.<sup>36</sup> All calculations are carried out on the Edison systems at the National Energy Research Scientific Computing Center (NERSC). Each node consists of two Intel “Ivy Bridge” processors with 24 cores in total and 64 gigabytes (GB) of memory. Our implementation only uses MPI. The number of cores is equal to the number of MPI ranks used in the simulation.

In this section, we demonstrate the performance of the ISDF-CVT method for accelerating hybrid functional calculations by using three types of systems.<sup>37</sup> They consist of bulk silicon systems ( $\text{Si}_{64}$ ,  $\text{Si}_{216}$ , and  $\text{Si}_{1000}$ ), a bulk water system with 64 molecules ( $(\text{H}_2\text{O})_{64}$ ), and a disordered silicon aluminum alloy system ( $\text{Al}_{176}\text{Si}_{24}$ ). Bulk silicon systems ( $\text{Si}_{64}$ ,  $\text{Si}_{216}$ , and  $\text{Si}_{1000}$ ) and the bulk water system ( $(\text{H}_2\text{O})_{64}$ ) are semiconducting with a relatively large energy gap  $E_{\text{gap}} > 1.0$  eV, and the  $\text{Al}_{176}\text{Si}_{24}$  system is metallic with a small energy gap  $E_{\text{gap}} < 0.1$  eV. All systems are closed shell systems, and the number of occupied bands is  $N_{\text{band}} = N_e/2$ , where  $N_e$  is the number of valence electrons. In order to compute the energy gap in the systems, we also include two unoccupied bands in all calculations.

**4.1. Accuracy:  $\text{Si}_{216}$  and  $\text{Al}_{176}\text{Si}_{24}$ .** We demonstrate the accuracy of the CVT-based ISDF decomposition in the hybrid

functional calculation for semiconducting  $\text{Si}_{216}$  and metallic  $\text{Al}_{176}\text{Si}_{24}$  systems, respectively. Although there is no general theoretical guarantee for the convergence of the K-Means algorithm and the convergence can depend sensitively on the initialization,<sup>38,39</sup> we find that, in the current context, initialization has little impact on the final accuracy of the approximation. Hence we use random initialization for the K-Means algorithm. In all calculations, the adaptively compressed exchange (ACE) technique is used to accelerate hybrid functional calculations without loss of accuracy.<sup>40</sup> The results obtained in this work are labeled as ACE-ISDF (CVT), which are compared against those obtained from the previous work based on the QRCP decomposition<sup>21</sup> labeled as ACE-ISDF (QRCP). In both cases, we introduce a rank parameter  $c$  to control the trade off between efficiency and accuracy, by setting the number of interpolation points  $N_\mu = cN_e$ . We measure the error using the valence band maximum (VBM) energy level, the conduction band minimum (CBM) energy level, the energy gap, the Hartree–Fock exchange energy, the total energy, and the atomic forces, respectively. We remark that, in ISDF-CVT and ISDF-QRCP, the atomic force is computed directly using the Hellmann–Feynman formula and, hence, neglects the Pulay force contribution from the change of the interpolation points. On the other hand, there is no Pulay contribution in the ACE formulation, and the Hellmann–Feynman force  $F_i^{\text{ACE}}$  can be used as the reference solution.

The last three quantities are defined as

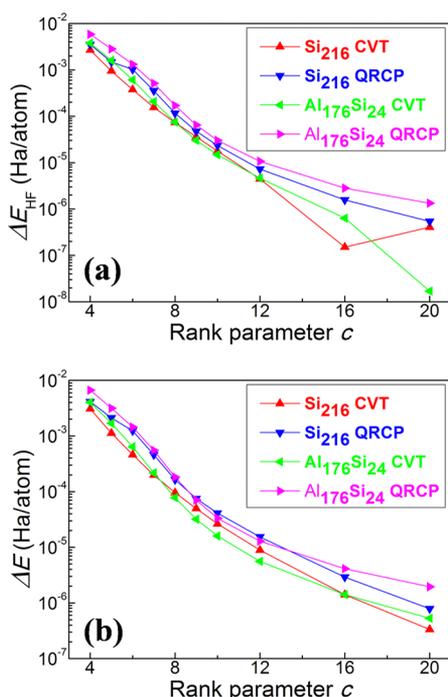
$$\Delta E_{\text{HF}} = |E_{\text{HF}}^{\text{ACE-ISDF(CVT)}} - E_{\text{HF}}^{\text{ACE}}|/N_A$$

$$\Delta E = |E^{\text{ACE-ISDF(CVT)}} - E^{\text{ACE}}|/N_A$$

$$\Delta F = \max_I \left\| F_I^{\text{ACE-ISDF(CVT)}} - F_I^{\text{ACE}} \right\|$$

where  $N_A$  is the number of atoms and  $I$  is the atom index.

Table 1 shows that the accuracy of the ACE-ISDF (CVT) method can systematically improve as the rank parameter  $c$  increases. When the rank parameter is large enough ( $\geq 20.0$ ), the results from ACE-ISDF (CVT) are fully comparable (the energy error is below  $10^{-6}$  Ha/atom and the force error is below  $10^{-5}$  Ha/Bohr) to those obtained from the benchmark calculations. Furthermore, for a moderate choice of the rank parameter  $c = 6.0$ , the error of the energy per atom reaches below the chemical accuracy of 1 kcal/mol ( $1.6 \times 10^{-3}$  Ha/atom), and the error of the force is around  $10^{-3}$  Ha/Bohr. This is comparable to the accuracy obtained from ACE-ISDF (QRCP) and to, e.g., linear scaling methods for insulating systems with a reasonable amount of truncation needed to achieve significant speedup.<sup>41</sup> In fact, when compared with ACE-ISDF (QRCP) in Figure 3, we find that the CVT based



**Figure 3.** Accuracy of ACE-ISDF based hybrid functional calculations (HSE06) obtained by using the CVT and QRCP procedures to select the interpolation points, with varying rank parameter  $c$  from 4 to 20 for  $\text{Si}_{216}$  and  $\text{Al}_{176}\text{Si}_{24}$  including the error of (a) Hartree–Fock exchange energy  $\Delta E_{\text{HF}}$  (Ha/atom) and (b) total energy  $\Delta E$  (Ha/atom).

ISDF decomposition achieves slightly higher accuracy, though there is no theoretical guarantee for this to hold in general. The last column of Table 1 shows the runtime of the K-Means algorithm. As  $c$  increases, the number of interpolation points as well as the number of cells increases proportionally. Hence we observe that the runtime of K-Means scales linearly with respect to  $c$ .

**4.2. Efficiency:  $\text{Si}_{1000}$ .** We report the efficiency of the ISDF-CVT method by performing hybrid DFT calculations for a bulk silicon system with 1000 atoms ( $N_{\text{band}} = 2000$ ) on 2000

computational cores as shown in Table 2, with respect to various choices of the kinetic energy cutoff ( $E_{\text{cut}}$ ). With the

**Table 2.** Wall Clock Time (in seconds) Spent in the Components of the ACE-ISDF and ACE Enabled Hybrid DFT Calculations Related to the Exchange Operator, for  $\text{Si}_{1000}$  on 2002 Edison Cores at Different  $E_{\text{cut}}$  Levels<sup>a</sup>

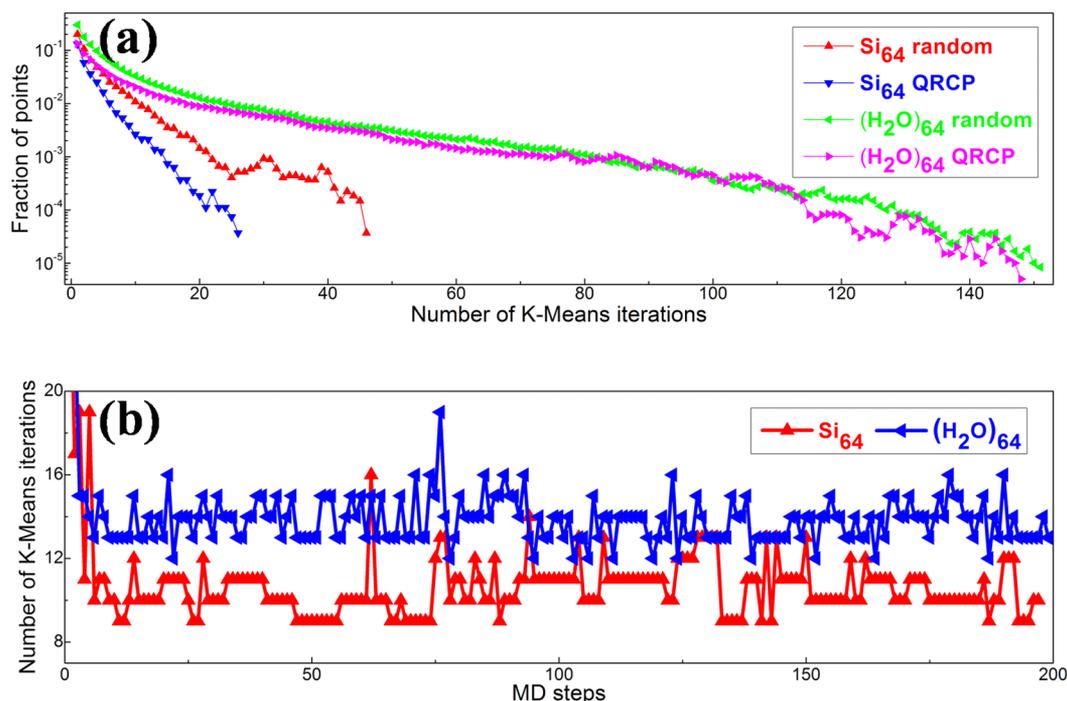
$\text{Si}_{1000}$		ACE-ISDF			ACE
$E_{\text{cut}}$	$N_g$	IP <sub>QRCP</sub>	IP <sub>KMEANS</sub>	IV (FFT)	FFT
10	74 <sup>3</sup>	38.06	0.70	12.48 (0.33)	85.15
20	104 <sup>3</sup>	126.39	1.24	36.48 (0.71)	143.54
30	128 <sup>3</sup>	240.87	2.03	68.50 (1.43)	268.88
40	148 <sup>3</sup>	434.16	3.26	108.18 (3.10)	783.27

<sup>a</sup>Interpolation points are selected via either the QRCP or the CVT procedure with the same rank parameter  $c = 6.0$ .  $N_g$  is the number of grid points in real space.

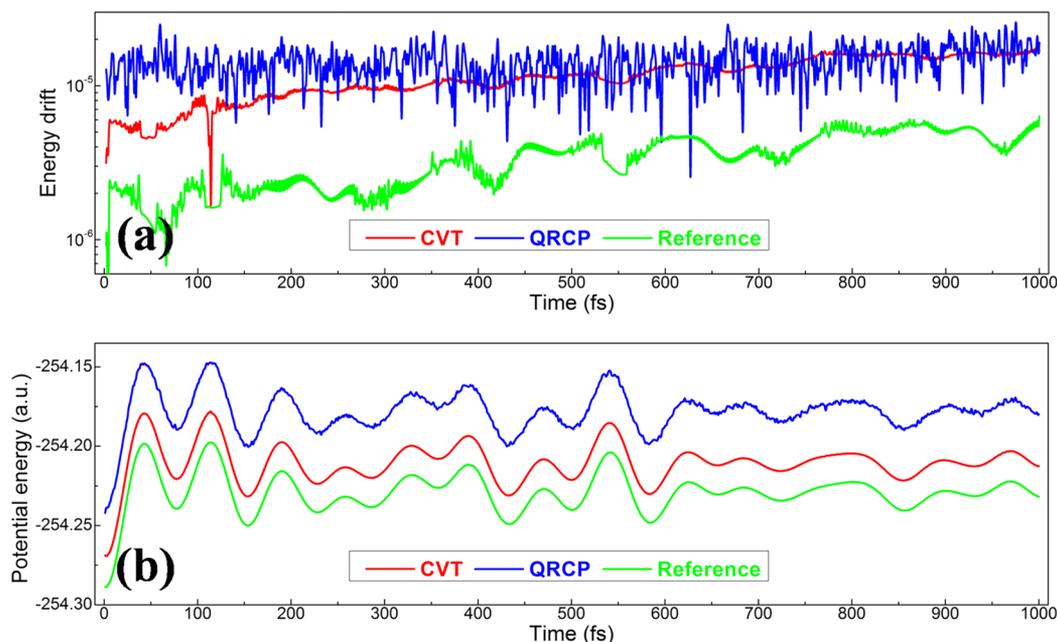
number of interpolation points fixed at  $N_\mu = 12\,000$ , both QRCP and K-Means scales linearly with the number of grid points  $N_g$ . Yet the runtime of K-Means is around 2 orders of magnitude faster than that of QRCP. The determination of interpolation vectors, which consists of solving a least-squares problem, previously costs a fifth of the ISDF runtime but now becomes the dominating component in CVT-based ISDF decomposition. Notice that the ISDF method allows us to reduce the number of Poisson-like equations from  $N_e^2 = 4 \times 10^6$  to  $N_\mu = 12\,000$ , which results in a significant speedup in terms of the cost of the FFT operations.

**4.3. AIMD:  $\text{Si}_{64}$  and  $(\text{H}_2\text{O})_{64}$ .** In this section, we demonstrate the accuracy of the ACE-ISDF (CVT) method in the context of AIMD simulations for a bulk silicon system  $\text{Si}_{64}$  under the NVE ensemble and a liquid water system  $(\text{H}_2\text{O})_{64}$  under the NVT ensemble, respectively. For the  $\text{Si}_{64}$  system, the initial MD structure (initial temperature  $T = 300$  K) is optimized by hybrid DFT calculations, and we perform the simulation ( $E_{\text{cut}} = 20$  Ha) for 1.0 ps with a MD time step of 1.0 fs. For the  $(\text{H}_2\text{O})_{64}$  system, we perform the simulation ( $E_{\text{cut}} = 60$  Ha) for 2.0 ps with a MD time step of 0.5 fs to sample the radial distribution function after equilibrating the system starting from a prepared initial guess.<sup>42</sup> In this case, the van der Waals (VdW) interaction is modeled at the level of the DFT-D2 method.<sup>43</sup> We use a single level Nose-Hoover thermostat<sup>44,45</sup> at  $T = 295$  K, and the choice of mass of the Nose-Hoover thermostat is 85 000 au.

In the AIMD simulation, the interpolation points need to be recomputed for each atomic configuration. At the initial MD step, although the initialization strategy does not impact the accuracy of the physical observable, it can impact the convergence rate of the K-Means algorithm. We measure the convergence in terms of the fraction of points that switch clusters during two consecutive iterations. Figure 4a shows the convergence of the K-Means algorithm with interpolation points initially chosen from a random distribution and from the QRCP solution, respectively. We find that the K-Means algorithm spends around half the number of iterations to wait for 0.1% of the points to settle on the respective clusters. However, these points often belong to the boundary of the clusters and have little effect on the positions of the centroids (interpolation points). Therefore, we decide to terminate the K-Means algorithm whenever the fraction of points that switch clusters falls below the 0.1% threshold. It is evident that QRCP initialization leads to faster convergence than random sampling.



**Figure 4.** Comparison of the ISDF-CVT method by using either random or QRCP initialization for hybrid DFT AIMD simulations on bulk silicon system  $\text{Si}_{64}$  and liquid water system  $(\text{H}_2\text{O})_{64}$ , including (a) the fraction of points of the switch cluster in each K-Means iteration and (b) the number of K-Means iterations during each MD step.



**Figure 5.** Comparison of hybrid HSE06 DFT AIMD simulations by using the ISDF-CVT and ISDF-QRCP methods as well as exact nested two-level SCF iteration procedure as the reference on the bulk silicon  $\text{Si}_{64}$ , including (a) relatively energy drift and (b) potential energy during MD steps.

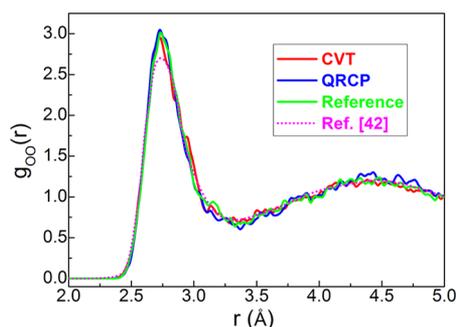
However, in the AIMD simulation, a very good initial guess of the interpolation points can be simply obtained from those from the previous MD step. Figure 4b shows that the number of K-Means iterations in the MD simulation can be very small, which demonstrates the effectiveness of this initialization strategy.

Figure 5a shows that both the CVT-based and QRCP-based ISDF decomposition lead to controlled energy drift, defined as  $E_{\text{drift}}(t) = (E_{\text{tot}}(t) - E_{\text{tot}}(0))/E_{\text{tot}}(0)$ . In the NVE simulation on

bulk silicon system  $\text{Si}_{64}$ , the energy drifts per atom are  $6.6 \times 10^{-5}$ ,  $7.5 \times 10^{-5}$ , and  $2.5 \times 10^{-5}$  Ha/ps, respectively, for the ISDF-CVT, ISDF-QRCP, and conventional nested two-level SCF iteration procedure, indicating that ISDF is a promising method for reducing the cost of hybrid functional calculations with controllable loss of accuracy. Figure 5b shows the total potential energy obtained by the three methods along the MD trajectory, and the difference among the three methods is more noticeable. This is due to the fact that ISDF decomposition is a

low rank decomposition for the pair product of orbitals, which leads to error in the Fock exchange energy and hence the total potential energy. Nonetheless, we find that such a difference mainly results in a shift of the potential energy surface along the MD trajectory and, hence, has little effect on physical observables defined via relative potential energy differences. Furthermore, the CVT method yields a potential energy trajectory that is much smoother compared to that obtained from QRCP. This is because the interpolation points obtained from CVT are driven by the electron density, which varies smoothly along the MD trajectory. Such properties do not hold for the QRCP method. This means that the CVT method can be more effective when a smooth potential energy surface is desirable, such as in the case of geometry optimization. The absolute error of the potential energy from the CVT method is coincidentally smaller than that from QRCP, but again we are not aware of any reason for this behavior to hold in general.

We also apply the ACE-ISDF (CVT) and ACE-ISDF (QRCP) methods for hybrid DFT AIMD simulations on the liquid water system  $(\text{H}_2\text{O})_{64}$  under the NVT ensemble to sample the radial distribution function in Figure 6. We find that



**Figure 6.** Oxygen–oxygen radial distribution functions  $g_{\text{OO}}(r)$  of the liquid water system  $(\text{H}_2\text{O})_{64}$  at  $T = 295$  K obtained from hybrid HSE06 + DFT-D2 AIMD simulations with the ISDF-CVT and ISDF-QRCP methods and exact nested two-level SCF iteration procedure (as the reference) as well as previous hybrid PBE0 + TS-vdW calculation.<sup>42</sup>

the results from all three methods agree very well, and our result is in quantitative agreement with previous hybrid functional calculations,<sup>42</sup> which uses a different exchange-correlation functional (PBE0) and van der Waals functional (TS-vdW<sup>46</sup>).

## 5. CONCLUSION

In this work, we demonstrate that the interpolative separable density fitting decomposition (ISDF) can be efficiently performed through a separated treatment of interpolation points and interpolation vectors. We find that the centroidal Voronoi tessellation method (CVT) provides an effective choice of interpolation points using only the electron density as the input information. The resulting interpolation points are by design inhomogeneous in the real space, concentrated at regions where the electron density is significant, and are well separated from each other. These are all key ingredients for obtaining a low rank decomposition that is accurate and a well conditioned set of interpolation vectors. We demonstrate that the CVT-based ISDF decomposition can be an effective strategy for reducing the cost hybrid functional calculations for large systems. The CVT-based method achieves similar

accuracy when compared with that obtained from QRCP, with significantly improved efficiency. Since the solution of the CVT method depends continuously on the electron density, we also find that the CVT method produces a smoother potential energy surface than that by the QRCP method in the context of ab initio molecular dynamics simulation. Our analysis indicates that it might be possible to further improve the quality of the interpolation points by taking into account the gradient information in the weight vector. We also expect that the CVT-based strategy can also be useful in other contexts where the ISDF decomposition is applicable, such as ground state calculations with rung-5 exchange-correlation functionals, and excited state calculations. These will be explored in future work.

## ■ APPENDIX A: MINIMIZATION OF THE APPROXIMATE RESIDUAL

The ISDF decomposition is a highly nonlinear process, and in general we cannot expect the choice of interpolation points from CVT decomposition to maximally reduce the error of the decomposition. Here we demonstrate that the choice of the interpolation points from the centroidal Voronoi tessellation algorithm approximately minimizes the residual for the ISDF decomposition and, hence, provides a heuristic solution to the problem of finding interpolation points.

For simplicity we assume  $\varphi_i = \psi_i$ , and, hence, each row of  $Z$  is  $Z(\mathbf{r}) = [\varphi_i(\mathbf{r})\varphi_j(\mathbf{r})]_{i,j=1}^N$ . Now suppose we cluster all matrix rows of  $Z$  into subcollections  $\{C_\mu\}_{\mu=1}^{N_\mu}$ , and for each  $C_\mu$  we choose a representative matrix row  $Z(\mathbf{r}_\mu)$ . Then the error of the ISDF can be approximately characterized as

$$R = \sum_{\mu=1}^{N_\mu} \sum_{\mathbf{r}_k \in C_\mu} \left\| Z(\mathbf{r}_k) - \text{Proj}_{\text{span}\{Z(\mathbf{r}_\mu)\}} Z(\mathbf{r}_k) \right\|^2 \quad (10)$$

where the projection is defined according to the  $L^2$  inner product as

$$\text{Proj}_{\text{span}\{Z(\mathbf{r}_\mu)\}} Z(\mathbf{r}_k) = \frac{Z(\mathbf{r}_k) \cdot Z(\mathbf{r}_\mu)}{Z(\mathbf{r}_\mu) \cdot Z(\mathbf{r}_\mu)} Z(\mathbf{r}_\mu) \quad (11)$$

Let  $\Phi$  be the  $N_g \times N$  matrix with each row  $\Phi(\mathbf{r}) = [\varphi_i(\mathbf{r})]_{i=1}^N$ , then the electron density  $\rho(\mathbf{r})$  is equal to  $\Phi(\mathbf{r}) \cdot \Phi(\mathbf{r})$ . Using the relation

$$Z(\mathbf{r}_\mu) \cdot Z(\mathbf{r}_\mu) = (\Phi(\mathbf{r}_\mu) \cdot \Phi(\mathbf{r}_\mu))^2 = \rho(\mathbf{r}_\mu)^2 \quad (12)$$

we have

$$\begin{aligned} R &= \sum_{\mu=1}^{N_\mu} \sum_{\mathbf{r}_k \in C_\mu} \rho(\mathbf{r}_k)^2 \left( 1 - \frac{(\Phi(\mathbf{r}_k) \cdot \Phi(\mathbf{r}_\mu))^4}{\rho(\mathbf{r}_k)^2 \rho(\mathbf{r}_\mu)^2} \right) \\ &= \sum_{\mu=1}^{N_\mu} \sum_{\mathbf{r}_k \in C_\mu} \rho(\mathbf{r}_k)^2 [1 - \cos^4(\theta(\mathbf{r}_k, \mathbf{r}_\mu))] \end{aligned} \quad (13)$$

Here  $\theta(\mathbf{r}_k, \mathbf{r}_\mu)$  is the angle between the vectors  $\Phi(\mathbf{r}_k)$  and  $\Phi(\mathbf{r}_\mu)$ . Use the fact that

$$\begin{aligned} &\rho(\mathbf{r}_k) [1 - \cos^4(\theta(\mathbf{r}_k, \mathbf{r}_\mu))] \\ &\leq 2\Phi(\mathbf{r}_k) \cdot \Phi(\mathbf{r}_k) \sin^2(\theta(\mathbf{r}_k, \mathbf{r}_\mu)) \\ &\leq 2\|\Phi(\mathbf{r}_k) - \Phi(\mathbf{r}_\mu)\|^2 \end{aligned} \quad (14)$$

we have

$$\begin{aligned}
 R &\leq 2 \sum_{\mu=1}^{N_{\mu}} \sum_{\mathbf{r}_k \in C_{\mu}} \rho(\mathbf{r}_k) \|\Phi(\mathbf{r}_k) - \Phi(\mathbf{r}_{\mu})\|^2 \\
 &\approx 2 \sum_{\mu=1}^{N_{\mu}} \sum_{\mathbf{r}_k \in C_{\mu}} \rho(\mathbf{r}_k) \|\nabla_{\mathbf{r}} \Phi(\mathbf{r}_{\mu})\|^2 \|\mathbf{r}_k - \mathbf{r}_{\mu}\|^2
 \end{aligned}
 \quad (15)$$

If we further neglect the spatial inhomogeneity of the gradient  $\Phi(\mathbf{r})$ , we arrive at the minimization criterion for the centroidal Voronoi decomposition.

## AUTHOR INFORMATION

### Corresponding Authors

\*(K.D.) E-mail: [kd383@cornell.edu](mailto:kd383@cornell.edu).

\*(W.H.) E-mail: [whu@lbl.gov](mailto:whu@lbl.gov).

\*(L.L.) E-mail: [linlin@math.berkeley.edu](mailto:linlin@math.berkeley.edu).

### ORCID

Wei Hu: 0000-0001-9629-2121

Lin Lin: 0000-0001-6860-9566

### Funding

This work was partly supported by the National Science Foundation under Grant No. DMS-1652330, the DOE under Grant No. DE-SC0017867, the DOE CAMERA project (L.L.), and the DOE Scientific Discovery through Advanced Computing (SciDAC) program (K.D., W.H., and L. L.).

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

The authors thank the National Energy Research Scientific Computing (NERSC) center and the Berkeley Research Computing (BRC) program at the University of California, Berkeley, for making computational resources available. We thank Anil Damle and Robert Saye for useful discussions.

## REFERENCES

- (1) Szabo, A.; Ostlund, N. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; McGraw-Hill: New York, 1989.
- (2) Martin, R. *Electronic Structure - Basic Theory and Practical Methods*; Cambridge University Press: West Nyack, NY, 2004.
- (3) Murphy, R. B.; Beachy, M. D.; Friesner, R. A.; Ringnalda, M. N. Pseudospectral localized Møller-Plesset methods: Theory and calculation of conformational energies. *J. Chem. Phys.* **1995**, *103*, 1481.
- (4) Reynolds, G.; Martinez, T. J.; Carter, E. A. Local weak pairs spectral and pseudospectral singles and doubles configuration interaction. *J. Chem. Phys.* **1996**, *105*, 6455.
- (5) Beebe, N. H. F.; Linderberg, J. Simplifications in the generation and transformation of two-electron integrals in molecular calculations. *Int. J. Quantum Chem.* **1977**, *12*, 683.
- (6) Koch, H.; Sánchez de Merás, A.; Pedersen, T. B. Reduced scaling in electronic structure calculations using Cholesky decompositions. *J. Chem. Phys.* **2003**, *118*, 9481–9484.
- (7) Aquilante, F.; Pedersen, T. B.; Lindh, R. Low-cost evaluation of the exchange Fock matrix from Cholesky and density fitting representations of the electron repulsion integrals. *J. Chem. Phys.* **2007**, *126*, 194106.
- (8) 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*, 024113.
- (9) Ren, X.; Rinke, P.; Blum, V.; Wieferink, J.; Tkatchenko, A.; Sanfilippo, A.; Reuter, K.; Scheffler, M. Resolution-of-Identity Approach to Hartree-Fock, Hybrid Density Functionals, RPA, MP2 and GW with Numeric Atom-Centered Orbital Basis Functions. *New J. Phys.* **2012**, *14*, 053020.

(10) Weigend, F. A fully direct RI-HF algorithm: Implementation, optimized auxiliary basis sets, demonstration of accuracy and efficiency. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285–4291.

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

(12) Parrish, R. M.; Hohenstein, E. G.; Martínez, T. J.; Sherrill, C. D. Discrete variable representation in electronic structure theory: Quadrature grids for least-squares tensor hypercontraction. *J. Chem. Phys.* **2013**, *138*, 194107.

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

(14) Bowler, D. R.; Miyazaki, T. O(N) methods in Electronic Structure Calculations. *Rep. Prog. Phys.* **2012**, *75*, 036503.

(15) Guidon, M.; Hutter, J.; Vandevondele, J. Auxiliary density matrix methods for Hartree-Fock exchange calculations. *J. Chem. Theory Comput.* **2010**, *6*, 2348–2364.

(16) 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.

(17) 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.

(18) Golub, G. H.; Van Loan, C. F. *Matrix computations*, 4th ed.; Johns Hopkins University Press: Baltimore, 2013.

(19) Lu, J.; Thicke, K. Cubic scaling algorithms for RPA correlation using interpolative separable density fitting. *J. Comput. Phys.* **2017**, *351*, 187–202.

(20) Lin, L.; Xu, Z.; Ying, L. Adaptively Compressed Polarizability Operator for Accelerating Large Scale Ab Initio Phonon Calculations. *Multiscale Model. Simul.* **2017**, *15*, 29–55.

(21) 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.

(22) Aurenhammer, F. Voronoi diagrams—a survey of a fundamental geometric data structure. *ACM Computing Surveys (CSUR)* **1991**, *23*, 345–405.

(23) Du, Q.; Gunzburger, M.; Ju, L.; Wang, X. Centroidal Voronoi tessellation algorithms for image compression, segmentation, and multichannel restoration. *Journal of Mathematical Imaging and Vision* **2006**, *24*, 177–194.

(24) Ogniewicz, R. L.; Kübler, O. Hierarchic voronoi skeletons. *Pattern recognition* **1995**, *28*, 343–359.

(25) Becke, A. D. A multicenter numerical integration scheme for polyatomic molecules. *J. Chem. Phys.* **1988**, *88*, 2547–2553.

(26) Chan, T. F.; Hansen, P. C. Some Applications of the Rank Revealing QR Factorization. *SIAM J. Sci. Statist. Comput.* **1992**, *13*, 727–741.

(27) MacQueen, J. Some methods for classification and analysis of multivariate observations. *Proc. Fifth Berkeley Symp. Math. Stat. Prob.* **1967**, 281–297.

(28) Medvedev, N. The algorithm for three-dimensional Voronoi polyhedra. *J. Comput. Phys.* **1986**, *67*, 223–229.

(29) Lloyd, S. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **1982**, *28*, 129–137.

(30) Lin, L.; Lu, J.; Ying, L.; E, W. Adaptive Local Basis Set for Kohn-Sham Density Functional Theory in a Discontinuous Galerkin Framework I: Total Energy Calculation. *J. Comput. Phys.* **2012**, *231*, 2140–2154.

(31) Hu, W.; Lin, L.; Yang, C. DGDF: A Massively Parallel Method for Large Scale Density Functional Theory Calculations. *J. Chem. Phys.* **2015**, *143*, 124110.

(32) Hu, W.; Lin, L.; Yang, C. Edge Reconstruction in Armchair Phosphorene Nanoribbons Revealed by Discontinuous Galerkin Density Functional Theory. *Phys. Chem. Chem. Phys.* **2015**, *17*, 31397–31404.

(33) Banerjee, A. S.; Lin, L.; Hu, W.; Yang, C.; Pask, J. E. Chebyshev Polynomial Filtered Subspace Iteration in the Discontinuous Galerkin Method for Large-Scale Electronic Structure Calculations. *J. Chem. Phys.* **2016**, *145*, 154101.

(34) Zhang, G.; Lin, L.; Hu, W.; Yang, C.; Pask, J. E. Adaptive Local Basis Set for Kohn-Sham Density Functional Theory in a Discontinuous Galerkin Framework II: Force, Vibration, and Molecular Dynamics Calculations. *J. Comput. Phys.* **2017**, *335*, 426–443.

(35) Hartwigsen, C.; Goedecker, S.; Hutter, J. Relativistic Separable Dual-Space Gaussian Pseudopotentials from H to Rn. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1998**, *58*, 3641.

(36) Heyd, J.; Scuseria, G. E.; Ernzerhof, M. Erratum: "Hybrid functionals based on a screened Coulomb potential" [*J. Chem. Phys.* **118**, 8207 (2003)]. *J. Chem. Phys.* **2006**, *124*, 219906.

(37) Hu, W.; Lin, L.; Yang, C. Projected Commutator DIIS Method for Accelerating Hybrid Functional Electronic Structure Calculations. *J. Chem. Theory Comput.* **2017**, *13*, 5458.

(38) Arthur, D.; Vassilvitskii, S. How slow is the k-means method? *Proceedings of the twenty-second annual symposium on Computational geometry* **2006**, 144–153.

(39) Arthur, D.; Vassilvitskii, S. k-means++: The advantages of careful seeding. *Proc. Eighteenth Annual ACM-SIAM Symp. Discrete Algorithms* **2007**, 1027–1035.

(40) Lin, L. Adaptively Compressed Exchange Operator. *J. Chem. Theory Comput.* **2016**, *12*, 2242–2249.

(41) Dawson, W.; Gygi, F. Performance and Accuracy of Recursive Subspace Bisection for Hybrid DFT Calculations in Inhomogeneous Systems. *J. Chem. Theory Comput.* **2015**, *11*, 4655–4663.

(42) DiStasio, R. A.; Santra, B.; Li, Z.; Wu, X.; Car, R. The individual and collective effects of exact exchange and dispersion interactions on the ab initio structure of liquid water. *J. Chem. Phys.* **2014**, *141*, 084502.

(43) Grimme, S. Semiempirical GGA-type Density Functional Constructed with a Long-Range Dispersion Correction. *J. Comput. Chem.* **2006**, *27*, 1787–1799.

(44) Nosé, S. A Unified Formulation of the Constant Temperature Molecular Dynamics Methods. *J. Chem. Phys.* **1984**, *81*, 511.

(45) Hoover, W. G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A: At, Mol, Opt. Phys.* **1985**, *31*, 1695.

(46) Tkatchenko, A.; Scheffler, M. Accurate Molecular van der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data. *Phys. Rev. Lett.* **2009**, *102*, 073005.