

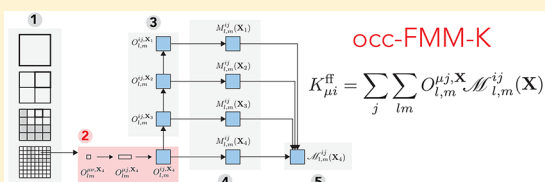
# Occupied-Orbital Fast Multipole Method for Efficient Exact Exchange Evaluation

Hai-Anh Le\*<sup>1</sup> and Toru Shiozaki

Department of Chemistry, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208, United States

**S** Supporting Information

**ABSTRACT:** We present an efficient algorithm for computing the exact exchange contributions in the Hartree–Fock and hybrid density functional theory models on the basis of the fast multipole method (FMM). Our algorithm is based on the observation that FMM with hierarchical boxes can be efficiently used in the exchange matrix construction, when at least one of the indices of the exchange matrix is constrained to be an occupied orbital. Timing benchmarks are presented for alkane chains (C<sub>400</sub>H<sub>802</sub> and C<sub>150</sub>H<sub>302</sub>), a graphene sheet (C<sub>150</sub>H<sub>30</sub>), a water cluster [(H<sub>2</sub>O)<sub>100</sub>], and a protein Crambin (C<sub>202</sub>H<sub>317</sub>O<sub>64</sub>N<sub>55</sub>S<sub>6</sub>). The computational cost of the far-field exchange evaluation for Crambin is roughly 3% that of a self-consistent field iteration when the multipoles up to rank 2 are used.



## 1. INTRODUCTION

Evaluating the electron–electron interaction in mean-field models, such as the Hartree–Fock method and hybrid density functional theory, is challenging because the bare electron–electron interaction is long-range. This is in contrast to the evaluation of screened interactions in dynamical electron correlation problems, which has been resolved to a great extent by local correlation approaches.<sup>1–3</sup> In particular, computation of the exact exchange contributions in the mean-field models remains an important challenge in quantum chemistry.<sup>4</sup> The exchange matrix elements are defined as

$$K_{\mu\nu} = \sum_{\lambda\sigma} (\mu\lambda\sigma\nu) D_{\lambda\sigma} = 2 \sum_i (\mu i \nu i) \quad (1)$$

$$(\mu\nu\lambda\sigma) = \iint d\mathbf{r}_1 d\mathbf{r}_2 \phi_\mu(\mathbf{r}_1) \phi_\nu(\mathbf{r}_1) \frac{1}{r_{12}} \phi_\lambda(\mathbf{r}_2) \phi_\sigma(\mathbf{r}_2) \quad (2)$$

where  $\mu, \nu, \lambda,$  and  $\sigma$  label atomic orbitals (AOs). Hereafter,  $i$  and  $j$  label occupied orbitals.  $D_{\lambda\sigma}$  are the density matrix elements, which become diagonal in the canonical molecular orbital (MO) representation. There have been extensive studies to optimize the exchange evaluation: for instance, the LinK method,<sup>5,6</sup> multipole accelerated algorithms,<sup>7,8</sup> rigorous integral screening,<sup>9–13</sup> density fitting with local domains,<sup>14,15</sup> truncated or short-range exchange kernels with and without the use of the resolution-of-the-identity (RI) approximation,<sup>16–19</sup> the chain-of-sphere exchange method based on quadrature,<sup>20</sup> the auxiliary density matrix method,<sup>21</sup> the pair-atomic RI approximation,<sup>22–25</sup> and the low-rank decomposition of the exchange operator.<sup>26,27</sup> For systems with sparse density matrices, combination of suitable integral prescreening and LinK can further speed up the evaluation of the exchange contribution, achieving linear scaling.

In this work, we report an efficient algorithm, termed occupied-orbital fast multipole method for exchange (occ-

FMM-K), for computing the exchange contributions based on the fast multipole method (FMM), where a system is partitioned into near-field and far-field regions.<sup>28–30</sup> Despite its tremendous success in Coulomb matrix construction,<sup>31–43</sup> FMM has been considered inapplicable to efficient computation of far-field exchange interactions. Our FMM-based algorithm for the exact exchange contributions neither relies on local orbitals nor makes any assumption about the decay properties of the density matrix (see below), making it amenable for future extensions of the algorithm to extended systems with small band gaps and to efficient computation of response properties. This algorithm is in fact complementary to existing exchange evaluation methods mentioned in the previous paragraph as they can be used to efficiently evaluate the exact integrals for the near-field contribution.

FMM was first introduced a few decades ago for evaluating the far-field Coulomb interaction energies between classical charges.<sup>28–30</sup> Many quantum chemical programs have since been developed for the Coulomb matrix evaluation.<sup>31–43</sup> In FMM, one approximates the two-electron Coulomb operator for separated charge distributions using the scaled regular and irregular solid harmonics

$$\frac{1}{r_{12}} = \sum_{l'mm'} (-1)^l O_{l,m}(\mathbf{r}_1 - \mathbf{X}) M_{l+l',m+m'}(\mathbf{X} - \mathbf{X}') \times O_{l',m'}(\mathbf{r}_2 - \mathbf{X}'), \quad (3)$$

in which the factor  $(-1)^l$  arises from the parity of the associated Legendre polynomials. The scaled regular and irregular solid harmonics (often referred to as multipoles and local expansions) are defined as

**Received:** August 17, 2017

**Published:** January 25, 2018

$$O_{l,m}(\mathbf{r}) = \epsilon_m \frac{r^l}{(l+|m|)!} P_{l,|m|}(\cos \theta) e^{-im\phi} \quad (4a)$$

$$M_{l,m}(\mathbf{r}) = \epsilon_m \frac{(l-|m|)!}{r^{l+1}} P_{l,|m|}(\cos \theta) e^{im\phi} \quad (4b)$$

in which  $\mathbf{r}$  is written in spherical coordinates  $(r, \theta, \phi)$  on the right-hand side.  $P_{l,m}$  is the associated Legendre polynomial, and  $\epsilon_m$  is a phase factor that is 1 if  $m \geq 0$  and  $(-1)^m$  otherwise.

The main idea of this work is to use FMM for computing only the exchange matrix elements that have (at least) one occupied-orbital index, taking advantage of the fact that the only matrix elements that are required to find the mean-field solution are  $K_{\mu\nu}$

$$K_{\mu i} = 2 \sum_j (\mu j | j i) \quad (5)$$

In other words, the virtual–virtual block of the exchange matrix  $K_{ab}$  is not strictly necessary. This is because self-consistent solutions minimize the mean-field energy with respect to orbital rotations between occupied orbitals  $i$  and virtual orbitals  $a$ , parametrized in terms of the exponential of an anti-Hermitian matrix  $\kappa_{ia}$ , at which the following energy gradients are made zero:

$$\frac{\partial E}{\partial \kappa_{ia}} = 2 \left( h_{ia} + J_{ia} - \frac{1}{2} K_{ia} \right) \quad (6)$$

Note that the final energy can be computed from  $K_{ij}$ . The use of the partial exchange matrix [eq 5] has been reported in a recent work by Manzer et al., who have introduced the so-called occ-RI-K algorithm.<sup>25</sup> As shown below, this trick is essential for utilizing a hierarchy of boxes with upward and downward translation of multipoles in the evaluation of far-field exchange contribution using FMM algorithm. We report an efficient, parallel implementation of the algorithm, which is publicly available as part of the BAGEL package.<sup>44,45</sup>

## 2. THEORY

When two basis-function pairs,  $\phi_\mu(\mathbf{r}_1)\phi_\nu(\mathbf{r}_1)$  and  $\phi_\lambda(\mathbf{r}_2)\phi_\sigma(\mathbf{r}_2)$ , are sufficiently separated, the electron repulsion integrals (eq 2) can be approximated using eq 3 as

$$(\mu\nu|\lambda\sigma) = \sum_{lm} (-1)^l O_{l,m}^{\mu\nu,\mathbf{X}} \sum_{l'm'} M_{l+l',m+m'}(\mathbf{X}-\mathbf{X}') O_{l',m'}^{\lambda\sigma,\mathbf{X}'} \quad (7)$$

The multipole integrals over atomic-orbital basis functions are defined as

$$O_{l,m}^{\mu\nu,\mathbf{X}} = \int d\mathbf{r} \phi_\mu(\mathbf{r}) O_{l,m}(\mathbf{r}-\mathbf{X}) \phi_\nu(\mathbf{r}) \quad (8)$$

The approximated integrals eq 7 now depend only on the multipole integrals and the separation between the expansion centers,  $\mathbf{X}$  and  $\mathbf{X}'$ . In our FMM implementation, the expansion centers are taken to be the center of the Cartesian box to which the basis pair belongs (note that  $\mathbf{X}$  is unique to each pair of  $\mu$  and  $\nu$ ).

When constructing the Coulomb matrix, the Coulomb potential at center  $\mathbf{X}$ ,  $\mathcal{M}_{l,m}(\mathbf{X})$ , due to the charge distributions associated with all distant basis function pairs  $\phi_\lambda(\mathbf{r}_2)\phi_\sigma(\mathbf{r}_2)$  is evaluated as follows. First, we contract the density matrix elements  $D_{\lambda\sigma}$  and the multipoles associated with  $\phi_\lambda(\mathbf{r}_2)\phi_\sigma(\mathbf{r}_2)$  that are centered at  $\mathbf{X}'$  ( $O_{l',m'}^{\lambda\sigma,\mathbf{X}'}$ ) to define multipole tensors  $O_{l',m'}^{\mathbf{X}'}$

for each box containing these distributions. Then, these multipole tensors are multiplied by the local expansions to give  $\mathcal{M}_{l,m}(\mathbf{X})$ ,

$$\mathcal{M}_{l,m}(\mathbf{X}) = (-1)^l \sum_{\mathbf{X}'} \sum_{l'm'} M_{l+l',m+m'}(\mathbf{X}-\mathbf{X}') O_{l',m'}^{\mathbf{X}'} \quad (9)$$

Note that the quantity  $\mathcal{M}_{l,m}(\mathbf{X})$  is defined for each of the boxes and includes all the distant Coulomb interactions. Using this, the far-field part of the Coulomb matrix is computed as

$$J_{\mu\nu}^{\text{ff}} = \sum_{lm} O_{l,m}^{\mu\nu,\mathbf{X}} \mathcal{M}_{l,m}(\mathbf{X}) \quad (10)$$

The Coulomb matrix elements associated with the neighboring charge distributions, that is, the near-field region, where the multipole expansion is no longer valid, are evaluated using standard algorithms. The cost of computing the near-field part is linearly scaling with respect to system size.

The summation in eq 9 is efficiently performed using a hierarchy of boxes that are constructed by first partitioning the system of interest into a number of boxes, each of which is further divided into smaller boxes and so on. This hierarchical structure allows the distant contributions to be computed at the coarse-grained levels (or higher levels, with larger and fewer boxes), and translated to the lower levels using the spherical harmonics addition theorem for scaled regular and irregular solid harmonics. The standard FMM algorithm consists of three steps: First, the multipoles are computed at the lowest level and translated upward<sup>28,31,40</sup>

$$O_{l,m}^{\mathbf{X}_p} = \sum_{l'm'} O_{l-l',m-m'}^{\mathbf{X}_c-\mathbf{X}_p} O_{l',m'}^{\mathbf{X}_c} \quad (11)$$

where  $\mathbf{X}_p$  and  $\mathbf{X}_c$  are the center of the parent and child boxes, respectively, and  $l \leq L_{\text{max}}$ . Second, the local expansions are obtained by translating the multipoles within the same level. To do so, each box has an interaction list that enumerates the non-neighboring boxes at the same level whose parents are its parent's neighbor. It reads<sup>31,40</sup>

$$\mathcal{M}_{l,m}(\mathbf{X}) = (-1)^l \sum_{l'm'} M_{l+l',m+m'}(\mathbf{X}-\mathbf{X}') O_{l',m'}^{\mathbf{X}'} \quad (12)$$

in which  $\mathbf{X}$  and  $\mathbf{X}'$  are the center of the box and that associated with those in the interaction list, respectively. Finally, the local expansions are translated downward,<sup>28,31,40</sup>

$$\mathcal{M}_{l,m}(\mathbf{X}_c) = \sum_{l'm'} M_{l',m'}(\mathbf{X}_p) O_{l-l',m-m'}^{\mathbf{X}_c-\mathbf{X}_p} \quad (13)$$

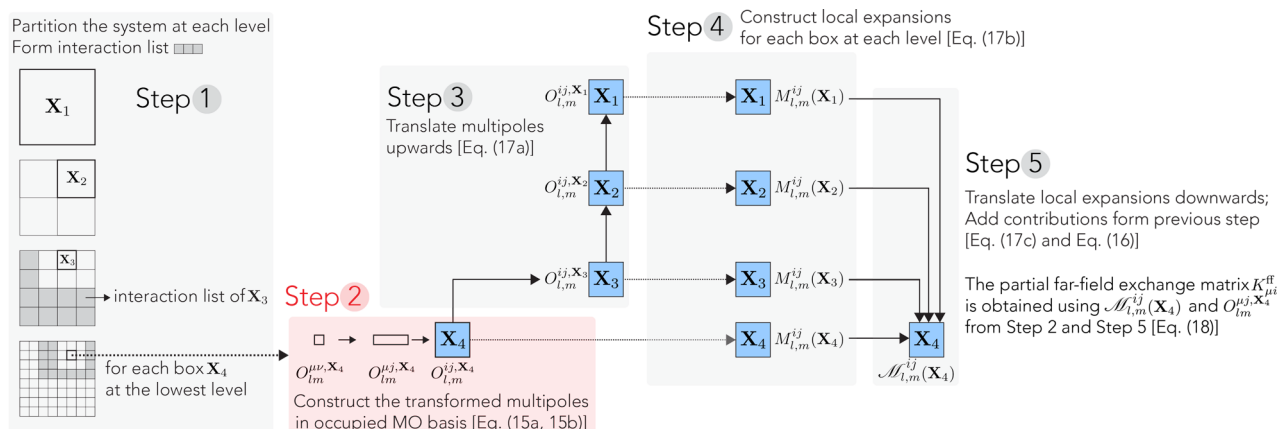
The local expansions containing the far-field interactions for all of the boxes at the lowest level are then collected to construct  $\mathcal{M}_{l,m}(\mathbf{X})$  in eq 9.

In this work, we have extended this algorithm to computation of partial exchange matrix elements (eq 5). Our new algorithm is termed occ-FMM-K. The molecular integrals that contribute to the occupied exchange matrix are written using the multipole approximation as

$$(\mu j | j i) = \sum_{lm} (-1)^l O_{l,m}^{\mu j,\mathbf{X}} \sum_{l'm'} M_{l+l',m+m'}(\mathbf{X}-\mathbf{X}') O_{l',m'}^{j i,\mathbf{X}'} \quad (14)$$

where MO transformed multipole integrals are defined as

$$O_{l,m}^{\mu j,\mathbf{X}} = \sum_{\mu} O_{l,m}^{\mu\nu,\mathbf{X}} C_{\nu j} \quad (15a)$$



**Figure 1.** Schematic representation of the occ-FMM-K algorithm for constructing  $K_{\mu\nu}^{ff}$  using the translation relations up and down the FMM hierarchy. Step 2 is the essence of occ-FMM-K.

$$O_{l,m}^{ij,X} = \sum_{\mu} O_{l,m}^{\mu j,X} C_{\mu} \quad (15b)$$

It is important to stress that one of the multipoles in eq 14 is fully transformed to the MO basis; therefore, its size remains the same at the coarse-grained level, allowing us to evaluate it using the FMM algorithm with a hierarchy of boxes. This step is necessary for FMM algorithm to be used in the evaluation of far-field exchange contribution, as the multipole tensor of the parent box is constructed by summing up the multipole tensors of the children. Without transforming to the MO basis, this cannot be achieved, which is the reason why FMM is traditionally considered incompatible with exchange evaluation, and only a simple multipole approximation without the hierarchy structure has been used in exchange evaluation.

The traditional FMM algorithm is modified as follows (see graphical explanation in Figure 1). First, for each box at the lowest level, we compute  $O_{l,m}^{\mu,\nu,X}$  and transform them to the MO basis,  $O_{l,m}^{ij,X}$  using eq 15. We then compute  $M_{l,m}^{ij}(X)$  that is analogous to eq 9,

$$M_{l,m}^{ij}(X) = (-1)^l \sum_{X'} \sum_{l'm'} M_{l+l',m+m'}(X-X') O_{l',m'}^{ij,X'} \quad (16)$$

Note that the summation over  $X'$  in this equation is essential for utilizing the translations in FMM discussed above. When computing  $M_{l,m}^{ij}(X)$ , we use the same algorithm as the traditional FMM, namely, those based on eqs 11–13 for each pair of  $i$  and  $j$

$$O_{l,m}^{ij,X_p} = \sum_{l'm'} O_{l-l',m-m'}^{ij,X_c} O_{l',m'}^{ij,X_c} \quad (17a)$$

$$M_{l,m}^{ij}(X) = (-1)^l \sum_{l'm'} M_{l+l',m+m'}(X-X') O_{l',m'}^{ij,X'} \quad (17b)$$

$$M_{l,m}^{ij}(X_c) = \sum_{l'm'} M_{l',m'}(X_p) O_{l-l',m-m'}^{ij,X_c} \quad (17c)$$

The occupied-orbital exchange matrix is then computed as

$$K_{\mu\nu}^{ff} = \sum_j \sum_{lm} O_{l,m}^{\mu j,X} M_{l,m}^{ij}(X) \quad (18)$$

The near-field contributions to the occupied-orbital exchange matrix can be computed simultaneously with those to the Coulomb matrix with marginal additional costs.

There are a number of parameters required to perform FMM calculations, and some are dependent on the system of interest. The number of levels or depth ( $N_s$ ) in FMM is typically chosen to be 4 or 5, such that the length of the smallest box is about 3.0 bohr. This number determines the total number of boxes as well as the size and number of boxes at the lowest level and, therefore, affects the efficiency of FMM. The definition of the near- and far-field regions depends of a number of parameters, for which interested readers can refer to refs 32, 40, and 35. Our implementation makes explicit use of contracted basis functions to optimize the computation of the electron repulsion integrals in the near-field region. The definition of the extent of each distribution used to determine the near- and far-field regions is described in refs 46 and 47. The “well-separatedness” index ( $ws$ ) is typically chosen to be 0 such that two charge distributions are considered non-overlapping if the distance between their centers is greater than the sum of their extents. However, this parameter can be tuned depending on the definition of the extents and the systems studied.

### 3. NUMERICAL RESULTS

In this section, we first show the convergence of the Coulomb and exchange energy contributions with respect to multipole ranks. We then present the parallel scaling of our algorithm, followed by the timing data using the optimized parameters.

#### 3.1. Convergence with Respect to Multipole Ranks.

We examined the convergence of the Hartree–Fock energy with respect to the ranks of multipole expansions,  $L_{\max}^J$  and  $L_{\max}^K$  for two graphene sheets  $C_{150}H_{30}$  and  $C_{96}H_{24}$ . We chose these systems because the exchange contributions in graphene sheets have been shown to be slowly decaying with distance and proved to be more challenging than systems frequently used in FMM studies, such as hydrocarbon chains, polyglycine, and water clusters.<sup>7</sup> The def2-SVP basis set was used. Note that the vanishing gap is observed as the sheet gets larger, and Hartree–Fock has been known to overestimate band gaps in molecular systems. We set the FMM parameters to be  $N_s = 5$ ,  $ws = 0.0$  for  $C_{150}H_{30}$  and  $N_s = 5$ ,  $ws = -0.1$  for  $C_{96}H_{24}$ . The Schwarz integral screening and SCF convergence thresholds were set to  $1.0 \times 10^{-8}$ . In the reference calculation, the multipole expansions were truncated at  $L_{\max}^J = 15$  and  $L_{\max}^K = 5$  for the far-field Coulomb and exchange interactions, respectively. The convergence was analyzed by comparing the reference energy and that computed from the Fock operator

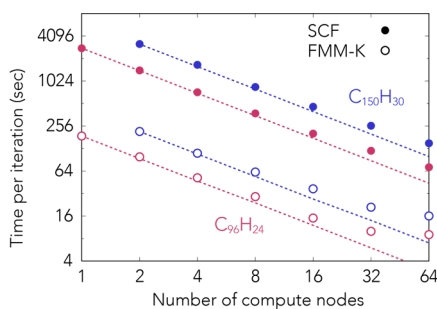
**Table 1.** Convergence of the Energy with Respect to Multipole Ranks  $L_{\max}^J$  and  $L_{\max}^K$  for the Graphene Sheet  $C_{150}H_{30}$  Using the def2-SVP Basis Set<sup>a</sup>

$L_{\max}^J$	no far-field exchange	$L_{\max}^K = 0$	$L_{\max}^K = 1$	$L_{\max}^K = 2$	$L_{\max}^K = 3$	$L_{\max}^K = 4$
0	-72782.004	-72785.684	-72785.818	-72785.811	-72785.811	-72785.811
1	-3717.595	-3721.276	-3721.409	-3721.402	-3721.402	-3721.402
2	-9.862	-13.543	-13.677	-13.670	-13.669	-13.669
3	20.290	16.610	16.476	16.483	16.483	16.483
4	2.832	-0.849	-0.982	-0.976	-0.975	-0.975
5	3.425	-0.256	-0.390	-0.383	-0.382	-0.382
6	3.786	0.106	-0.028	-0.021	-0.021	-0.021
7	3.812	0.132	-0.002	0.005	0.005	0.005
8	3.809	0.128	-0.005	0.002	0.002	0.002
9	3.807	0.127	-0.007	0.000	0.000	0.000
10	3.807	0.126	-0.007	0.001	0.000	0.000

<sup>a</sup>Errors are shown in  $mE_h$  with respect to the reference energy,  $E - E_{\text{ref}}$ , computed using  $L_{\max}^J = 15$  and  $L_{\max}^K = 5$  ( $E_{\text{ref}} = -5694.58434391 E_h$ ).

constructed using the reference MO coefficients and different values for  $L_{\max}^J$  and  $L_{\max}^K$ . The results for  $C_{150}H_{30}$  are shown in Table 1. Similar results were obtained for  $C_{96}H_{24}$ .<sup>48</sup> The errors decay quickly for both the Coulomb and exchange contributions as higher-rank multipoles are included. However, since the magnitude of the far-field exchange contribution ( $\sim 4 mE_h$ ) is a few orders of magnitude smaller than that of the far-field Coulomb contribution,  $L_{\max}^K$  can be smaller than  $L_{\max}^J$ , thus significantly reducing the computational cost at almost no loss in accuracy. It is worth noting that the error in the far-field exchange contributions is around  $1 \mu E_h$  with  $L_{\max}^K = 2$  for this challenging system. From these results, we concluded that the multipole series should be truncated at  $L_{\max}^J = 10$  for the Coulomb interaction and at  $L_{\max}^K = 2$  for the exchange interaction to achieve  $\mu E_h$  accuracy.

**3.2. Parallel Scalability.** Our algorithm can be trivially parallelized with very high efficiency, making it useful for large-scale problems. We measured the strong parallel scaling using the graphene sheets  $C_{96}H_{24}$  and  $C_{150}H_{30}$ . The FMM parameters used for these calculations were  $N_s = 5$ ,  $ws = 0$ ,  $L_{\max}^J = 10$ , and  $L_{\max}^K = 2$ . The Schwarz integral screening threshold was set to  $1.0 \times 10^{-8}$ . The results are shown in Figure 2. Calculations

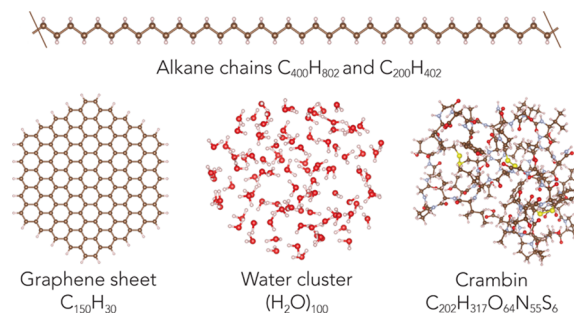


**Figure 2.** Timings for graphene sheets  $C_{96}H_{24}$  and  $C_{150}H_{30}$  using def2-SVP and the parameters  $N_s = 5$ ,  $L_{\max}^J = 10$ , and  $L_{\max}^K = 2$ . Each compute node consists of 2 Xeon E5-2650 CPUs (Sandy Bridge 2.0 GHz).

were performed using the def2-SVP basis set on a 64-node computer cluster, where each compute node consists of 2 Xeon E5-2650 CPUs (Sandy Bridge 2.0 GHz). Total timings for an SCF iteration and timings for the far-field exchange evaluation were averaged over the first 5 iterations. The cost of the far-field exchange evaluation for  $C_{96}H_{24}$ , which was about 10% of the total cost per SCF iteration, was 188 s with 1 compute node,

and reduced to 99, 52, 29, 15, 10, and 9 s using 2, 4, 8, 16, 32, and 64 compute nodes. The timings for  $C_{150}H_{30}$  were 216, 111, 62, 37, 21, and 16 s using 2, 4, 8, 16, 32, and 64 compute nodes. For  $C_{96}H_{24}$ , the strong scalings from 1 to 64 compute nodes for an SCF iteration and far-field exchange evaluation were found to be 61% and 33%, respectively. Those for  $C_{150}H_{30}$  from 2 to 64 nodes were 66% and 42%.

The good scaling for the far-field exchange evaluation is due to the fact that the transformation of the multipole tensors from the AO basis  $O_{l,m}^{\mu,\nu,X}$  to the occupied MO basis  $O_{l,m}^{j,X}$  in step 2 of the occ-FMM-K algorithm (Figure 1) can be done independently in batches of occupied-orbital index  $j$ . As a result, the upward and downward translations of the multipoles and local expansions in the occupied MO basis in steps 3–5 are well distributed. The construction of the partial exchange matrix from the multipoles and local expansions (eq 18) is also similarly parallelized. Larger systems would exhibit better parallel scaling because the matrices become large enough that the peak performance of linear algebra routines is achieved even when a large number of nodes are used. The near-field Coulomb and exchange contributions are calculated with exact four-center integrals at the moment and accounts for most of the differences between the total timing for an SCF iteration and the time taken for the far-field exchange evaluation.



**Figure 3.** Systems used for timing benchmarks in this work.

**3.3. Timing Data.** The performance of our occ-FMM-K implementation is assessed for a number of molecular systems (shown in Figure 3) using the def2-SVP basis set. The results are compiled in Table 2. The parameters used in all of the timing benchmark calculations are  $ws = 0$ ,  $L_{\max}^J = 10$ , and  $L_{\max}^K = 2$ . We used  $N_s = 5$  for the graphene sheet and crambin,  $N_s = 4$  for the water cluster, while for the alkane chains, we fixed the smallest box size at the lowest level to be 3.0 bohr. The Schwarz

Table 2. Wall Time (min) for Calculating the Far-Field Exchange and Coulomb Contributions Using the FMM Algorithms<sup>a</sup>

system		atoms	electrons	basis <sup>b</sup>	far-field <i>K</i>	far-field <i>J</i>	near-field	diag.	SCF iter.
alkane chain	C <sub>200</sub> H <sub>402</sub>	602	1602	4810	0.36	0.10	0.66	0.44	1.50
alkane chain	C <sub>400</sub> H <sub>802</sub>	1202	3202	9610	2.76	0.61	1.59	1.94	5.50
graphene sheet	C <sub>150</sub> H <sub>30</sub>	180	930	2250	0.26	0.09	1.98	0.15	2.52
water cluster	(H <sub>2</sub> O) <sub>100</sub>	300	1000	2400	0.75	0.53	0.64	0.23	2.54
crambin	C <sub>202</sub> H <sub>317</sub> O <sub>64</sub> N <sub>55</sub> S <sub>6</sub>	644	2522	6187	1.38	0.16	38.75	0.99	41.23

<sup>a</sup>The timings for near-field and diagonalization, as well as the total timing for an SCF iteration are also shown. 128 Xeon E5-2650 CPUs (Sandy Bridge 2.0 GHz, total 1024 cores) with InfiniBand QDR were used. <sup>b</sup>The def2-SVP basis set was used.

integral screening threshold was set to  $1.0 \times 10^{-8}$ . In principle, the sets of the FMM parameters used for the evaluation of far-field Coulomb and exchange contributions can be different. We have not yet investigated how the parameters besides  $L_{\max}$  can be optimized to achieve maximum efficiency without loss of accuracy. It is, however, expected that the optimal parameters used for the Coulomb interaction will be different from those used for the exchange interaction as the Coulomb interaction is longer-range, and the cost of evaluating the Coulomb contribution is significantly smaller. This will be investigated in the future.

We included the one-dimensional alkane chains C<sub>200</sub>H<sub>402</sub> and C<sub>400</sub>H<sub>802</sub> as examples, because FMM is known to perform most efficiently for one-dimensional systems (even though the far-field exchange contributions to the total energies for these particular systems are negligible). This efficiency is due to the fact that the fraction of boxes in the near field does not change with system size in one-dimension. From a 200-carbon chain (4810 basis functions) to a 400-carbon chain (9610 basis functions), the total timing for an SCF iteration increased from 1.5 to 5.5 min. In the first case C<sub>200</sub>H<sub>402</sub>, the cost of computing the far-field exchange contribution was only a fraction of that for the near-field contributions, amounting to 24% of the total timing per SCF interaction. Evaluation of the exchange contribution becomes more costly for the longer chain C<sub>400</sub>H<sub>802</sub>, taking 50% of the total timing per SCF iteration. The cost of computing the far-field Coulomb contribution was small.

Next we performed a calculation for a two-dimensional graphene sheet C<sub>150</sub>H<sub>30</sub>. As mentioned previously, this is considered among the most challenging systems for exchange computation, because the exchange interaction decays slowly with distance. Note that this example was the largest two-dimensional system used in the benchmarks by Burant and Scuseria<sup>7</sup> for their NFX method that accounts for the far-field exchange contributions by simply increasing the size of the near-field FMM. For this example, the cost of the far-field exchange evaluation using our algorithm was around 0.3 min, which was 12% of the total cost for an SCF iteration (2.5 min). The remaining cost is largely due to the near-field four-center integral evaluation and diagonalization of the Fock matrix.

Finally, the timings are reported for a water cluster (H<sub>2</sub>O)<sub>100</sub> (ref 48.) and a small protein Crambin C<sub>202</sub>H<sub>317</sub>O<sub>64</sub>N<sub>55</sub>S<sub>6</sub> to assess the performance of our algorithm for three-dimensional systems. The latter was previously used to benchmark the DFT and DLPNO-CCSD(T) methods.<sup>3,49,50</sup> The cost of far-field exchange evaluation was 32% and 3% of that of an SCF iteration for the water cluster and crambin, respectively. Similar to the previous examples, a large portion of the remaining cost is attributed to the near-field four-center integral evaluation.

These results, together with the excellent parallel scaling of our algorithm, are highly encouraging. It is also worth noting

that (1) the cost of the near-field computation can be further reduced using, for example, the RI approximation; and (2) the use of localized orbitals and screening of occupied-orbital pairs would significantly reduce the cost of the far-field exchange evaluation. The memory requirement for large calculations is determined at the moment by the size of the multipole and local expansion tensors for each box at the lowest level.

#### 4. CONCLUSION

In this Article, we introduced an efficient FMM-based algorithm (named occ-FMM-K) for evaluating the exact exchange matrix elements that contribute to the energy and orbital-rotation gradient at the mean-field level. This is done by constructing the partial exchange matrix  $K_{\mu\nu}$  where all matrix elements have at least one occupied-orbital index. The multipole and local expansion tensors are first transformed into the occupied-orbital basis. The upward and downward translations of these tensors are then performed in exactly the same manner as conventional FMM for the Coulomb interaction. Efficient parallelization and the fact that there is no assumption on the sparsity of the density or multipole matrices make this algorithm attractive for large and extended systems, especially those with small band gaps. It is important to note that existing exchange algorithms which take advantage of this sparsity are expected to perform as well in nondelocalized systems.

There are, however, a number of ways to further improve our algorithm, which currently scales cubically with system sizes. First, it is possible to reduce the cost of the far-field exchange evaluation for many systems by using localized molecular orbitals and screening occupied-orbital pairs. This would significantly mitigate the cost of storage and basis transformations. Our preliminary results show that scaling can be improved (see Supporting Information). Second, the expensive near-field integral evaluation can be replaced by an algorithm based on the RI approximation. In addition, extensions of our algorithm to complete active space self-consistent field (CASSCF) and configuration interaction singles (CIS) should be straightforward. These improvements and extensions will be investigated in the near future.

Another direction is application of occ-FMM-K with large basis sets, which is currently hampered due to a technical reason (i.e., numerical instability with the current implementation). This is interesting, because the cost of the upward and downward translations in the occ-FMM-K algorithm (steps 3 and 4 in Figure 1) only depends on the number of occupied orbitals (and not the number of basis functions), implying that our algorithm is expected to be more competitive for larger basis sets. The FMM algorithms, however, become less effective when highly diffuse functions are present, and so does the occ-FMM-K algorithm. This will be addressed in future works.

**■ ASSOCIATED CONTENT****📄 Supporting Information**

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.7b00880.

Cartesian coordinates for all the systems studied in this work, as well as additional results (PDF)

**■ AUTHOR INFORMATION****Corresponding Author**

\*E-mail: anh@u.northwestern.edu.

**ORCID**

Hai-Anh Le: 0000-0002-7492-2310

**Notes**

The authors declare no competing financial interest.

**■ ACKNOWLEDGMENTS**

We thank Dr. Jae Woo Park for providing the geometry of the water cluster. This work has been supported by National Science Foundation ACI-1550481 (H.-A.L.) and CHE-1351598 (T.S.). T.S. is an Alfred P. Sloan Fellow.

**■ REFERENCES**

- (1) Pulay, P. Localizability of dynamic electron correlation. *Chem. Phys. Lett.* **1983**, *100*, 151–154.
- (2) Schütz, M.; Werner, H.-J. Low-order scaling local electron correlation methods. IV. Linear scaling local coupled-cluster (LCCSD). *J. Chem. Phys.* **2001**, *114*, 661–681.
- (3) Riplinger, C.; Pinski, P.; Becker, U.; Valeev, E. F.; Neese, F. Sparse maps-A systematic infrastructure for reduced-scaling electronic structure methods. II. Linear scaling domain based pair natural orbital coupled cluster theory. *J. Chem. Phys.* **2016**, *144*, 024109.
- (4) Rebolini, E.; Izsák, R.; Reine, S. S.; Helgaker, T.; Pedersen, T. B. Comparison of Three Efficient Approximate Exact-Exchange Algorithms: The Chain-of-Spheres Algorithm, Pair-Atomic Resolution-of-the-Identity Method, and Auxiliary Density Matrix Method. *J. Chem. Theory Comput.* **2016**, *12*, 3514–3522.
- (5) Ochsenfeld, C.; White, C. A.; Head-Gordon, M. Linear and sublinear scaling formation of Hartree–Fock-type exchange matrices. *J. Chem. Phys.* **1998**, *109*, 1663–1669.
- (6) Ochsenfeld, C. Linear scaling exchange gradients for Hartree–Fock and hybrid density functional theory. *Chem. Phys. Lett.* **2000**, *327*, 216–223.
- (7) Burant, J. C.; Scuseria, G. E.; Frisch, M. J. A linear scaling method for Hartree–Fock exchange calculations of large molecules. *J. Chem. Phys.* **1996**, *105*, 8969–8972.
- (8) Schwegler, E.; Challacombe, M. Linear scaling computation of the Fock matrix. IV. Multipole accelerated formation of the exchange matrix. *J. Chem. Phys.* **1999**, *111*, 6223–6229.
- (9) Schwegler, E.; Challacombe, M. Linear scaling computation of the Hartree–Fock exchange matrix. *J. Chem. Phys.* **1996**, *105*, 2726–2734.
- (10) Schwegler, E.; Challacombe, M.; Head-Gordon, M. Linear scaling computation of the Fock matrix. II. Rigorous bounds on exchange integrals and incremental Fock build. *J. Chem. Phys.* **1997**, *106*, 9708–9717.
- (11) Maurer, S. A.; Lambrecht, D. S.; Flaig, D.; Ochsenfeld, C. Distance-dependent Schwarz-based integral estimates for two-electron integrals: Reliable tightness vs. rigorous upper bounds. *J. Chem. Phys.* **2012**, *136*, 144107.
- (12) Maurer, S. A.; Lambrecht, D. S.; Kussmann, J.; Ochsenfeld, C. Efficient distance-including integral screening in linear-scaling Møller-Plesset perturbation theory. *J. Chem. Phys.* **2013**, *138*, 014101.
- (13) Thompson, T. H.; Ochsenfeld, C. Distance-including rigorous upper bounds and tight estimates for two-electron integrals over long- and short-range operators. *J. Chem. Phys.* **2017**, *147*, 144101.
- (14) 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.
- (15) Köppl, C.; Werner, H.-J. Parallel and Low-Order Scaling Implementation of Hartree–Fock Exchange Using Local Density Fitting. *J. Chem. Theory Comput.* **2016**, *12*, 3122–3134.
- (16) Jung, Y.; Sodt, A.; Gill, P. M. W.; Head-Gordon, M. Auxiliary basis expansions for large-scale electronic structure calculations. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6692–6697.
- (17) Izmaylov, A. F.; Scuseria, G. E.; Frisch, M. J. Efficient evaluation of short-range Hartree–Fock exchange in large molecules and periodic systems. *J. Chem. Phys.* **2006**, *125*, 104103.
- (18) Paier, J.; Diaconu, C. V.; Scuseria, G. E.; Guidon, M.; VandeVondele, J.; Hutter, J. Accurate Hartree–Fock energy of extended systems using large Gaussian basis sets. *Phys. Rev. B* **2009**, *80*, 174114.
- (19) Guidon, M.; Hutter, J.; VandeVondele, J. Robust Periodic Hartree–Fock Exchange for Large-Scale Simulations Using Gaussian Basis Sets. *J. Chem. Theory Comput.* **2009**, *5*, 3010–3021.
- (20) 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.
- (21) Guidon, M.; Hutter, J.; VandeVondele, J. Auxiliary Density Matrix Methods for Hartree–Fock Exchange Calculations. *J. Chem. Theory Comput.* **2010**, *6*, 2348–2364.
- (22) Merlot, P.; Kjargaard, 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.
- (23) Hollman, D. S.; Schaefer, H. F.; Valeev, E. F. Semi-exact concentric atomic density fitting: Reduced cost and increased accuracy compared to standard density fitting. *J. Chem. Phys.* **2014**, *140*, 064109.
- (24) 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.
- (25) 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.
- (26) Lewis, C. A.; Calvin, J. A.; Valeev, E. F. Clustered Low-Rank Tensor Format: Introduction and Application to Fast Construction of Hartree–Fock Exchange. *J. Chem. Theory Comput.* **2016**, *12*, 5868–5880.
- (27) Lin, L. Adaptively Compressed Exchange Operator. *J. Chem. Theory Comput.* **2016**, *12*, 2242–2249.
- (28) Greengard, L. *The Rapid Evaluation of Potential Fields in Particle Systems*; MIT Press: Cambridge, MA, 1987.
- (29) Greengard, L.; Rokhlin, V. A Fast Algorithm for Particle Simulations. *J. Comput. Phys.* **1987**, *73*, 325–348.
- (30) Greengard, L. Fast Algorithms for Classical Physics. *Science* **1994**, *265*, 909–914.
- (31) White, C. A.; Head-Gordon, M. Derivation and efficient implementation of the fast multipole method. *J. Chem. Phys.* **1994**, *101*, 6593–6605.
- (32) White, C. A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M. Linear scaling density functional calculations via the continuous fast multipole method. *Chem. Phys. Lett.* **1994**, *230*, 8–16.
- (33) Petersen, H. G.; Soelvason, D.; Perram, J. W.; Smith, E. R. The very fast multipole method. *J. Chem. Phys.* **1994**, *101*, 8870–8876.
- (34) Kutteh, R.; Aprà, E.; Nichols, J. A generalized fast multipole approach for Hartree–Fock and density functional computations. *Chem. Phys. Lett.* **1995**, *238*, 173–179.
- (35) Strain, M. C.; Scuseria, G. E.; Frisch, M. J. Achieving Linear Scaling for the Electronic Quantum Coulomb Problem. *Science* **1996**, *271*, 51–53.
- (36) Pérez-Jordá, J. M.; Yang, W. A concise redefinition of the solid spherical harmonics and its use in fast multipole methods. *J. Chem. Phys.* **1996**, *104*, 8003–8006.

- (37) 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.
- (38) Challacombe, M.; White, C.; Head-Gordon, M. Periodic boundary conditions and the fast multipole method. *J. Chem. Phys.* **1997**, *107*, 10131–10140.
- (39) Kudin, K. N.; Scuseria, G. E. Revisiting infinite lattice sums with the periodic fast multipole method. *J. Chem. Phys.* **2004**, *121*, 2886–2890.
- (40) Choi, C. H.; Ruedenberg, K.; Gordon, M. S. New Parallel Optimal-Parameter Fast Multipole Method (OPFMM). *J. Comput. Chem.* **2001**, *22*, 1484–1501.
- (41) Watson, M. A.; Salek, P.; Macak, P.; Helgaker, T. Linear-scaling formation of Kohn-Sham Hamiltonian: Application to the calculation of excitation energies and polarizabilities of large molecular systems. *J. Chem. Phys.* **2004**, *121*, 2915–2931.
- (42) Rudberg, E.; Salek, P. Efficient implementation of the fast multipole method. *J. Chem. Phys.* **2006**, *125*, 084106.
- (43) Toivanen, E. A.; Losilla, S. A.; Sundholm, D. The grid-based fast multipole method—a massively parallel numerical scheme for calculating two-electron interaction energies. *Phys. Chem. Chem. Phys.* **2015**, *17*, 31480–31490.
- (44) BAGEL, Brilliantly Advanced General Electronic-structure Library. <http://www.nubakery.org> under the GNU General Public License.
- (45) Shiozaki, T. BAGEL: Brilliantly Advanced General Electronic-structure Library. *WIREs Comput. Mol. Sci.* **2018**, *8*, e1331.
- (46) Pérez-Jordá, J. M.; Yang, W. Fast evaluation of the Coulomb energy for electron densities. *J. Chem. Phys.* **1997**, *107*, 1218–1226.
- (47) 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*, 9136–9148.
- (48) The geometry of all the systems studied in this work as well as additional data can be found in [Supporting Information](#).
- (49) Riplinger, C.; Sandhoefer, B.; Hansen, A.; Neese, F. Natural triple excitations in local coupled cluster calculations with pair natural orbitals. *J. Chem. Phys.* **2013**, *139*, 134101.
- (50) Furche, F.; Perdew, J. P. The performance of semilocal and hybrid density functionals in transition-metal chemistry. *J. Chem. Phys.* **2006**, *124*, 044103.