

Multicomponent Cholesky Decomposition: Application to Nuclear–Electronic Orbital Theory

Aodong Liu,[§] Tianyuan Zhang,[§] Sharon Hammes-Schiffer,^{*} and Xiaosong Li^{*}Cite This: *J. Chem. Theory Comput.* 2023, 19, 6255–6262

Read Online

ACCESS |



Metrics & More

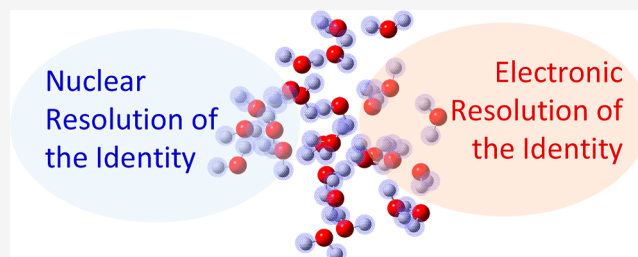


Article Recommendations



Supporting Information

ABSTRACT: The Cholesky decomposition technique is commonly used to reduce the memory requirement for storing two-particle repulsion integrals in quantum chemistry calculations that use atomic orbital bases. However, when quantum methods use multicomponent bases, such as nuclear–electronic orbitals, additional challenges are introduced due to asymmetric two-particle integrals. This work proposes several multicomponent Cholesky decomposition methods for calculations using nuclear–electronic orbital density functional theory. To analyze the errors in different Cholesky decomposition components, benchmark calculations using water clusters are carried out. The largest benchmark calculation is a water cluster $(\text{H}_2\text{O})_{27}$ where all 54 protons are treated quantum mechanically. This study provides energetic and complexity analyses to demonstrate the accuracy and performance of the proposed multicomponent Cholesky decomposition method.



1. INTRODUCTION

Ab initio electronic structure methods require storing all four-index, two-particle integrals in the in-core algorithm. However, the $M(N^4)$ scaling, where N is the number of basis functions, results in a major bottleneck for in-core *ab initio* electronic structure calculations due to this storage requirement. To address this issue, resolution-of-identity (RI) methods such as the Cholesky decomposition (CD)^{1–9} and density fitting (DF)^{10–12} techniques are used to approximate a four-index tensor with a product of two three-index tensors, thus reducing the need to store four-index integrals.^{13–23} Although both CD and DF techniques result in similar density-integral contractions, DF requires a preoptimized auxiliary basis, whereas CD can build the auxiliary basis on-the-fly to any order of accuracy for any basis set and any chemical system. In other words, the CD approach is more versatile and can also generate and benchmark auxiliary bases for any basis sets. However, while the DF approach is easy to code for electronic structure methods when provided with a set of preoptimized auxiliary bases, the implementation of CD is complex and may require optimization to achieve optimal performance.

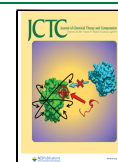
The primary challenge in developing an efficient CD approach is that determining auxiliary bases or Cholesky pivots requires traversing all two-particle integrals. This requirement creates a dilemma because the CD method aims to avoid storing $M(N^4)$ two-particle integrals. However, recent advances in the two-step algorithm have significantly improved the efficiency of the CD method.^{8,9} In the first step, the auxiliary bases or CD pivots are determined on the fly, as the two-particle integrals are generated. In the second step,

three-index integrals are computed, stored, and used in *ab initio* calculations. The dynamic two-step CD algorithm tracks and reuses integrals, leading to optimal performance with a minimal floating-point operation count.⁹ This approach represents a significant improvement over previous methods that require storing all two-particle integrals or computing auxiliary bases in a separate step, thus, reducing the memory and computational requirements for CD-based calculations.

Although CD and DF have proven successful in electronic structure calculations, their applicability and utilization in multicomponent methods, such as nuclear–electronic orbital (NEO) theory, have remained largely unexplored. In NEO methods,^{24,25} the Coulombic interactions include those between electrons and quantum nuclei, as well as those between electrons and those between quantum nuclei, resulting in three different types of two-particle integrals. Since integrals between electrons and between quantum nuclei are symmetric in nature, conventional CD and DF can be directly applied. However, integrals between electrons and quantum nuclei are asymmetric due to the different characteristics of the electronic and nuclear degrees of freedom. Although multicomponent DF approaches have been recently implemented within the framework of NEO density functional

Received: June 22, 2023

Published: September 12, 2023



theory,²⁶ NEO coupled-cluster theory,²⁷ asymmetric magnetic integrals,^{28,29} and relativistic integrals,³⁰ strategies to accurately approximate the asymmetric integrals with CD remain unexplored. In this work, we introduce a multicomponent CD method with three different strategies for constructing asymmetric two-particle NEO integrals. We conduct systematic studies along with time and storage complexity analyses to benchmark the different multicomponent CD strategies. This approach represents a significant advance toward extending the applicability of CD to multicomponent methods, enabling more accurate and efficient calculations within the NEO framework.

2. METHOD

We use the following notations throughout this work:

- Lowercase letters $p, q, r, s, \dots$ are electronic basis functions;
- Uppercase letters $P, Q, R, S, \dots$ are quantum nuclear basis functions;
- Lowercase Greek letters $\alpha, \beta, \dots$ are electronic Cholesky bases;
- Uppercase Greek letters $\Gamma, \Theta, \dots$ are quantum nuclear Cholesky bases;
- κ, λ are the combined two-component Cholesky bases.

2.1. NEO Integrals. Within the NEO framework, in addition to electrons, specified nuclei (typically protons) are also treated quantum mechanically. Assuming there are N_e and N_n electronic and quantum nuclear basis functions, respectively, all NEO methods require three types of two-particle integrals, written in chemist's notation as follows:

$$(pq|rs) \equiv M_{pq,rs} \quad (1)$$

$$(PQ|RS) \equiv M_{PQ,RS} \quad (2)$$

$$(pq|RS) \equiv M_{pq,RS} \quad (3)$$

where lower p, q, r, s and upper-case letters P, Q, R, S represent electronic and nuclear basis functions, respectively. Equation 1 (dimension $N_e^2 \times N_e^2$) and eq 2 (dimension $N_n^2 \times N_n^2$) are symmetric rank-four tensors and are related to the repulsion between electrons and between quantum nuclei, respectively. Equation 3 gives rise to the Coulombic attraction between electrons and quantum nuclei and has an asymmetric dimension of $N_e^2 \times N_n^2$.

2.2. Multicomponent CD. Given a symmetric four-index matrix $\mathbf{M}$, such as those representing repulsion integrals between electrons (eq 1) and between quantum nuclei (eq 2), the CD, like other RI approaches, seeks to approximate the full rank-four tensor as a product of rank-three tensors

$$(pq|rs) = M_{pq,rs} \approx \sum_{\alpha \in \mathcal{B}} L_{pq,\alpha} L_{rs,\alpha}^* \quad (4)$$

where $\mathcal{L} = \{L_\alpha, \dots\}$ are the Cholesky vectors and $\mathcal{B} = \{\alpha, \dots\}$ are the Cholesky bases (auxiliary bases). In the two-step CD algorithm, the first step determines the Cholesky bases $\mathcal{B}$ and the second step computes the Cholesky vectors $\mathcal{L}$. To compute the Cholesky vectors, the Coulombic interaction matrix of the Cholesky basis, $J_{\alpha\beta} = (\alpha|\beta)$, is Cholesky-decomposed according to $\mathbf{J} = \mathbf{K}\mathbf{K}^T$, where $\mathbf{K}$ is a lower triangular matrix. Note that for a complex-valued matrix, the transpose becomes the conjugate transpose.

Using the RI relationship, the four-index two-particle repulsion integrals can be written as follows:

$$(pq|rs) \approx \sum_{\alpha, \beta \in \mathcal{B}} (pq|\alpha)(\mathbf{J}^{-1})_{\alpha\beta}(\beta|rs) \quad (5)$$

Since the inverse of the Cholesky-decomposed $\mathbf{J}$ can be written as $\mathbf{J}^{-1} = \mathbf{K}^{-T}\mathbf{K}^{-1}$, where $\mathbf{K}^{-T} \equiv (\mathbf{K}^T)^{-1}$, the Cholesky vectors in eq 4 can be formed as follows:

$$L_{pq,\alpha} = \sum_{\beta \in \mathcal{B}} (pq|\beta)(\mathbf{K}^{-T})_{\beta,\alpha} \quad (6)$$

2.2.1. Symmetric NEO Integrals. For NEO systems, the two-particle integral matrices $M_{pq,rs}$ (eq 1, $N_e^2 \times N_e^2$) and $M_{PQ,RS}$ (eq 2, $N_n^2 \times N_n^2$) are symmetric rank-four tensors and are related to the repulsion between electrons and between quantum nuclei, respectively. Application of the CD method to symmetric rank-four tensors is straightforward, leading to electronic ($\mathcal{B}_e$ and $\mathcal{L}_e$) and nuclear ($\mathcal{B}_n$ and $\mathcal{L}_n$) Cholesky bases and vectors, which can be used to build the electronic and nuclear parts of the Hamiltonian. See ref 9 for details on the two-step CD algorithm for symmetric rank-four tensors.

2.2.2. Asymmetric Multicomponent NEO Integrals. The two-particle integrals between electrons and quantum nuclei (eq 3) form an asymmetric rank-four tensor $M_{pq,RS}$ with dimension $N_e^2 \times N_n^2$, which makes the RI relationship (eq 5) not directly applicable. In this work, we explore three RI strategies to generate rank-three CD vectors for asymmetric NEO integrals.

2.2.2.1. Single One-Component RI. In the single one-component RI approach, either electronic or nuclear Cholesky bases, but not both, are used in the RI equation for asymmetric integrals. For example, when electronic Cholesky bases are used, the asymmetric integral can be approximated as follows:

$$(pq|RS) = M_{pq,RS} \approx \sum_{\alpha, \beta \in \mathcal{B}_e} (pq|\alpha)(\mathbf{J}^{-1})_{\alpha\beta}(\beta|RS) \quad (7)$$

$$= \sum_{\alpha \in \mathcal{B}_e} L_{pq,\alpha} L_{RS,\alpha}^* \quad (8)$$

where the nuclear Cholesky 3-index tensor is defined as follows:

$$L_{RS,\alpha} = \sum_{\beta \in \mathcal{B}_e} (RS|\beta)(\mathbf{K}^{-T})_{\beta,\alpha} \quad (9)$$

This strategy has been previously employed in NEO coupled-cluster calculations, utilizing the DF technique.²⁷

2.2.2.2. Double One-Component RI. The second algorithm that we propose to approximate the asymmetric $(pq|RS)$ integrals uses both the electronic and the nuclear Cholesky bases. Since the dimensions of the two Cholesky basis sets can be different, a second RI matrix is needed. As such, this approach is referred to as the double one-component RI method. In this approach, both $(pq|rs)$ and $(PQ|RS)$ integrals are Cholesky-decomposed, and the resulting Cholesky bases $\mathcal{B}_e$ and $\mathcal{B}_n$ are used *via* a double one-component RI procedure to compute the asymmetric $(pq|RS)$ NEO integrals

$$\begin{aligned}
 (pq|RS) &\approx \sum_{\alpha, \beta \in \mathcal{B}_e} \sum_{\Gamma, \Theta \in \mathcal{B}_n} (pq|\beta)(\mathbf{J}^{-1})_{\beta\alpha}(\alpha|\Theta)(\mathbf{J}^{-1})_{\Theta\Gamma}(\Gamma|RS) \\
 J_{\beta\alpha} &= (\beta|\alpha) \\
 J_{\Theta\Gamma} &= (\Theta|\Gamma)
 \end{aligned}
 \quad (10)$$

which can be rewritten as the following working expressions:

$$(pq|RS) \approx \sum_{\alpha \in \mathcal{B}_e} \sum_{\Gamma \in \mathcal{B}_n} L_{pq,\alpha} A_{\alpha\Gamma} L_{RS,\Gamma}^* \quad (11)$$

$$L_{pq,\alpha} = \sum_{\beta \in \mathcal{B}_e} (pq|\beta)(\mathbf{K}^{-T})_{\beta,\alpha} \quad (12)$$

$$L_{RS,\Gamma} = \sum_{\Theta \in \mathcal{B}_n} (RS|\Theta)(\mathbf{K}^{-T})_{\Theta,\Gamma} \quad (13)$$

$$A_{\alpha\Gamma} = \sum_{\beta \in \mathcal{B}_e} \sum_{\Theta \in \mathcal{B}_n} (\mathbf{K}^{-1})_{\alpha\beta}(\beta|\Theta)(\mathbf{K}^{-T})_{\Theta\Gamma} \quad (14)$$

It should be noted that both $\mathcal{L}_e$ and $\mathcal{L}_n$ are precomputed from Cholesky-decomposed $(pq|rs)$ and $(PQ|RS)$ integrals, and thus only $A_{\alpha\Gamma}$ needs to be evaluated in the double one-component RI approach for approximating $(pq|RS)$.

2.2.2.3. Two-Component RI. In the previous two asymmetric RI approaches, only a one-component (either electronic or nuclear) Cholesky basis is used at a time. The third algorithm that we propose is a two-component RI, where the union of the electronic and nuclear Cholesky bases is used in a single RI procedure. The union of the electronic and nuclear Cholesky bases is defined as $\mathcal{B}_c = \mathcal{B}_e \cup \mathcal{B}_n$, giving rise to the two-component Cholesky basis. The asymmetric NEO integrals can be expressed in this two-component Cholesky basis as follows:

$$(pq|RS) \approx \sum_{\kappa, \lambda \in \mathcal{B}_c} (pq|\lambda)(\mathbf{J}^{-1})_{\lambda\kappa}(\kappa|RS) \quad (15)$$

$$= \sum_{\kappa \in \mathcal{B}_c} L_{pq,\kappa} L_{RS,\kappa}^* \quad (16)$$

$$L_{pq,\kappa} = \sum_{\lambda \in \mathcal{B}_c} (pq|\lambda)(\mathbf{K}^{-T})_{\lambda,\kappa} \quad (17)$$

$$L_{RS,\kappa} = \sum_{\lambda \in \mathcal{B}_c} (RS|\lambda)(\mathbf{K}^{-T})_{\lambda,\kappa} \quad (18)$$

Here, the two-component Coulombic interaction matrix $\mathbf{J}$ is defined as follows:

$$\begin{pmatrix}
 (\alpha|\beta) & (\alpha|\Gamma) \\
 (\Gamma|\alpha) & (\Gamma|\Theta)
 \end{pmatrix}
 \quad (19)$$

and the only additional quantity that needs to be computed is the $(\alpha|\Gamma)$ matrix.

2.3. Storage and Time Complexity. One of the main benefits of RI-based methods is their ability to reduce integral storage from a 4-index tensor to a 3-index tensor, which can be kept in memory and reused throughout the calculation without the need for recomputation, unlike the AO-direct algorithm. Table 1 enumerates the theoretical storage complexity of the RI-based NEO methods, focusing on multicomponent asymmetric $(pq|RS)$ integrals, as symmetric electronic and

Table 1. Additional Storage Requirement for Storing Cholesky-Decomposed Asymmetric NEO $(pq|RS)$ Integrals

RI algorithm	nuclear–electronic integral storage ^a
conventional without RI	$\mathcal{M}(N_e^2 N_n^2)$
one-component electronic RI	$\mathcal{M}(N_e^2 \mathcal{B}_e)$
one-component nuclear RI	$\mathcal{M}(N_e^2 \mathcal{B}_n)$
double one-component RI	$\mathcal{M}(\mathcal{B}_e \mathcal{B}_n)$
two-component RI	$\mathcal{M}((N_e^2 + N_n^2)(\mathcal{B}_e + \mathcal{B}_n))$

^a $|\mathcal{B}|$ is the number of Cholesky bases.

nuclear integrals can be computed using well-established CD approaches. Note that the storage requirements for pure nuclear and electronic CD vectors are $\mathcal{M}(N_n^2 |\mathcal{B}_n|)$ and $\mathcal{M}(N_e^2 |\mathcal{B}_e|)$, respectively, for all RI methods. Table 1 suggests that storing asymmetric nuclear–electronic integrals in a NEO calculation requires significantly less memory compared to the $\mathcal{M}(N_e^2 N_n^2)$ storage required for conventional 4-index integrals. This reduction in storage is observed when $|\mathcal{B}_e| \ll N_e^2$ and $|\mathcal{B}_n| \ll N_n^2$ for one-component RI methods, and when $|\mathcal{B}_e| + |\mathcal{B}_n| \ll N_e^2$ and $|\mathcal{B}_e| + |\mathcal{B}_n| \ll N_n^2$ for the two-component RI approach.

The time complexity analysis for the multicomponent CD method for mean-field calculations, such as Hartree–Fock and density functional theory (DFT) in the NEO framework, is presented in Table 2. This table compares the time complexity

Table 2. Time Complexity for Building the Nuclear–Electronic Part of the Fock Matrix

RI algorithm	nuclear–electronic Fock-build FLOPS ^a
conventional without RI	$\mathcal{O}(N_e^2 N_n^2)$
one-component electronic RI	$\mathcal{O}(N_e^2 \mathcal{B}_e + N_n^2 \mathcal{B}_e)$
one-component nuclear RI	$\mathcal{O}(N_e^2 \mathcal{B}_n + N_n^2 \mathcal{B}_n)$
double one-component RI	$\mathcal{O}(N_e^2 \mathcal{B}_e + \mathcal{B}_e \mathcal{B}_n + N_n^2 \mathcal{B}_n)$
two-component RI	$\mathcal{O}(N_e^2 (\mathcal{B}_e + \mathcal{B}_n) + N_n^2 (\mathcal{B}_e + \mathcal{B}_n))$

^a $|\mathcal{B}|$ is the number of Cholesky bases.

of the CD-based methods to the conventional Fock-build approach that does not use the RI approximation. The time complexities for constructing the pure nuclear and electronic Coulomb matrices using the RI-based methods are $\mathcal{O}(N_n^2 |\mathcal{B}_n|)$ and $\mathcal{O}(N_e^2 |\mathcal{B}_e|)$, respectively. Similarly, the time complexities for building the pure nuclear and electronic exchange matrices are $\mathcal{O}(N_n^3 |\mathcal{B}_n|)$ and $\mathcal{O}(N_e^3 |\mathcal{B}_e|)$, respectively, for the RI-based methods.

In Table 2, we analyze only the computational cost associated with building the nuclear–electronic part of the Fock matrix. However, the computational cost (FLOP count) for building the full Hamiltonian matrix using RI-based methods may not be more advantageous than the conventional 4-index integral approach. This is mainly due to the increased computational cost for the exchange part of the Hamiltonian build,³¹ which cannot be avoided for the pure electronic and nuclear parts of the Hamiltonian in a Hartree–Fock or hybrid DFT NEO calculation. For the nuclear–electronic block of the

Hamiltonian, there is no exchange contribution, and Table 2 shows that the computational cost (Coulomb only) is significantly reduced when $|\mathcal{B}_e| \ll N_e^2$ and $|\mathcal{B}_n| \ll N_n^2$ for one-component RI methods. The two-component RI method exhibits computational advantages when the sizes of sets $\mathcal{B}_e$ and $\mathcal{B}_n$ satisfy the conditions: $|\mathcal{B}_e| + |\mathcal{B}_n| \ll N_e^2$ and $|\mathcal{B}_e| + |\mathcal{B}_n| \ll N_n^2$. Thus, RI-based methods can be more advantageous for the nuclear–electronic block of the Fock matrix than for the full Hamiltonian matrix.

Based on the analyses presented in Tables 1 and 2, it is evident that RI-based integral storage and Hamiltonian build can significantly accelerate large NEO calculations when the number of Cholesky bases is much smaller than the number of orbital pairs. One of the unique advantages of the CD-based RI approach is that the accuracy of the integral approximation can be tuned up to the exact condition by setting a CD threshold of τ during the evaluation of the CD basis. Therefore, the time and storage complexity of the CD-based method also depends on the desired level of accuracy. Overall, these findings suggest that the CD-based RI approach is a promising method for accelerating large-scale NEO calculations while maintaining high accuracy.

3. RESULTS AND DISCUSSION

All three proposed multicomponent CD algorithms are implemented in a development version of the Chronus Quantum software.³² We performed benchmark calculations using the NEO density functional theory (NEO-DFT) approach on the eigen isomer of the protonated water tetramer (H_9O_4^+ , Figure 1) for the complexity and accuracy analyses.

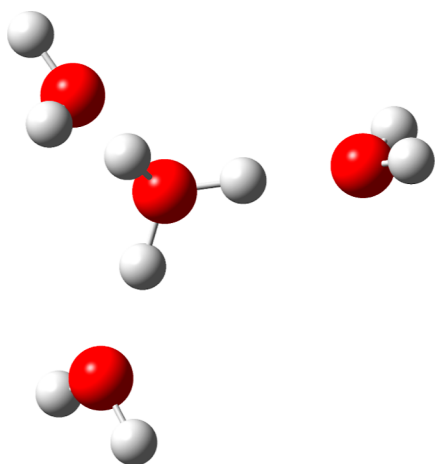


Figure 1. Eigen isomer of a protonated water tetramer.

All nine protons in this system were treated quantum mechanically with the PBS-G protonic basis set,³³ and the correlation consistent cc-pVQZ basis set³⁴ was used for the electrons. We also calculated the formation energy of a $(\text{H}_2\text{O})_{27}$ cluster using NEO-DFT to showcase the computational capabilities of the multicomponent CD approach. The PBE electronic correlation functional³⁵ and the epc17-2 electron–proton correlation functional³⁶ were used for all NEO-DFT calculations. The self-consistent-field procedure is considered converged when the root-mean-square electronic/protonic density difference falls below 10^{-8} a.u. and the energy difference between two consecutive steps falls below 10^{-10} a.u.

3.1. Storage Analysis. Table 3 presents the integral storage requirements for the NEO calculations of the

Table 3. Integral Storage for the Eigen Isomer of the Protonated Water Tetramer H_9O_4^+

Conventional without RI								
4-index tensor storage (GB)								
(pq rs)	(PQ RS)		(pq RS)		total			
492.0	480.3		486.1		1470.2			
One-Component Electronic/Protonic RI								
τ	$ \mathcal{B}_e $	$ \mathcal{B}_n $	CD tensor storage (GB)					
			$L_{pq,\alpha}$	$L_{PQ,\Gamma}$	$L_{RS,\alpha}$	$L_{pq,\Gamma}$	total _e	total _n
10^{-4}	2226	2592	4.4	5.1	4.4	5.1	13.9	14.6
10^{-5}	2783	2880	5.5	5.6	5.5	5.7	16.6	16.9
10^{-6}	3665	3159	7.3	6.2	7.2	6.3	20.6	19.7
Double One-Component RI								
τ_e	τ_n	$ \mathcal{B}_e $	$ \mathcal{B}_n $	CD tensor storage (GB)				
				$L_{pq,\alpha}$	$L_{PQ,\Gamma}$	$A_{\alpha,\Gamma}$	total	
10^{-4}	10^{-4}	2266	2592	4.4	5.1	0.05	9.5	
	10^{-5}		2880		5.6	0.05	10.1	
	10^{-6}		3159		6.2	0.06	10.7	
10^{-5}	10^{-4}	2783	2592	5.5	5.1	0.06	10.7	
	10^{-5}		2880		5.6	0.06	11.2	
	10^{-6}		3159		6.2	0.07	11.8	
10^{-6}	10^{-4}	3665	2592	7.3	5.1	0.08	12.4	
	10^{-5}		2880		5.6	0.08	13.0	
	10^{-6}		3159		6.2	0.09	13.6	
Two-Component RI								
τ_e	τ_n	$ \mathcal{B}_e $	$ \mathcal{B}_n $	CD tensor storage (GB)				
				$L_{pq,\alpha}$	$L_{PQ,\Gamma}$	$L_{pq,\kappa}$	$L_{RS,\kappa}$	total
10^{-4}	10^{-4}	2266	2592	4.4	5.1	9.6	9.4	28.5
	10^{-5}		2880		5.6	10.1	10.0	30.2
	10^{-6}		3159		6.2	10.7	10.6	31.8
10^{-5}	10^{-4}	2783	2592	5.5	5.1	10.7	10.5	31.8
	10^{-5}		2880		5.6	11.2	11.1	33.5
	10^{-6}		3159		6.2	11.8	11.6	35.2
10^{-6}	10^{-4}	3665	2592	7.3	5.1	12.4	12.3	37.0
	10^{-5}		2880		5.6	13.0	12.8	38.7
	10^{-6}		3159		6.2	13.5	13.4	40.4

protonated water tetramer (H_9O_4^+ , Figure 1), where all nine protons are treated quantum mechanically. The system consists of 498 electronic and 495 protonic basis functions, necessitating 1470 GB of memory to store all four-index two-particle integrals. In contrast, by setting the electronic and nuclear CD threshold to $\tau = 10^{-6}$, the storage requirements for all NEO CD vectors are significantly reduced to 19.7–20.6, 13.6, and 40.4 GB for one-component RI, double one-component RI, and two-component RI approaches, respectively. By increasing the CD threshold, the number of selected Cholesky bases decreases, leading to a decrease in the memory required to store the CD 3-index tensor. The CD storage requirement is reduced by almost 2 orders of magnitude for the one-component RI approach at $\tau = 10^{-4}$ compared to storing the conventional 4-index tensors.

3.2. Numerical Accuracy Analysis. The previous section demonstrates that using the CD approach can significantly reduce the memory required for storing NEO integrals.

Table 4. Signed Energy Error Using Different RI Methods with Different Thresholds for the Eigen Isomer of the Protonated Water Tetramer $\text{H}_9\text{O}_4^{+\alpha}$

One-Component Electronic/Protonic RI						
τ	$ \mathcal{B}_e $	$ \mathcal{B}_n $	δE (a.u.)			
			$(pq rs)$	$(PQ RS)$	$(pq RS)_e$	$(pq RS)_p$
10^{-4}	2226	2592	-4.3×10^{-5}	-4.9×10^{-9}	-1.5×10^{-3}	-6.7×10^{-4}
10^{-5}	2783	2880	-4.0×10^{-6}	-1.1×10^{-9}	1.0×10^{-4}	2.9×10^{-4}
10^{-6}	3665	3159	-4.6×10^{-7}	-9.5×10^{-10}	-3.7×10^{-6}	4.3×10^{-4}
Double One-Component RI						
τ_e	τ_n	$ \mathcal{B}_e $	$ \mathcal{B}_n $	δE (a.u.)		
				$(pq rs)$	$(PQ RS)$	$(pq RS)$
10^{-4}	10^{-4}	2266	2592	-4.3×10^{-5}	-4.9×10^{-9}	-2.2×10^{-3}
	10^{-5}		2880	-1.1×10^{-9}	-1.2×10^{-3}	
	10^{-6}		3159	-9.5×10^{-10}	-1.1×10^{-3}	
10^{-5}	10^{-4}	2783	2592	-4.0×10^{-6}	-4.9×10^{-9}	-5.7×10^{-4}
	10^{-5}		2880	-1.1×10^{-9}	3.9×10^{-4}	
	10^{-6}		3159	-9.5×10^{-10}	5.3×10^{-4}	
10^{-6}	10^{-4}	3665	2592	-4.6×10^{-7}	-4.9×10^{-9}	-6.8×10^{-4}
	10^{-5}		2880	-1.1×10^{-9}	2.9×10^{-4}	
	10^{-6}		3159	-9.5×10^{-10}	4.2×10^{-4}	
Two-Component RI						
τ_e	τ_n	$ \mathcal{B}_e $	$ \mathcal{B}_n $	δE (a.u.)		
				$(pq rs)$	$(PQ RS)$	$(pq RS)$
10^{-4}	10^{-4}	2266	2592	-4.3×10^{-5}	-4.9×10^{-9}	1.0×10^{-7}
	10^{-5}		2880	-1.1×10^{-9}	-7.4×10^{-9}	
	10^{-6}		3159	-9.5×10^{-10}	-4.7×10^{-8}	
10^{-5}	10^{-4}	2783	2592	-4.0×10^{-6}	-4.9×10^{-9}	7.9×10^{-8}
	10^{-5}		2880	-1.1×10^{-9}	1.6×10^{-8}	
	10^{-6}		3159	-9.5×10^{-10}	1.4×10^{-8}	
10^{-6}	10^{-4}	3665	2592	-4.6×10^{-7}	-4.9×10^{-9}	1.6×10^{-8}
	10^{-5}		2880	-1.1×10^{-9}	1.2×10^{-8}	
	10^{-6}		3159	-9.5×10^{-10}	9.8×10^{-9}	

^a δE is the difference between the total NEO-DFT energy with the specified Cholesky-decomposed integrals and the total energy computed with full 4-index integrals using the AO-direct algorithm.

However, a crucial question that arises is how the use of CD integrals impacts the numerical accuracy of NEO calculations. To address this question, we conducted a comprehensive analysis of the accuracy of NEO calculations for the protonated water tetramer (H_9O_4^+ , see Figure 1), as a function of the CD threshold.

The accuracy of the CD-based method for NEO calculations is assessed in Table 4 by comparing the total NEO-DFT energy obtained using the CD method with that obtained using the AO-direct algorithm with full 4-index integrals. Various calculations were performed, utilizing only electronic (*pqlrs*), protonic (*PQlRS*), or electronic/protonic (*pqlRS*) CD integrals while keeping the other integral types in the full 4-index form. For instance, when assessing the accuracy of the electronic/protonic (*pqlRS*) CD integral approximation, the pure electronic (*pqlrs*) and protonic (*PQlRS*) integrals were maintained in their exact full 4-index form. This investigation aimed to analyze the error magnitudes associated with the CD approximation for different types of integrals. The energy error presented in Table 4 was evaluated using the total NEO-DFT energy calculated using the same fully converged electronic and protonic density matrices computed with 4-index integrals. This procedure ensures that the same set of converged electronic and protonic densities are used for each calculation

to enable a fair and consistent comparison of the CD-based methods. We also present the errors using fully converged SCF results for all calculations in the Supporting Information.

Table 4 shows that the difference in total NEO-DFT energy decreases as the CD threshold τ is lowered for calculations using only electronic (*pqlrs*) or protonic (*PQlRS*) CD integrals while keeping the electronic/protonic integrals in their exact 4-index form. Notably, the effect of the electronic (*pqlrs*) CD threshold τ_e on the total energy is more pronounced than that of the protonic (*PQlRS*) CD integrals. This is because electronic densities are more delocalized than protonic densities and are, therefore, more sensitive to the quality of the underlying RI approximation. For protonic CD, the error in energy is already below 10^{-8} a.u. with $\tau_n = 10^{-4}$. Both the electronic (*pqlrs*) and protonic (*PQlRS*) CD approximations exhibit an asymptotic behavior, approaching the exact energy from below as the CD threshold decreases. This behavior can be attributed to the positive-definite nature of the CD error in the electronic and protonic repulsion terms. As the CD threshold is tightened, the error in the CD approximation decreases, resulting in a more accurate representation of the electronic and protonic repulsion energies.

We also investigated the accuracy of the CD approximation for the asymmetric multicomponent electronic/protonic (*pql*

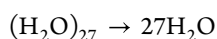
RS) integrals. In Table 4, we report the results of NEO-DFT calculations using only the electronic/protonic ($pqlRS$) CD integrals while keeping the pure electronic and protonic integrals in their exact 4-index forms. The accuracy of the ($pqlRS$) CD integral approximation varies depending on the RI method used.

For the one-component RI approach, the electronic CD basis $[(prlRS)_e]$ performs slightly better than the protonic CD basis $[(prlRS)_p]$. The accuracy is not improved by the double one-component approach. The complete CD basis for the multicomponent electronic/protonic integrals is the union of electronic and protonic basis pairs. As a result, using the one-component (either electronic or protonic) or double-component RI approach is insufficient for accurately describing the projection between electronic and protonic basis pairs. Hence, all one-component RI approaches typically exhibit large errors and slow convergence as the CD threshold is tightened. To overcome this issue, it is necessary to use a multicomponent RI approach that fully accounts for the intercomponent interactions and accurately describes the projection between electronic and protonic basis pairs.

In Table 4, we also evaluate the performance of the two-component RI approach for NEO-DFT calculations using only electronic/protonic ($pqlRS$) CD integrals while keeping the pure electronic and protonic integrals in their exact 4-index forms. As both electronic and protonic CD bases are used for the asymmetric electronic/protonic ($pqlRS$) integral, the mapping between the electronic and protonic CD bases is more complete, which leads to better accuracy. The two-component RI approach is able to converge toward the exact 4-index result, as seen in the table. At $\tau = 10^{-6}$, the error in the NEO-DFT energy using only the ($pqlRS$) two-component CD basis is only 10^{-8} a.u. Thus, the two-component RI approach is a viable option for reducing the memory requirements of NEO calculations without compromising accuracy.

However, as the CD threshold is tightened, the NEO-DFT results obtained using the asymmetric multicomponent electronic/protonic ($pqlRS$) integrals exhibit nonmonotonic convergence, with no definitive upper or lower bound on the energy error compared to the exact result. This behavior is observed across all of the RI strategies investigated in this study. The reason for this phenomenon is that the energy contribution arising from the asymmetric multicomponent integrals is first order, as opposed to being quadratic in relation to both the electronic and protonic density matrices. As a consequence, the error associated with the electronic/protonic CD integrals may not consistently decrease in a monotonic manner as the CD threshold is tightened.

3.3. Large Case Study—27 H_2O Cluster. One advantage of using RI-based methods for NEO calculations is that they allow for the storage of integrals in memory, making it feasible to perform large-scale computational studies. In this work, we showcase the effectiveness of the multicomponent CD approach by computing the formation energy of a water cluster consisting of 27 H_2O molecules, as shown in Figure 2. The formation energy is obtained by computing the difference in energy between the cluster and the constituent H_2O molecules, using the following reaction:



$$\Delta E = 27 \times E(H_2O) - E((H_2O)_{27})$$

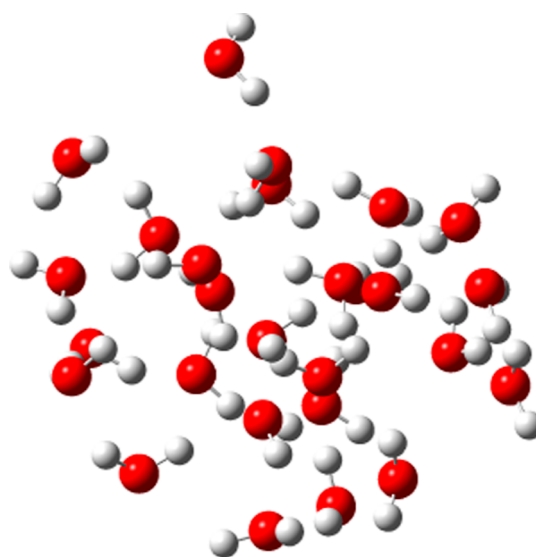


Figure 2. Twenty-seven water cluster model system.

where E is the converged NEO-DFT energy. To generate the initial structure of the water cluster, we used a snapshot from a classical molecular dynamics simulation of bulk water using the SPC/E model,³⁷ which has a physically reasonable hydrogen bond distribution.^{38,39}

The formation energy of the water cluster $(H_2O)_{27}$ was computed using different electronic and protonic basis sets, with a CD threshold τ set to 10^{-4} for all integrals. The two-component RI approach was utilized, and the computed results are presented in Table 5. For reference, the formation energy was also computed using exact 4-index integrals in an AO-direct algorithm with a Schwarz screening threshold of 10^{-12} . The total formation energy of the water cluster $(H_2O)_{27}$, computed using the two-component RI method, exhibits a difference of $\sim 10^{-3}$ kcal/mol compared to the results obtained using the 4-index exact integrals across all tested basis sets. However, the memory requirement for storing 4-index integrals in the largest calculation (cc-pVTZ + PB4-F1) is 154,700 GB, making this problem impractical for in-core computation, compared to only 858 GB for the two-component RI approach. This test suggests that the RI-based approach is computationally advantageous over the AO-direct approach for the system size studied here. However, the AO-direct Fock-build approach may eventually outperform the RI-based Fock-build approach in the large system limit.

4. CONCLUSIONS

The goal of this study was to investigate the application of the multicomponent CD methods for NEO-DFT calculations. We introduced and implemented three different types of CD methods, including one-component, double-component, and two-component RI algorithms. Benchmark calculations were carried out on a protonated water tetramer and a large $(H_2O)_{27}$ cluster, where all protons were treated quantum mechanically.

Our numerical analyses revealed that the energy errors introduced by electronic and protonic CD approximations exhibit an asymptotic behavior that is bound from above as the CD threshold is tightened. This behavior can be attributed to the positive definite nature of the CD error and the quadratic energy dependence on electronic or protonic density matrices.

Table 5. Formation Energy for a (H₂O)₂₇ Water Cluster Computed with the Two-Component RI Method Using Different Electronic and Protonic Basis Sets

basis set						ΔE (kcal/mol)	
electronic	N_e	$ B_e $	protonic	N_n	$ B_n $	conventional	two-component RI
def2-TZVP	1161	5489	PB4-D	1262	4860	9.10797	9.10646
cc-pVTZ	1566	7120	PB4-D	1262	4860	10.18563	10.18329
def2-TZVP	1161	5489	PB4-F1	1620	6966	9.04314	9.04151
cc-pVTZ	1566	7120	PB4-F1	1620	6966	10.10621	10.10396

In contrast, the energy error arising from multicomponent nuclear–electronic CD approximations is not bound from below or above by the exact solution due to its linear dependence on the electronic and protonic density matrices.

Our analysis revealed that while all CD methods were useful in reducing storage requirements for in-core calculations of two-particle integrals, only the two-component RI approach achieved high accuracy. By using the two-component CD integrals, our NEO-DFT calculations for the large (H₂O)₂₇ cluster exhibited a formation energy error of only $\sim 10^{-3}$ kcal/mol and a significantly decreased computational cost compared to AO-direct calculations utilizing 4-index integrals. These results showcase the effectiveness of the two-component CD approach for both precise and efficient calculations within the NEO framework.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.3c00686>.

Molecular structures and SCF energy errors using different RI methods (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Sharon Hammes-Schiffer – Department of Chemistry, Yale University, New Haven, Connecticut 06520, United States; orcid.org/0000-0002-3782-6995;

Email: sharon.hammes-schiffer@yale.edu

Xiaosong Li – Department of Chemistry, University of Washington, Seattle, Washington 98195, United States;

orcid.org/0000-0001-7341-6240; Email: xsli@uw.edu

Authors

Aodong Liu – Department of Chemistry, University of Washington, Seattle, Washington 98195, United States

Tianyuan Zhang – Department of Chemistry, University of Washington, Seattle, Washington 98195, United States

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.jctc.3c00686>

Author Contributions

[§]A.L. and T.Z. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The development of the Chronus Quantum computational software is supported by the Office of Advanced Cyberinfrastructure, National Science Foundation (grant no. OAC-2103717 to X.L., supporting T.Z., and OAC-2103902 to

S.H.-S.). The development of solvated NEO methods is supported by the Department of Energy in the Computational Chemical Science program (grant no. DE-SC0023284 to S.H.-S. and X.L., supporting A.L.). Computations were facilitated through the use of advanced computational, storage, and networking infrastructure provided by the shared facility supported by the University of Washington Molecular Engineering Materials Center (DMR-1719797) via the Hyak supercomputer system.

■ REFERENCES

- (1) 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*, 683–705.
- (2) Røeggen, I.; Wisløff-Nilssen, E. On the Beebe-Linderberg Two-Electron Integral Approximation. *Chem. Phys. Lett.* **1986**, *132*, 154–160.
- (3) Aquilante, F.; Lindh, R.; Pedersen, T. B. Unbiased Auxiliary Basis Sets for Accurate Two-Electron Integral Approximations. *J. Chem. Phys.* **2007**, *127*, 114107.
- (4) Aquilante, F.; Gagliardi, L.; Pedersen, T. B.; Lindh, R. Atomic Cholesky Decompositions: A Route to Unbiased Auxiliary Basis Sets for Density Fitting Approximation With Tunable Accuracy and Efficiency. *J. Chem. Phys.* **2009**, *130*, 154107.
- (5) 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.
- (6) Boman, L.; Koch, H.; Sánchez de Merás, A. Method Specific Cholesky Decomposition: Coulomb and Exchange Energies. *J. Chem. Phys.* **2008**, *129*, 134107.
- (7) Røeggen, I.; Johansen, T. Cholesky Decomposition of the Two-Electron Integral Matrix in Electronic Structure Calculations. *J. Chem. Phys.* **2008**, *128*, 194107.
- (8) Folkestad, S. D.; Kjønsstad, E. F.; Koch, H. An Efficient Algorithm for Cholesky Decomposition of Electron Repulsion Integrals. *J. Chem. Phys.* **2019**, *150*, 194112.
- (9) Zhang, T.; Liu, X.; Valeev, E. F.; Li, X. Toward the Minimal Floating Operation Count Cholesky Decomposition of Electron Repulsion Integrals. *J. Phys. Chem. A* **2021**, *125*, 4258–4265.
- (10) Dunlap, B. I.; Connolly, J. W. D.; Sabin, J. R. On Some Approximations in Applications of X α Theory. *J. Chem. Phys.* **1979**, *71*, 3396–3402.
- (11) Kendall, R. A.; Früchtel, H. A. The Impact of the Resolution of the Identity Approximate Integral Method on Modern Ab Initio Algorithm Development. *Theor. Chem. Acc.* **1997**, *97*, 158–163.
- (12) Whitten, J. L. Coulombic Potential Energy Integrals and Approximations. *J. Chem. Phys.* **1973**, *58*, 4496–4501.
- (13) Freitag, L.; Knecht, S.; Angeli, C.; Reiher, M. Multireference Perturbation Theory with Cholesky Decomposition for the Density Matrix Renormalization Group. *J. Chem. Theory Comput.* **2017**, *13*, 451–459.
- (14) Lesiuk, M. Implementation of the Coupled-Cluster Method with Single, Double, and Triple Excitations using Tensor Decompositions. *J. Chem. Theory Comput.* **2020**, *16*, 453–467.
- (15) Bozkaya, U. Efficient Implementation of the Second-Order Quasidegenerate Perturbation Theory with Density-Fitting and

Cholesky Decomposition Approximations: Is It Possible To Use Hartree-Fock Orbitals for a Multiconfigurational Perturbation Theory? *J. Chem. Theory Comput.* **2019**, *15*, 4415–4429.

(16) Hannon, K. P.; Li, C.; Evangelista, F. A. An Integral-Factorized Implementation of the Driven Similarity Renormalization Group Second-Order Multireference Perturbation Theory. *J. Chem. Phys.* **2016**, *144*, 204111.

(17) Zhang, T.; Li, C.; Evangelista, F. A. Improving the Efficiency of the Multireference Driven Similarity Renormalization Group via Sequential Transformation, Density Fitting, and the Noninteracting Virtual Orbital Approximation. *J. Chem. Theory Comput.* **2019**, *15*, 4399–4414.

(18) Motta, M.; Shee, J.; Zhang, S.; Chan, G. K.-L. Efficient Ab Initio Auxiliary-Field Quantum Monte Carlo Calculations in Gaussian Bases via Low-Rank Tensor Decomposition. *J. Chem. Theory Comput.* **2019**, *15*, 3510–3521.

(19) Fosso-Tande, J.; Nguyen, T.-S.; Gidofalvi, G.; DePrince, A. E., III Large-Scale Variational Two-Electron Reduced-Density-Matrix-Driven Complete Active Space Self-Consistent Field Methods. *J. Chem. Theory Comput.* **2016**, *12*, 2260–2271.

(20) Epifanovsky, E.; Zuev, D.; Feng, X.; Khistyayev, K.; Shao, Y.; Krylov, A. I. General Implementation of the Resolution-Of-The-Identity and Cholesky Representations of Electron Repulsion Integrals Within Coupled-Cluster and Equation-Of-Motion Methods: Theory and Benchmarks. *J. Chem. Phys.* **2013**, *139*, 134105.

(21) Sherrill, C. D. Frontiers in Electronic Structure Theory. *J. Chem. Phys.* **2010**, *132*, 110902.

(22) Hohenstein, E. G.; Sherrill, C. D. Density Fitting and Cholesky Decomposition Approximations in Symmetry-Adapted Perturbation Theory: Implementation and Application to Probe the Nature of π - π Interactions in Linear Acenes. *J. Chem. Phys.* **2010**, *132*, 184111.

(23) Aquilante, F.; Pedersen, T. B.; Lindh, R.; Roos, B. O.; Sánchez de Merás, A.; Koch, H. Accurate Ab Initio Density Fitting for Multiconfigurational Self-Consistent Field Methods. *J. Chem. Phys.* **2008**, *129*, 024113.

(24) Webb, S. P.; Iordanov, T.; Hammes-Schiffer, S. Multiconfigurational Nuclear-Electronic Orbital Approach: Incorporation of Nuclear Quantum Effects in Electronic Structure Calculations. *J. Chem. Phys.* **2002**, *117*, 4106–4118.

(25) Pavošević, F.; Culpitt, T.; Hammes-Schiffer, S. Multicomponent Quantum Chemistry: Integrating Electronic and Nuclear Quantum Effects via the Nuclear-Electronic Orbital Method. *Chem. Rev.* **2020**, *120*, 4222–4253.

(26) Mejía-Rodríguez, D.; de la Lande, A. Multicomponent Density Functional Theory With Density Fitting. *J. Chem. Phys.* **2019**, *150*, 174115.

(27) Pavošević, F.; Tao, Z.; Hammes-Schiffer, S. Multicomponent Coupled Cluster Singles and Doubles with Density Fitting: Protonated Water Tetramers with Quantized Protons. *J. Phys. Chem. Lett.* **2021**, *12*, 1631–1637.

(28) Nottoli, T.; Burger, S.; Stopkiewicz, S.; Gauss, J.; Lipparini, F. Computation of NMR Shieldings at the CASSCF Level Using Gauge-Including Atomic Orbitals and Cholesky Decomposition. *J. Chem. Phys.* **2022**, *157*, 084122.

(29) Gauss, J.; Blaschke, S.; Burger, S.; Nottoli, T.; Lipparini, F.; Stopkiewicz, S. Cholesky Decomposition of Two-Electron Integrals in Quantum-Chemical Calculations with Perturbative or Finite Magnetic Fields Using Gauge-Including Atomic Orbitals. *Mol. Phys.* **2023**, *121*, No. e2101562.

(30) Kelley, M. S.; Shiozaki, T. Large-Scale Dirac-Fock-Breit Method Using Density Fitting and 2-Spinor Basis Functions. *J. Chem. Phys.* **2013**, *138*, 204113.

(31) Früchtel, H. A.; Kendall, R. A.; Harrison, R. J.; Dyall, K. G. An Implementation of RI-SCF on Parallel Computers. *Int. J. Quantum Chem.* **1997**, *64*, 63–69.

(32) Stetina, T.; Clark, A.; Li, X. X-ray Absorption Signatures of Hydrogen-bond Structure in Water-Alcohol Solutions. *Int. J. Quantum Chem.* **2019**, *119*, No. e25802.

(33) Yu, Q.; Pavošević, F.; Hammes-Schiffer, S. Development of Nuclear Basis Sets for Multicomponent Quantum Chemistry Methods. *J. Chem. Phys.* **2020**, *152*, 244123.

(34) Dunning, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

(35) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(36) Brorsen, K. R.; Yang, Y.; Hammes-Schiffer, S. Multicomponent Density Functional Theory: Impact of Nuclear Quantum Effects on Proton Affinities and Geometries. *J. Phys. Chem. Lett.* **2017**, *8*, 3488–3493.

(37) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91*, 6269–6271.

(38) Ozkanlar, A.; Clark, A. E. Chemnetworks: A Complex Network Analysis Tool for Chemical Systems. *J. Comput. Chem.* **2014**, *35*, 495–505.

(39) Lu, L.; Wildman, A.; Jenkins, A. J.; Young, L.; Clark, A. E.; Li, X. The “Hole” Story in Ionized Water from the Perspective of Ehrenfest Dynamics. *J. Phys. Chem. Lett.* **2020**, *11*, 9946–9951.