OptOrbFCI-CD: Combining optimal orbital selection and Cholesky decomposition for accurate and efficient full configuration interaction

Zhiyuan Zhang,[1] Zhenlin Zhang,[1] Yingzhou Li,[2, a)] Wei Hu,[3, b)] and Jinlong Yang[3, c)]
[1)]*School of Emerging Technology, Hefei National Research Center for Physical Sciences at the Microscale, and Hefei National Laboratory, University of Science and Technology of China, Hefei, Anhui 230026, China*
[2)]*School of Mathematical Sciences, Fudan University, Shanghai, China*
[3)]*State Key Laboratory of Precision and Intelligent Chemistry, University of Science and Technology of China, Hefei, Anhui 230026, China*

Full configuration interaction (FCI) remains an indispensable method for strongly correlated electronic structure calculations due to its high accuracy, but its steeply scaling computational cost limits applications with large basis sets. We propose an improved optimal orbital selection algorithm for full configuration interaction with Cholesky decompostion (OptOrbFCI-CD) that significantly enhances computational efficiency without sacrificing accuracy in large basis set calculations. The computational complexities of the algorithm are systematically analyzed before and after applying Cholesky decomposition, demonstrating significant time reduction for large systems. Numerical results for H_4 molecules are presented, validating the algorithm's effectiveness across varying basis set sizes. Our findings suggest that although the energy continues to decrease with increasing basis set size, the convergence rate progressively slows without reaching full saturation. The improved algorithm exhibits strong potential for scaling to larger systems while maintaining an optimal balance between computational cost and accuracy.

I. INTRODUCTION

Understanding the quantum behavior of many-body systems is a cornerstone of quantum chemistry and condensed matter physics. Accurate solutions to the electronic structure problem are essential for applications in material design, drug discovery, and molecular simulations. While approximate methods such as density functional theory (DFT)[1,2] and coupled cluster (CC)[3,4] theory are effective for weakly correlated systems, strongly correlated systems pose significant challenges due to their inherent complexity.

Over the past few decades, significant advancements have been achieved in developing methods to tackle the full configuration interaction (FCI) problem, which remains a cornerstone of quantum chemistry due to its ability to provide an exact solution within a given basis set. These methods have significantly accelerated FCI computations compared to traditional approaches, often achieving orders-of-magnitude improvements in efficiency. Among the most prominent FCI solvers, the density matrix renormalization group (DMRG)[5–7] has shown remarkable success by representing the variational wave function as a tensor train. While DMRG is highly effective for low-dimensional systems, its extension to three-dimensional systems is severely limited by the exponential increase in computational complexity. Stochastic approaches, such as full configuration interaction quantum Monte Carlo (FCIQMC)[8–10], approximate the ground-state variational wave function using an empirical distribution of a large number of stochastic walkers. To reduce the variance in energy estimations and lower the required number of walkers, advanced variants like initiator FCIQMC (iFCIQMC)[11,12] and semistochastic FCIQMC (S-FCIQMC)[13,14] have been developed, striking a balance between variance and bias. Dynamic configuration selection methods, including configuration interaction by perturbatively selecting iteration (CIPSI)[15,16], adaptive configuration interaction (ACI)[17,18], adaptive sampling configuration interaction (ASCI)[19,20], heat-bath configuration interaction (HCI)[21], and stochastic HCI (SHCI)[22], dynamically identify and select the most important configurations based on various perturbation approximations. These methods use traditional eigensolvers combined with post-perturbation corrections to efficiently estimate the ground-state energy. Optimization-based methods have also emerged as powerful tools for FCI problems. For example, the coordinate descent full configuration interaction (CDFCI)[23,24] reformulates the FCI problem as an unconstrained optimization task and solves it variationally using a coordinate descent method with hard thresholding. The systematic full configuration interaction fast randomized iteration (sFCI-FRI)[25] framework introduces a hierarchical factorization and randomized iteration scheme to further reduce computational costs.

In addition to these specialized methods, several algorithms[26–33] inspired by advancements in numerical linear algebra have been developed to address the FCI problem. Despite their diversity and significant progress, all these approaches face a fundamental limitation: for extremely large basis sets, such as those exceeding 100 orbitals, the computational cost becomes prohibitively high, often rendering calculations infeasible due to the exponential scaling with respect to the basis set size. The full CI energy is invariant to the choice of orbitals, however when considering all excitations within an active space such as Complete Active Space CI (CASCI) the choice of orbitals affects the total energy, recent advancements in the complete active space self consistent field method (CASSCF)[34–53] have shown its efficiency in quantum chemistry calculations. By combining matrix factorization techniques with CASSCF, further improvements in computational efficiency can be achieved. This approach[44,54–67] is motivated

[a)]Electronic mail: yingzhouli@fudan.edu.cn
[b)]Electronic mail: whuustc@ustc.edu.cn
[c)]Electronic mail: jlyang@ustc.edu.cn

by the inherent low-rank structure of two-electron integrals, which allows for significant dimension reduction while preserving numerical accuracy. This combined approach helps optimize matrix operations during the calculation process, particularly when dealing with large basis sets, significantly reducing the computational resources and time required. The optimal orbital selection for FCI (OptOrbFCI)[68] algorithm addresses this challenge by projecting the original orbital space onto a reduced subspace. This projection is achieved by optimizing a matrix that minimizes the ground-state energy in the reduced variational space. While OptOrbFCI is effective in reducing the size of the basis set, the computational cost associated with optimizing the projection matrix remains a significant bottleneck, as it involves tensor contractions.

In this work, we introduce a Cholesky decompositionm (CD) strategy, a matrix factorization method that decomposes a symmetric positive definite matrix into the product of a lower triangular matrix and its transpose to accelerate the optimization phase of OptOrbFCI, thereby enhancing computational efficiency and ensuring numerical stability in linear algebra operations. By exploiting the low-rank structure of the two-electron integrals, the computational complexity of the projection matrix optimization is significantly reduced. Numerical experiments show this improvement makes computations several times faster while maintaining the original algorithm's accuracy

The remainder of this paper is organized as follows. Section 2 details the mathematical formulation of the OptOrbFCI algorithm and proposes the Cholesky decomposition strategy. Section 3 presents numerical results that validate the efficiency and accuracy of the improved algorithm. Finally, Section 4 concludes the paper and discusses potential future directions.

II. METHOD

In this section, we introduce the optimal orbital selection for FCI (OptOrbFCI) and analyze the extent to which this algorithm reduces computational cost compared to previous methods.

A. Preparation

For a given large orbital set $\{\psi_1,\ldots,\psi_M\}$, the associated Hamiltonian operator in the second quantization (M and N denote the number of the given molecular orbitals and the computationally affordable number of orbitals ($N<M$)) is

$$\hat{H}=\sum_{p,q=1}^{M} h_{pq}\hat{c}_p^\dagger\hat{c}_q+\frac{1}{2}\sum_{p,q,r,s=1}^{M} v_{pqrs}\hat{c}_p^\dagger\hat{c}_q^\dagger\hat{c}_s\hat{c}_r, \tag{1}$$

where $\hat{c}_p^\dagger$ and $\hat{c}_q$ are creation and annihilation operators associated with ψ_p and ψ_q repectively. The one-electron and two-electron integrals, h_{pq} and v_{pqrs}, admit the following expressions:

$$h_{pq}=\int \mathrm{d}\boldsymbol{x_1}\psi_p^*(\boldsymbol{x_1})h(\boldsymbol{x_1})\psi_q(\boldsymbol{x_1}), \tag{2}$$

$$v_{pqrs}=\int \mathrm{d}\boldsymbol{x_1}\mathrm{d}\boldsymbol{x_2}\psi_p^*(\boldsymbol{x_1})\psi_q^*(\boldsymbol{x_2})v(\boldsymbol{x_1},\boldsymbol{x_2})\psi_s(\boldsymbol{x_2})\psi_r(\boldsymbol{x_1}), \tag{3}$$

where $h(\boldsymbol{x_1})$ and $v(\boldsymbol{x_1},\boldsymbol{x_2})$ are the one-body and two-body operators respectively. Due to the limited memory and computational power, we are only able to solve FCI problems under N orbitals. Hence, we introduce a partial unitary matrix $U\in\mathscr{U}(M,N)$, where $\mathscr{U}(M,N)$ is the space of all partial unitary matrix of size M by N, i.e.,

$$\mathscr{U}(M,N)=\{U\in\mathbb{R}^{M\times N}|U^\top U=I_N\}, \tag{4}$$

and I_N denotes the identity matrix of size N by N. The transformed orbitals from $\{\psi_1,\ldots,\psi_M\}$ via U are denoted as $\{\phi_1,\ldots,\phi_N\}$ such that

$$\phi_i=\sum_{j=1}^{M}\psi_j U_{ji}, \tag{5}$$

where U_{ji} denotes the element of j-th row and i-th column of U. The Hamiltonian operator associated with $\{\phi_1,\ldots,\phi_N\}$ is

$$\tilde{H}=\sum_{p',q'=1}^{N}\tilde{h}_{p'q'}\hat{d}_{p'}^\dagger\hat{d}_{q'}+\frac{1}{2}\sum_{p',q',r',s'=1}^{N}\tilde{v}_{p'q'r's'}\hat{d}_{p'}^\dagger\hat{d}_{q'}^\dagger\hat{d}_{s'}\hat{d}_{r'}, \tag{6}$$

where $\hat{d}_{p'}^\dagger$ and $\hat{d}_{q'}$ are the creation and annihilation operators associated with $\phi_{p'}$ and $\phi_{q'}$ respectively, the one-electron integral $\tilde{h}_{p'q'}$ is

$$\begin{aligned}\tilde{h}_{p'q'}&=\int \mathrm{d}\boldsymbol{x_1}\phi_{p'}^*(\boldsymbol{x_1})h(\boldsymbol{x_1})\phi_{q'}(\boldsymbol{x_1})\\&=\sum_{p,q=1}^{M}h_{pq}U_{pp'}U_{qq'}\end{aligned} \tag{7}$$

and the two-electron integral $\tilde{v}_{p'q'r's'}$ is

$$\begin{aligned}\tilde{v}_{p'q'r's'}&=\int \mathrm{d}\boldsymbol{x_1}\mathrm{d}\boldsymbol{x_2}\phi_{p'}^*(\boldsymbol{x_1})\phi_{q'}^*(\boldsymbol{x_2})v(\boldsymbol{x_1},\boldsymbol{x_2})\phi_{s'}(\boldsymbol{x_2})\phi_{r'}(\boldsymbol{x_1})\\&=\sum_{p,q,r,s=1}^{M}v_{pqrs}U_{pp'}U_{qq'}U_{ss'}U_{rr'}.\end{aligned} \tag{8}$$

The connection between orbital set $\{\psi_1,\ldots,\psi_M\}$ and $\{\phi_1,\ldots,\phi_N\}$ implies the connection between creation and annihilation operators,

$$\hat{d}_{q'}=\sum_{q=1}^{M}\hat{c}_q U_{qq'},\hat{d}_{p'}^\dagger=\sum_{p=1}^{M}\hat{c}_p^\dagger U_{pp'}. \tag{9}$$

Moreover, we denote the variational space for wave function as $\mathscr{D}[(\phi_1,\ldots,\phi_N)]=\mathscr{D}[(\psi_1,\ldots,\psi_M)U]$, which is the span of all Slater determinants constructed from $\{\phi_1,\ldots,\phi_N\}$.

B. Optimal orbital selection for FCI

The core concept of OptOrbFCI lies in projecting a large basis set onto a reduced subspace, followed by iterative optimization of both the projection matrix and the Full Configuration Interaction (FCI) problem within the projected subspace. With all notations defined above, the algorithm OptOrFCI solve this problem

$$\begin{aligned} &\underset{|\Phi\rangle \in \mathscr{D}[(\psi_1,\ldots,\psi_M)]\,U}{\text{minimize}} && \langle\Phi|\hat{H}|\Phi\rangle \\ &\langle\Phi|\Phi\rangle = 1 \\ &U \in \mathscr{U}(M,N) \end{aligned} \tag{10}$$

The derivation of OptOrbFCI is presented in the Supplementary Material. Here, we simply list the two subproblems that need to be optimized.

Sub-problem with fixed U. When we fix U, the orbital set $\{\phi_1,\ldots,\phi_N\}$ is also fixed. The optimization problem is then simplified as

$$\begin{aligned} &\underset{|\Phi\rangle \in \mathscr{D}[(\phi_1,\ldots,\phi_N)]}{\text{minimize}} && \langle\Phi|\tilde{H}|\Phi\rangle \\ &\langle\Phi|\Phi\rangle = 1 \end{aligned} \tag{11}$$

which is a standard FCI problem under the orbital set $\{\phi_1,\ldots,\phi_N\}$.

Sub-problem with fixed $|\Phi\rangle$. When we fix $|\Phi\rangle$, the objective function can be written as

$$\begin{aligned} \langle\Phi|\tilde{H}[U]|\Phi\rangle &= \sum_{p',q'=1}^{N} \tilde{h}_{p'q'}\,\langle\Phi|\hat{d}^\dagger_{p'}\hat{d}_{q'}|\Phi\rangle \\ &\quad + \sum_{p',q',r',s'=1}^{N} \tilde{v}_{p'q'r's'}\,\langle\Phi|\hat{d}^\dagger_{p'}\hat{d}^\dagger_{q'}\hat{d}_{s'}\hat{d}_{r'}|\Phi\rangle \\ &= \sum_{p',q'=1}^{N}\sum_{p,q=1}^{M} h_{pq}U_{pp'}U_{qq'}\,{}^1\mathrm{D}^{p'}_{q'} \\ &\quad + \sum_{p',q',r',s'=1}^{N}\sum_{p,q,r,s=1}^{M} v_{pqrs}U_{pp'}U_{qq'}U_{rr'}U_{ss'}\,{}^2\mathrm{D}^{p'q'}_{r's'} \\ &\doteq P_4(U), \end{aligned} \tag{12}$$

where ${}^1\mathrm{D}^{p'}_{q'} = \langle\Phi|\hat{d}^\dagger_{p'}\hat{d}_{q'}|\Phi\rangle$ and ${}^2\mathrm{D}^{p'q'}_{r's'} = \langle\Phi|\hat{d}^\dagger_{p'}\hat{d}^\dagger_{q'}\hat{d}_{s'}\hat{d}_{r'}|\Phi\rangle$ are the standard one-body reduced density matrix (1RDM) and two-body reduced density matrix (2RDM), respectively. The objective function, denoted as $P_4(U)$, is then a fourth order polynomial of U. Notice that h_{pq} and v_{pqrs} are given coefficients associated with the original molecular orbital set $\{\psi_1,\ldots,\psi_M\}$, and ${}^1\mathrm{D}^{p'}_{q'}$ and ${}^2\mathrm{D}^{p'q'}_{r's'}$ are also independent of U as long as we fix $|\Phi\rangle$. Hence the sub-problem can be summarized as

$$\min_{U\in\mathscr{U}(M,N)} P_4(U), \tag{13}$$

which minimizes a fourth order polynomial of U with an orthonormality constraint.

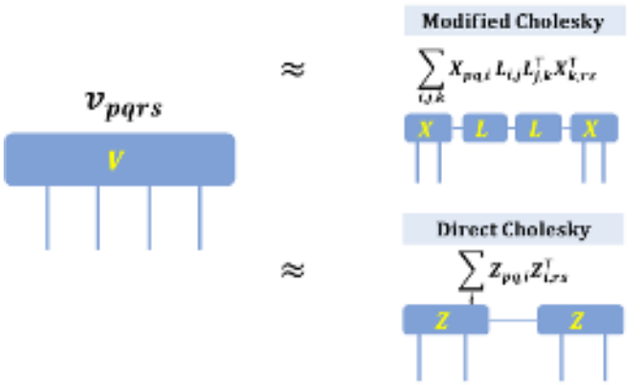

FIG. 1. Schematic representation of decompositions. Rectangular blocks denote matrix entities, with solid connecting lines representing tensor contraction operations. The left and right panels correspond to Algorithm 1, modified Cholesky decomposition and Algorithm 2, Cholesky decomposition, respectively. The difference between the two lies in whether the matrix L and X are contracted.

C. Cholesky decomposition for OptOrbFCI

In Section III, our computational profiling reveals that the numerical evaluation of eq (13) accounts for approximately 40% to 90% of the total processing time across various molecular systems. This substantial computational overhead originates from the high-dimensional tensor contraction operation expressed as $\langle pq|rs\rangle$, where the four-index two-electron integrals $\langle pq|rs\rangle$ must be contracted with multiple molecular orbital coefficient matrices. Through dimensional analysis of eq 13, we identify that the two-electron integral tensor $\langle pq|rs\rangle$ inherently possesses a low-rank structure due to the spatial locality of atomic orbitals. To exploit this mathematical structure, we implement a Cholesky decomposition-based acceleration strategy[44,54–67] that factorizes the two-electron integral tensor into rank-reduced Cholesky vectors, effectively decoupling the four-index contraction into lower-dimensional operations. Specifically, the decomposition approximates the four-center integral matrix as $V \approx LL^\top$, where L represents a set of Cholesky vectors with rank significantly smaller than the original dimensionality. This rank-reduction strategy effectively transforms the $\mathscr{O}(M^4)$ scaling contraction into a sequence of $\mathscr{O}(M^3)$ operations, while maintaining chemical accuracy through rigorous error control. Here, we propose the following two strategies. Figure 1 shows the schematic diagrams of decompositions, rectangular blocks denote matrix entities, with solid connecting lines representing tensor contraction operations. The top and bottom panels correspond to Algorithm 1 modified Cholesky decomposition and Algorithm 2 direct Cholesky decomposition, respectively. Analytical expressions for the decomposition have been explicitly provided within the figure to elucidate the mathematical workflow.

Algorithm 1 Modified Cholesky decomposition $v_{pqrs} \approx \sum_{i,k}\sum_{j=1}^{M_r} X_{pq,i}L_{i,j}L^{\top}_{j,k}X^{\top}_{k,rs}$

1: Map v_{pqrs} into a two-dimensional matrix, $V_{(pq),(rs)} = v_{pqrs}$.
2: Sample $k\times M_r$ rows of V, denoted as the set of row indices $\{row\}$.
3: Compute eigenvalue decomposition, $V(\{row\},\{row\}) = USU^{\top}$. The smaller eigenvalues are truncated using a tolerance hyperparameter, and the indices of the selected eigenvalues are denoted as the set $\{idx\}$.
4: Compute $L = S(idx,idx)^{-1/2}\times U(:,\{idx\})^{\top}, X = V(row,:)$

Algorithm 2 Direct Cholesky decomposition $v_{pqrs} \approx \sum_{i=1}^{M_r} Z_{pq,i}Z^{\top}_{i,rs}$

1: Map v_{pqrs} into a two-dimensional matrix, $V_{(pq),(rs)} = v_{pqrs}$.
2: Sample $k\times M_r$ rows of V, denoted as the set of row indices $\{row\}$.
3: Compute the row-pivoted QR decomposition. $V(row,:)P = QR$.
4: Choose the main M_r rows, resample $(k-1)\times M_r$ rows, and update the set of row indices $\{row\}$;
5: Repeat $3\sim 4$ for $l-1$ times.
6: Compute eigenvalue decomposition, $V(\{row\},\{row\}) = USU^{\top}$. The smaller eigenvalues are truncated using a tolerance hyperparameter, and the indices of the selected eigenvalues are denoted as the set $\{idx\}$.
7: Compute $Z = S(idx,idx)^{-1/2}\times U(:,\{idx\})^{\top}V(row,:)$

Here, M_r represents the effective rank obtained through numerical truncation when considering the two-electron integral V_{pqrs} as a second-order matrix. It is evident that the two strategies are essentially the same, with the subtle difference being whether the contraction of the small matrices is performed during the decomposition process or in subsequent iterative cycles. As seen from the Table I, the decomposition part of Algorithm 1 has a complexity of $\mathscr{O}(M^3)$. When $\mathscr{O}\left(\frac{M}{N}\right)\to\infty$, Algorithm 1 exhibits better theoretical performance. However, in practical terms, the current systems being computed do not reach such a scale, and Algorithm 1 modified Cholesky decomposition takes several times longer than Algorithm 2 direct Cholesky decomposition to execute. On the other hand, Algorithm 2 direct Cholesky decomposition offers a good balance between complexity and accuracy. In our software, we provide implementations of both algorithms, but the subsequent discussion is based on Algorithm 2. The second part of $P_4(U)$ admits the following form

$$\begin{aligned} W &\doteq \sum_{p',q',r',s'=1}^{N}\sum_{p,q,r,s=1}^{M} v_{pqrs}U_{pp'}U_{qq'}U_{rr'}U_{ss'}{}^2\mathrm{D}^{p'q'}_{r's'} \\ &= \sum_{p',q',r',s'=1}^{N}\sum_{p,q,r,s=1}^{M}\sum_{i=1}^{M_r} Z_{pq,i}Z^{\top}_{i,rs}U_{pp'}U_{qq'}U_{rr'}U_{ss'}{}^2\mathrm{D}^{p'q'}_{r's'}, \end{aligned} \tag{14}$$

We contract the above expression in the following order

$$\begin{aligned} W = &\sum_{p',q'=1}^{N}\sum_{i=1}^{M_r}((\sum_{p=1}^{M}(\sum_{q=1}^{M}Z_{pq,i}U_{qq'})U_{pp'}) \\ &\times(\sum_{r',s'=1}^{N}(\sum_{r=1}^{M}(\sum_{s=1}^{M}Z^{\top}_{i,rs}U_{ss'})U_{rr'}))^2\mathrm{D}^{p'q'}_{r's'}), \end{aligned} \tag{15}$$

In the computation of the alternating Barzilai-Borwein step size, which involves the $\nabla_U P_4(U)$, we have employed a similar strategy

$$\nabla_U W = \sum_{q'=1}^{N}\sum_{i=1}^{M_r}(\sum_{q=1}^{M}Z_{pq,i}U_{qq'})(\sum_{r',s'=1}^{N}(\sum_{r=1}^{M}(\sum_{s=1}^{M}Z^{\top}_{i,rs}U_{ss'})U_{rr'}))^2\mathrm{D}^{p'q'}_{r's'}, \tag{16}$$

In the Table I, we present an analysis of the computational complexity before and after the application of Cholesky decomposition. The contraction strategy follows three steps: First, contract matrix U to replace index M with N. Second, contract the 2RDM. Finally, contract remaining indices. Exploiting symmetry in this process significantly reduces computational costs. Only the highest-order terms are retained for each item in the table.

In the Table II, we list the scaling relationship between the decomposition time and the basis size. In figure 2, under logarithmic coordinates, we can clearly see that there is a quartic relationship between the time and the size of the base set.

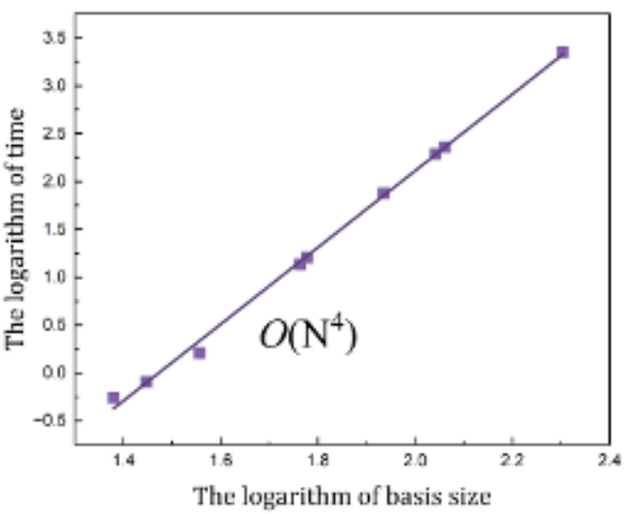

FIG. 2. Scaling relationship between the decomposition time and the basis size. It can be clearly seen that it is a quartic decomposition.

Finally, we provide a brief example to illustrate the impact of the tolerance hyper parameter on the runtime of Cholesky decomposition, the iterative optimization time, and the precision of convergence energy. In the example, the water molecule is in its equilibrium position, with an OH bond length of 1.84345 Å and a HOH bond angle of 110.6°. The basis set used is cc-pVQZ, from which 13 orbitals are selected. As can be seen from the Table III, as the tolerance

TABLE I. Computational complexity of each entry in every Algorithm. W represents the energy of the two-electron part. M is original basis size, N is the number of active orbitals, M_r represents effective rank of Cholesky decomposition, k and l are constants. $\mathcal{O}(M) = \mathcal{O}(M_r) > \mathcal{O}(N)$.

Quantity	Without decomposition	Algorithm 1 Modified Cholesky decomposition	Algorithm 2 Direct Cholesky decomposition
Decomposition	—	$\mathcal{O}(k^3M_r^3)$	$\mathcal{O}(lk^2M^2M_r^2)$
W	$\mathcal{O}(M^4N)$	$\mathcal{O}(M^2M_rN)$	$\mathcal{O}(M^2M_rN)$
$\nabla_U W$	$\mathcal{O}(M^4N)$	$\mathcal{O}(M^2M_rN)$	$\mathcal{O}(M^2M_rN)$

TABLE II. A test of decomposition time. Basis size is number of orbitals, rank is rank of decomposition.

Basis size	Rank	Average time (sec)
24	165	0.55
28	183	0.81
36	233	1.61
58	412	13.67
60	412	16.03
86	637	75.66
110	744	193.27
115	761	228.24
201	1296	2250.35

TABLE III. Impact of the tolerance hyper parameter on matrix size, rank of Cholesky decomposition, decomposition time, iterative optimization time, and ground-state energy convergence. The smaller the tolerance, the larger the rank, decomposition time, and iteration time will be, and the more accurate the energy will be. 10^{-6} is a relatively balanced value.

Tolerance	Rank	Dec time (sec)	Iter time (sec)	GS energy
10^{-2}	18	256.36	3.70	-76.16458
10^{-3}	137	293.23	8.84	-76.24720
10^{-4}	452	289.73	24.30	-76.24914
10^{-5}	633	303.52	37.22	-76.24927
10^{-6}	761	376.76	41.56	-76.24928
10^{-7}	1100	413.23	44.74	-76.24928
10^{-8}	1100	399.08	51.24	-76.24928

decreases, the eigenvalue decomposition matrix gradually increases, and the decomposition time exhibits a corresponding upward trend. Due to the randomness of the decomposition, there is some fluctuation. The iterative optimization time increases accordingly, while the converged ground-state energy has very few significant digits at larger truncations, showing no difference within at least five decimal places beyond 10^{-6}. For the subsequent calculations, considering the balance between efficiency and accuracy, a tolerance of 10^{-6} is chosen.

III. NUMERICAL RESULTS

In this section, we demonstrate that the improved algorithm OptOrbFCI is both highly efficient and maintains a high level of accuracy through several numerical experiments. First, we compared the algorithm before the improvement and demonstrated that the improved version does not result in any loss of accuracy. Then, we tested the computation time for a single H_2O molecule under different basis sets and compression conditions to demonstrate the high efficiency of the algorithm.

In all the numerical experiments, the original given orbitals (one-body and two-body integrals) are calculated via the restricted Hartree-Fock (RHF) in the PSI4[69] package. All energies are reported in the unit of Hartree (Ha). All computations were performed on a single compute node of an Intel server, Intel Xeon Gold 6226R CPU @ 2.90GHz equipped with two CPUs. Our algorithm is compatible with all Configuration Interaction algorithms. In this paper, we use an efficient FCI software called CDFCI. The communication between CDFCI and the OptOrbFCI is done via file system, i.e., the FCIDUMP file and RDM files.

A. Accuracy analysis

In this section, we compare the converged ground state energy of the improved algorithm with that of the original algorithm to verify that the decomposition does not compromise accuracy. The tests were conducted on a single H_2O molecule using the cc-pVDZ and cc-pVQZ basis sets. The geometric structure of the water molecule is the same as that mentioned earlier. The comparison is detailed in Table IV and Table V. From these tables, it can be observed that in most cases, the number of iterations is the same and the difference in converged energy is on the order of 10^{-6}. In a few instances, there are fluctuations in the number of iterations, but the energy differences remain within the chemical accuracy. Such minor fluctuations in iteration count are considered normal.

B. Efficiency analysis

In this section, we validate the improvement in computational efficiency of the new algorithm compared to the old one. The tests were conducted on a single H_2O molecule using the cc-pVTZ and cc-pVQZ basis sets. The comparison is detailed in Table VI and Figure 3 4. The table's "time per 1000 iterations" entry clearly shows a significant efficiency gap before and after the improvement, consistent with our earlier calculations of computational complexity. This is also reflected in the Figure 3, The heights of the columns in different colors respectively represent the percentages of the iteration time of each part relative to the total time of the program. From this, it can be clearly seen that the improved program has significantly outperformed the program before the improvement. The additional decomposition time introduced is negligible

TABLE IV. Accuracy comparison of improved algorithm, original algorithm, and CASSCF[70] for H_2O under the cc-pVDZ basis set. Orbs is the number of active orbitals, and Iter is the number of macro iteration. In most cases, the iteration count is identical with converged energy differences on the 10^{-6} order. Minor iteration count fluctuations occur in few instances but energy differences stay within chemical accuracy, which is deemed normal.

Orbs	Improved algorithm		Original algorithm		CASSCF	
	GS energy	Iter	GS energy	Iter	GS energy	Iter
12	-76.18435	6	-76.18435	6	-76.1734	7
13	-76.19841	6	-76.19841	6	-76.1888	7
14	-76.21770	6	-76.21769	9	-76.2029	7
15	-76.22222	7	-76.22222	7	-76.2223	7
16	-76.22693	8	-76.22693	8	-76.2247	7
17	-76.22927	7	-76.22927	7	-76.2295	7
18	-76.23208	12	-76.23213	13	-76.2314	6
19	-76.23397	4	-76.23396	4	-76.2341	5
20	-76.23583	3	-76.23583	3	-76.2360	4

TABLE V. Accuracy comparison of improved algorithm, original algorithm and CASSCF for H_2O under the cc-pVQZ basis set

Orbs	Improved algorithm		Original algorithm		CASSCF	
	GS energy	Iter	GS energy	Iter	GS energy	Iter
12	-76.23385	16	-76.22925	5	-76.2353	19
13	-76.24928	7	-76.24928	7	-76.2356	6
14	-76.27061	10	-76.27061	10	-76.2566	6
15	-76.27664	6	-76.27665	6	-76.2780	6
16	-76.28989	15	-76.28989	15	-76.2914	19
17	-76.29559	20	-76.29501	17	-76.2889	8

compared to the total time. Figure 4 presents the iteration acceleration ratios of different basis sets and orbital, it demonstrates that the speedup is substantial with large basis sets, and our main goal is to utilize large basis sets like cc-pVQZ and cc-pV5Z to achieve high precision.

C. Results of H_4

In this section, we will present the computational results for H_4. In the H_4 molecule, the four hydrogen atoms lie in the same plane and are arranged in a rectangular configuration with a bond length of 1.23 Å. The hyper parameters for H_4 is listed here. For H_4, in CDFCI, the compression threshold is 5×10^{-5}, the tolerance for convergence is 5×10^{-6}, and the maximum number of iterations is 3×10^7. In the projection method, the convergence tolerance is 10^{-7} and the maximum number of iterations is 10^4. In OptOrbFCI, the convergence tolerance is 5×10^{-5} and the maximum number of iterations is 20.

Table VII presents the numerical results for H_4. We can observe that for the cc-pVNZ series of basis sets, the energy continues to decrease as the basis set size increases. Although the rate of decrease slows down, there is no clear indication of complete convergence. This is consistent with the trends in previous literature[71].

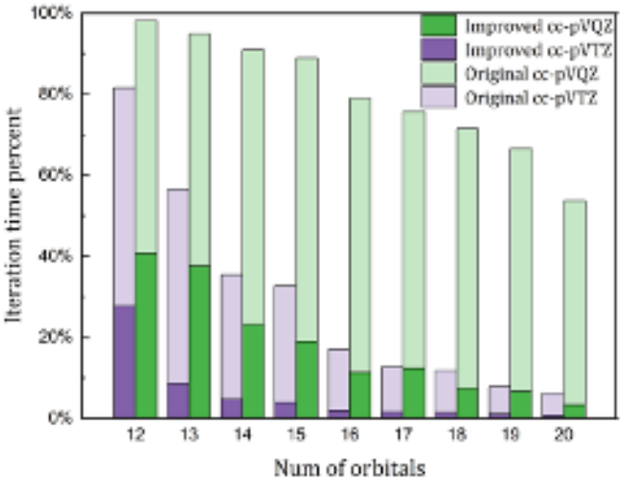

FIG. 3. The iteration time percentage to the total time of different basis sets and number of active orbitals. The colored columns' heights represent the percentage of iteration time for each part relative to the total program time. Clearly, the improved program outperforms the original, with negligible added decomposition time when compared to total time.

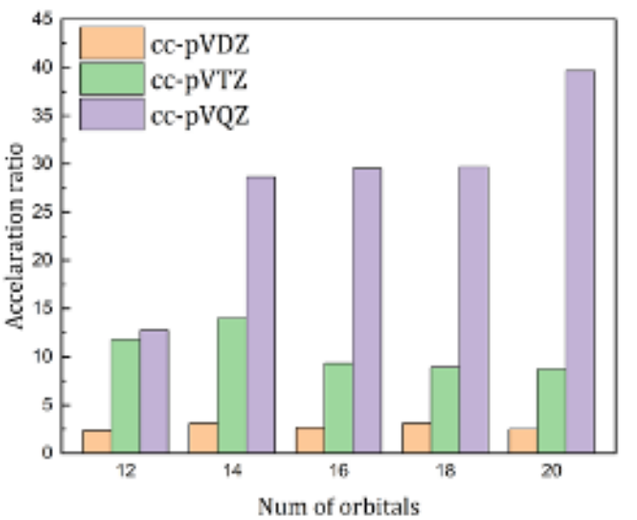

FIG. 4. Acceleration ratios of different basis sets and number of active orbitals. Figure shows iteration acceleration ratios for different basis sets and orbitals, demonstrates substantial speedup with large basis sets.

IV. CONCLUSION AND OUTLOOK

The proposed OptOrbFCI-CD method, which combines optimal orbital selection and Cholesky decomposition, significantly enhances the computational efficiency of full configuration interaction (FCI) while maintaining high accuracy. Through systematic analysis of the algorithm's computational complexity before and after applying Cholesky decomposition, the study demonstrates its effectiveness in reducing time costs for large basis sets. Numerical experiments on sys-

TABLE VI. Time comparison of improved algorithm and original algorithm for H_2O under the cc-pVTZ and cc-pVQZ basis set, the unit for each entry is seconds. Orbs is the number of active orbitals, Total is the total time of OptOrbFCI program, per 1000 iterations is the average time for a small loop to run a thousand times in each optimization iteration, Decomposition is the decomposition time. Clearly, the improved program significantly outperforms the original. The added decomposition time is negligible when compared to the total time. Rarely, longer total times in the improved algorithm stem from iteration count fluctuations.

Basis set	Orbs	Improved algorithm			Original algorithm	
		Total	Per 1000 iterations	Decomposition	Total	Per 1000 iterations
cc-pVTZ	12	451	10.33	19.62	1807	136.36
	13	1970	11.38	19.28	3634	128.30
	14	3711	11.38	19.28	6774	129.63
	15	3660	13.25	19.33	5983	145.16
	16	17790	14.65	19.67	19114	132.44
	17	9001	14.84	19.89	11664	149.84
	18	9430	17.43	14.68	9864	144.83
	19	13072	17.27	13.45	13307	147.94
	20	28480	17.17	14.26	29194	150.86
cc-pVQZ	12	4348	41.34	350.50	23031	1521.23
	13	2729	44.04	350.94	32381	1529.85
	14	6030	45.90	351.13	43922	1474.84
	15	4539	49.00	351.13	43922	1474.84
	16	20441	48.30	352.91	25390	1448.03
	17	33262	48.47	354.36	111091	1407.97
	18	39632	50.82	353.70	122438	1460.87
	19	85454	50.12	349.61	138122	1488.37
	20	118561	56.27	352.48	288699	1549.20

TABLE VII. Basis Sets and Numerical Results for H_4. As the basis set increases, the energy continuously decreases and shows a convergent trend, but it has not fully converged.

Molecule	Basis	Electrons	Orbitals	HF energy	Active orbitals	Iteration number	OptOrbFCI GS energy
H_4	cc-pVDZ	4	20	-1.9523027	20	-	-2.09718
	cc-pVTZ	4	56	-1.9563948	40	2	-2.10819
	cc-pVQZ	4	120	-1.9574441	40	3	-2.11063
	cc-pV5Z	4	220	-1.9578371	40	4	-2.11141
	cc-pV6Z	4	364	-1.9579159	40	5	-2.11154

tems such as H_4 validate the method's accuracy and efficiency across varying basis set sizes. The study observed that while energy continues to decrease with increasing basis set size, the convergence rate progressively slows without reaching full saturation. The improved algorithm achieves an optimal balance between computational cost and accuracy, showing strong potential for scaling to larger systems.

There are a list of future works of OptOrbFCI-CD. First, while the Cholesky decomposition algorithm is highly optimized for Gaussian-type orbitals (GTOs), its application to plane-wave basis sets necessitates to redesign algorithm. Second, the current framework neglects spin and spatial symmetry constraints. Integrating symmetry-adapted orbital optimization could reduce the variational space dimension, enhancing both convergence speed and robustness. Third, explore extensions to low-lying excited-state calculations. By modifying the objective function to simultaneously address the ground state and low-lying excited states, optimizing the rotation matrix could balance errors across states, enabling more accurate computation of excitation energies compared to the current algorithm.

SUPPLEMENTARY MATERIAL

The derivation and algorithm of OptOrbFCI are omitted in the main text and provided in full in the Supplementary Material.

ACKNOWLEDGMENTS

This work is partly supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB1170000, XDB0450101), the Innovation Program for Quantum Science and Technology (2021ZD0303306), the National Natural Science Foundation of China (42550106, 22173093, 22373096, 21688102), University of Science and Technology of China-Southwest University of Science and Technology Counterpart Cooperation and Development Joint Fund (KY2490002501) and Shanghai City's 2024 Targeted Commissioned Project in the Field of Quantum Technology (25LZ2601200), the Anhui Provincial Key Research and Development Program (2022a05020052), the Na-

tional Key Research and Development Program of China (2021YFB0300600), the CAS Project for Young Scientists in Basic Research (YSBR-005), the Fundamental Research Funds for the Central Universities (WK2320000061), and the Anhui Province Science and Technology Innovation Project (202423k09020010). The authors thank the Hefei Advanced Computing Center, the Supercomputing Center of Chinese Academy of Sciences, the ORISE Supercomputing Center, the Supercomputing Center of USTC, the National Supercomputing Centers in Wuxi, Tianjin, Shanghai, and Guangzhou for the computational resources. The authors acknowledge the Beijing Beilong Super Cloud Computing Co., Ltd for providing HPC resources (http://www.blsc.cn/).

[1]P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[2]J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[3]I. Purvis, George D. and R. J. Bartlett, J. Chem. Phys. **76**, 1910 (1982).
[4]R. J. Bartlett and M. Musiał, Rev. Mod. Phys. **79**, 291 (2007).
[5]U. Schollwöck, Rev. Mod. Phys. **77**, 259 (2005).
[6]G. K.-L. Chan and S. Sharma, Annu. Rev. Phys. Chem. **62**, 465 (2011).
[7]R. Olivares-Amaya, W. Hu, N. Nakatani, S. Sharma, J. Yang, and G. K.-L. Chan, J. Chem. Phys. **142**, 034102 (2015).
[8]G. H. Booth, A. J. W. Thom, and A. Alavi, J. Chem. Phys. **131**, 054106 (2009).
[9]G. H. Booth, A. Grüneis, G. Kresse, and A. Alavi, Nature **493**, 365 (2013).
[10]N. S. Blunt, S. D. Smart, G. H. Booth, and A. Alavi, J. Chem. Phys. **143**, 134117 (2015).
[11]D. Cleland, G. H. Booth, and A. Alavi, J. Chem. Phys. **132**, 041103 (2010).
[12]N. S. Blunt, J. Chem. Phys. **148**, 221101 (2018).
[13]F. R. Petruzielo, A. A. Holmes, H. J. Changlani, M. P. Nightingale, and C. J. Umrigar, Phys. Rev. Lett. **109**, 230201 (2012).
[14]N. S. Blunt, S. D. Smart, J. A. F. Kersten, J. S. Spencer, G. H. Booth, and A. Alavi, J. Chem. Phys. **142**, 184107 (2015).
[15]B. Huron, J. P. Malrieu, and P. Rancurel, J. Chem. Phys. **58**, 5745 (1973).
[16]E. Giner, A. Scemama, and M. Caffarel, Can. J. Chem. **91**, 879 (2013).
[17]J. B. Schriber and F. A. Evangelista, J. Chem. Phys. **144**, 161106 (2016).
[18]J. B. Schriber and F. A. Evangelista, J. Chem. Theory Comput. **13**, 5354 (2017).
[19]N. M. Tubman, J. Lee, T. Y. Takeshita, M. Head-Gordon, and K. B. Whaley, J. Chem. Phys. **145**, 044112 (2016).
[20]N. M. Tubman, C. D. Freeman, D. S. Levine, D. Hait, M. Head-Gordon, and K. B. Whaley, J. Chem. Theory Comput. **16**, 2139 (2020).
[21]A. A. Holmes, N. M. Tubman, and C. J. Umrigar, J. Chem. Theory Comput. **12**, 3674 (2016).
[22]S. Sharma, A. A. Holmes, G. Jeanmairet, A. Alavi, and C. J. Umrigar, J. Chem. Theory Comput. **13**, 1595 (2017).
[23]Z. Wang, Y. Li, and J. Lu, J. Chem. Theory Comput. **15**, 3558 (2019).
[24]Z. Wang, Z. Zhang, J. Lu, and Y. Li, J. Chem. Theory Comput. **19**, 7731 (2023).
[25]S. M. Greene, R. J. Webber, J. Weare, and T. C. Berkelbach, J. Chem. Theory Comput. **15**, 4834 (2019).
[26]L.-H. Lim and J. Weare, SIAM Rev. **59**, 547 (2017).
[27]Y. Li, J. Lu, and Z. Wang, SIAM J. Sci. Comput. **41**, A2681 (2019).
[28]T. M. Hernandez, R. Van Beeumen, M. A. Caprio, and C. Yang, Numer. Linear Algebr. Appl. **28**, e2341 (2021).
[29]W. Gao, Y. Li, and B. Lu, J. Sci. Comput. **93**, 63 (2022).
[30]K. D. Vogiatzis, D. Ma, J. Olsen, L. Gagliardi, and W. A. de Jong, J. Chem. Phys. **147**, 184111 (2017).
[31]H. Gao, S. Imamura, A. Kasagi, and E. Yoshida, J. Chem. Theory Comput. **20**, 1185 (2024).
[32]H. Hu, S. Upadhyay, L. Lu, A. J. Jenkins, T. Zhang, A. Shayit, S. Knecht, and X. Li, Chem. Phys. Rev. **5**, 041404 (2024).
[33]A. Shayit, C. Liao, S. Upadhyay, H. Hu, T. Zhang, A. E. DePrince III, C. Yang, and X. Li, Nat. Commun. **16**, 11016 (2025).
[34]P. Siegbahn, A. Heiberg, B. Roos, and B. Levy, Phys. Scr. **21**, 323 (1980).
[35]B. O. Roos, P. R. Taylor, and P. E. Sigbahn, Chem. Phys. **48**, 157 (1980).
[36]P. E. Siegbahn, J. Almlöf, A. Heiberg, and B. O. Roos, J. Chem. Phys. **74**, 2384 (1981).
[37]P. J. Knowles and H.-J. Werner, Chem. Phys. Lett. **115**, 259 (1985).
[38]D. Zgid and M. Nooijen, J. Chem. Phys. **128**, 144116 (2008).
[39]D. Ghosh, J. Hachmann, T. Yanai, and G. K. Chan, J. Chem. Phys. **128**, 144117 (2008).
[40]T. Yanai, Y. Kurashige, D. Ghosh, and G. K.-L. Chan, Int. J. Quantum Chem. **109**, 2178 (2009).
[41]J. Olsen, Int. J. Quantum Chem. **111**, 3267 (2011).
[42]S. Wouters, T. Bogaerts, P. Van Der Voort, V. Van Speybroeck, and D. Van Neck, J. Chem. Phys. **140**, 241103 (2014).
[43]G. Li Manni, S. D. Smart, and A. Alavi, J. Chem. Theory Comput. **12**, 1245 (2016).
[44]L. Freitag, S. Knecht, C. Angeli, and M. Reiher, J. Chem. Theory Comput. **13**, 451 (2017).
[45]Y. Ma, S. Knecht, S. Keller, and M. Reiher, J. Chem. Theory Comput. **13**, 2533 (2017).
[46]J. E. Smith, B. Mussard, A. A. Holmes, and S. Sharma, J. Chem. Theory Comput. **13**, 5468 (2017).
[47]Q. Sun, J. Yang, and G. K.-L. Chan, Int. J. Quantum Chem. **683**, 291 (2017).
[48]L. Freitag, Y. Ma, A. Baiardi, S. Knecht, and M. Reiher, J. Chem. Theory Comput. **15**, 6724 (2019).
[49]D. A. Kreplin, P. J. Knowles, and H.-J. Werner, J. Chem. Phys. **150** (2019).
[50]D. S. Levine, D. Hait, N. M. Tubman, S. Lehtola, K. B. Whaley, and M. Head-Gordon, J. Chem. Theory Comput. **16**, 2340 (2020).
[51]B. G. Levine, A. S. Durden, M. P. Esch, F. Liang, and Y. Shu, J. Chem. Phys. **154** (2021).
[52]O. J. Fajen and K. R. Brorsen, J. Chem. Theory Comput. **17**, 965 (2021).
[53]W. Dobrautz, O. Weser, N. A. Bogdanov, A. Alavi, and G. Li Manni, J. Chem. Theory Comput. **17**, 5684 (2021).
[54]H. Harbrecht, M. Peters, and R. Schneider, Appl. Numer. Math. **62**, 428 (2012), third Chilean Workshop on Numerical Analysis of Partial Differential Equations (WONAPDE 2010).
[55]S. D. Folkestad, E. F. Kjønstad, and H. Koch, J. Chem. Phys. **150**, 194112 (2019).
[56]T. B. Pedersen, S. Lehtola, I. Fdez. Galván, and R. Lindh, Wiley Interdiscip. Rev. Comput. Mol. Sci. **14**, e1692 (2024).
[57]H. Koch, A. Sánchez de Merás, and T. B. Pedersen, J. Chem. Phys. **118**, 9481 (2003).
[58]A. Liu, T. Zhang, S. Hammes-Schiffer, and X. Li, J. Chem. Theory Comput. **19**, 6255 (2023).
[59]T. Nottoli, J. Gauss, and F. Lipparini, J. Chem. Theory Comput. **17**, 6819 (2021).
[60]T. Zhang, X. Liu, E. F. Valeev, and X. Li, J. Phys. Chem. A **125**, 4258 (2021).
[61]U. Bozkaya, J. Chem. Theory Comput. **12**, 1179 (2016).
[62]J. F. H. Lew-Yee, R. Flores-Moreno, J. L. Morales, and J. M. del Campo, J. Chem. Theory Comput. **16**, 1597 (2020).
[63]A. Luenser, H. F. Schurkus, and C. Ochsenfeld, J. Chem. Theory Comput. **13**, 1647 (2017).
[64]D. B. Krisiloff, C. M. Krauter, F. J. Ricci, and E. A. Carter, J. Chem. Theory Comput. **11**, 5242 (2015).
[65]K. Brandhorst and M. Head-Gordon, J. Chem. Theory Comput. **7**, 351 (2011).
[66]J. Boström, M. Pitoňák, F. Aquilante, P. Neogrády, T. B. Pedersen, and R. Lindh, J. Chem. Theory Comput. **8**, 1921 (2012).
[67]A. E. I. DePrince and C. D. Sherrill, J. Chem. Theory Comput. **9**, 2687 (2013).
[68]Y. Li and J. Lu, J. Chem. Theory Comput. **16**, 6207 (2020).
[69]D. G. A. Smith, L. A. Burns, A. C. Simmonett, R. M. Parrish, M. C. Schieber, R. Galvelis, P. Kraus, H. Kruse, R. Di Remigio, A. Alenaizan, A. M. James, S. Lehtola, J. P. Misiewicz, M. Scheurer, R. A. Shaw, J. B. Schriber, Y. Xie, Z. L. Glick, D. A. Sirianni, J. S. O'Brien, J. M. Waldrop, A. Kumar, E. G. Hohenstein, B. P. Pritchard, B. R. Brooks, I. Schaefer, Henry F., A. Y. Sokolov, K. Patkowski, I. DePrince, A. Eugene, U. Bozkaya, R. A. King, F. A. Evangelista, J. M. Turney, T. D. Crawford, and C. D. Sherrill, J. Chem. Phys. **152**, 184108 (2020).
[70]Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, S. Wouters, and G. K.-L. Chan, Wiley Interdiscip. Rev. Comput. Mol. Sci. **8**, e1340 (2018).

[71] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, and J. Olsen, Chem. Phys. Lett. **302**, 437 (1999).