Article

Robust Approximation of Tensor Networks: Application to Grid-Free Tensor Factorization of the Coulomb Interaction

Karl Pierce, Varun Rishi, and Edward F. Valeev*

Cite This: https://doi.org/10.1021/acs.jctc.0c01310

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Approximation of a tensor network by approximating (e.g., factorizing) one or more of its constituent tensors can be improved by canceling the leading-order error due to the constituents' approximation. The utility of such robust approximation is demonstrated for robust canonical polyadic (CP) approximation of a (density-fitting) factorized two-particle Coulomb interaction tensor. The resulting algebraic (grid-free) approximation for the Coulomb tensor, closely related to the factorization appearing in pseudospectral and tensor hypercontraction approaches, is efficient and accurate, with significantly reduced rank compared to the naive (nonrobust) approximation. Application of the robust approximation to the particle—particle ladder term in the coupled-cluster singles and doubles reduces the size complexity from $O(N^6)$ to $O(N^5)$ with robustness ensuring negligible errors in chemically relevant energy differences using CP ranks approximately equal to the size of the density-fitting basis.
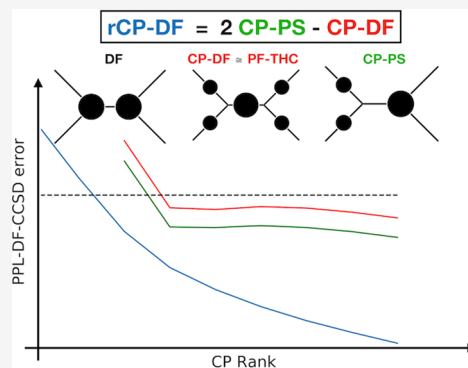
1. INTRODUCTION

Numerical approximation of the (matrix elements of the) Hamiltonian is a ubiquitous strategy for decreasing the cost and complexity of quantum simulation of, e.g., electronic structure in both real space and spectral representations. Examples in spectral representations include density fitting (DF: also referred to in quantum chemistry as the resolution-of-the-identity (RI), in global[1,2] and local[3−5]), the pseudospectral[6−15] (PS) approach, Cholesky decomposition (CD),[16−19] the fast multipole method (FMM),[20−22] tensor hypercontraction (THC),[23−31] the canonical polyadic (CP) decomposition (also known as CANDECOMP/PARAFAC[32,33],[34−40] and many others.[41−50] These approaches can be coarsely classified as (a) abstract (algebraic) approximations of the Hamiltonian tensor (e.g., CD, CP, global DF, algebraic FMM[51,52]) and (b) approximations that utilize physical context (e.g., use of grids in pseudospectral and THC, domain decomposition in FMM and local DF).

It is common to wish to approximate tensors in a tensor network. In such a case, it may be possible to construct a better network approximation to the original tensor network than obtained by approximating the individual tensors in the network. Inspired by these basic observations, we consider the robust[a] approximation of tensor networks, in which the leading-order error due to the approximation of the network constituents is canceled. Here, we demonstrate the utility of the idea by constructing a robust CP (rCP) approximation for a simple network of 2 order-3 tensors obtained by the DF-factorization of the two-particle Coulomb interaction tensor. Unlike DF-factorization alone, the rCP-DF decomposition reduces the complexity of the ladder-type diagrams in many-

body electronic structure methods. The robustness of the approximation ensures a favorable prefactor; in this work, cost savings are observed for systems with as few as three atoms, as demonstrated for the particle—particle ladder (PPL) diagram in the coupled-cluster method with single and double excitations (CCSD).

The rest of the paper is organized as follows. In Section 2 of this paper, we introduce the idea of robust approximation of tensor networks, use it to construct an efficient algebraic approximation to a two-particle interaction tensor, and discuss how to utilize the proposed factorization to evaluate the particle—particle ladder (PPL) diagram with reduced complexity. Section 3 describes the details of the computational experiments. Section 4 compares the performances of nonrobust and robust approximations applied to the CCSD PPL diagram using standard benchmark sets of noncovalent interaction energies and reaction energies. Section 5 summarizes our findings and discusses other possible applications of the idea.

2. FORMALISM

2.1. Robust Approximation of Tensor Networks.
Consider a tensor network composed of a sequence of tensors, $\{\mathcal{T}_1, ..., \mathcal{T}_k\} \equiv \{\mathcal{T}_i\}$, $i = 1, ..., k$. For our purposes the

A

(a) $g_{ab,cd}$

(b) Factorization of $g_{ab,cd}$ via Eq. (5)

(c) CP factorization of $B_{pq,X}$ via Eq. (10)

(d) CP-PS factorization of $g_{ab,cd}$ via Eq. (17)

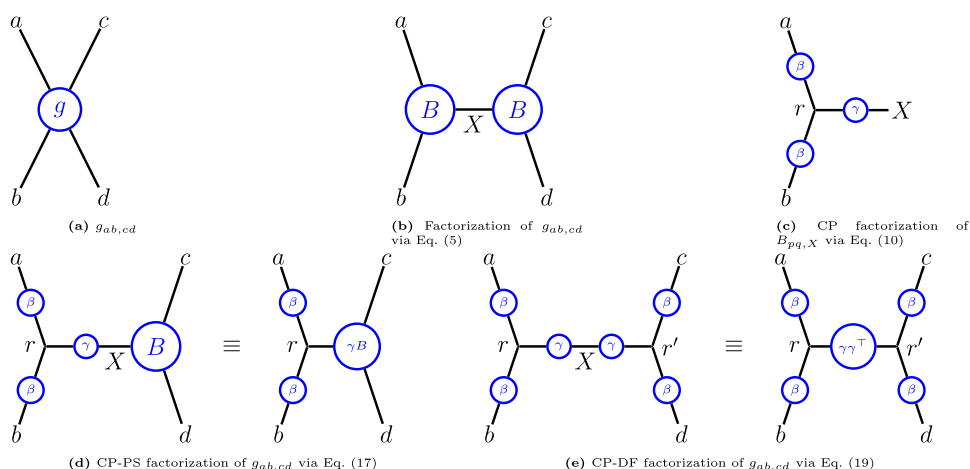(e) CP-DF factorization of $g_{ab,cd}$ via Eq. (19)

Figure 1. Graphical representation of the two-particle interaction tensor (eq 4) and factorizations thereof considered in this work.

network can have arbitrary topology, it does not even need to be connected. Our objective is to minimize the error in the network due to replacing tensors $\mathcal{T}_i$ by their approximants $\hat{\mathcal{T}}_i$. Assuming that the approximation error in each tensor

$$\delta_i \equiv \mathcal{T}_i - \hat{\mathcal{T}}_i \tag{1}$$

is "small", i.e., $\| \delta_i \| = O(\epsilon)$, the tensor network can be accurately represented in terms of tensor approximants by including terms linear in the error

$$\begin{aligned} \{\mathcal{T}_1, \, ..., \, \mathcal{T}_k\} = & \{\hat{\mathcal{T}}_1, \, ..., \, \hat{\mathcal{T}}_k\} \\ & + \sum_j \{\hat{\mathcal{T}}_1, \, ..., \, \hat{\mathcal{T}}_{j-1}, \, \delta_j, \, \hat{\mathcal{T}}_{j+1}, \, ..., \, \hat{\mathcal{T}}_k\} + O(\epsilon^2) \end{aligned} \tag{2}$$

Note that the naive approximation of the network, given by the first term on the right-hand side, is only accurate to $O(\epsilon)$. A *robust* approximation, accurate to $O(\epsilon^2)$, is obtained by plugging eq 1 into eq 2

$$\begin{aligned} \{\mathcal{T}_1, \, ..., \, \mathcal{T}_k\} = & (1 - k)\{\hat{\mathcal{T}}_1, \, ..., \, \hat{\mathcal{T}}_k\} \\ & + \sum_j \{\hat{\mathcal{T}}_1, \, ..., \, \hat{\mathcal{T}}_{j-1}, \, \mathcal{T}_j, \, \hat{\mathcal{T}}_{j+1}, \, ..., \, \hat{\mathcal{T}}_k\} + O(\epsilon^2) \end{aligned} \tag{3}$$

Clearly, the robust approximation is only applicable to tensor networks, not individual tensors.

In the context of numerical tensor approximations, the robust approximation has enjoyed a long use by the electronic structure community.[45,53,55,56] Despite its simplicity and/or apparent lack of novelty, in the context of tensor computation, the idea has potentially significant unexplored utility. Its utility came as a real surprise to us when we stumbled on its novel application, described below.

2.2. Robust Approximation of a Factorized Two-Particle Interaction Tensor. Consider tensor representation of a two-particle interaction[b] in a generic basis of size n

$$g_{ab,cd} \equiv \iint \phi_a^*(\mathbf{r}_1)\phi_b(\mathbf{r}_1)g(\mathbf{r}_1, \mathbf{r}_2)\phi_c^*(\mathbf{r}_2)\phi_d(\mathbf{r}_2) \, d\mathbf{r}_1 \, d\mathbf{r}_2 \tag{4}$$

The comma separator between indices defines the default matricization; namely, matrix O will refer to the matricized form of tensor O, with element $O_{p_1 p_2 ... q_1 q_2 ...}$ located in row $p_1 p_2 ...$ and column $q_1 q_2 ...$ of the matrix. It is also useful to convey tensor

expressions diagrammatically; in Penrose notation tensor, g is represented as a single node (Figure 1).

To efficiently approximate g, it is important to retain the analytic properties, such as symmetries and positivity. In this work, specifically, we must consider the properties of the Poisson kernel, $g(\mathbf{r}_1, \mathbf{r}_2) = |\mathbf{r}_1 - \mathbf{r}_2|^{-1}$, which is "positive" in both two-particle and one-particle senses, i.e., both $\hat{g}_2 f(\mathbf{r}_1, \mathbf{r}_2) \equiv g(\mathbf{r}_1, \mathbf{r}_2) \times f(\mathbf{r}_1, \mathbf{r}_2)$ and $\hat{g}_1 f(\mathbf{r}_1) \equiv \int g(\mathbf{r}_1, \mathbf{r}_2)f(\mathbf{r}_2) \, d\mathbf{r}_2$, respectively, are positive-definite operators.

For positive-definite kernels, the tensor g can be factorized into a symmetric form

$$g_{ab,cd} \approx \sum_X B_{ab,X}B_{cd,X} \tag{5}$$

which, in its matrix form, is recognized as the ubiquitous, symmetric particle-wise factorization

$$\mathbf{g} \approx \mathbf{B}\mathbf{B}^{\mathrm{T}} \tag{6}$$

Such "generalized square-root" factorization is not unique. One way to compute the factorization efficiently is by a (rank-revealing) Cholesky decomposition (CD);[19] for any finite precision, the CD rank (i.e., the number of columns of $\mathbf{B}$) is $O(n)$. Another way to compute this symmetric factorization is via DF, where

$$B_{ab,X} = C_{ab,Y}(\mathbf{G}^{1/2})_{Y,X} \tag{7}$$

the fitting coefficients $C_{pq,Y}$ are determined by weighted least-squares fitting,[1−5] typically, using the Coulomb "metric"

$$(\mathbf{G})_{X,Y} \equiv \iint \phi_X(1)g(1, 2)\phi_Y(2) \, d1 \, d2 \tag{8}$$

and the square root of $\mathbf{G}$ is defined by eq 6, rather than the conventional, principal square root. The size of the fitting basis $\{\phi_X\}$, denoted here by X, is in practice proportional to n.

For large systems, CD and DF approaches lead to sparse $\mathbf{B}$; however, on large basis sets, the onset of sparsity can be slow and thus difficult to exploit. Hence, it may be worthwhile to seek more general data sparsity in $\mathbf{B}$ by further factorization. For example, consider the approximate CP factorization of $\mathbf{B}$

$$B_{ab,X} \approx \sum_r^R \beta_{a,r}\kappa_{b,r}\gamma_{X,r} \tag{9}$$

For real basis functions, $g_{ab,cd}$ and, hence, $B_{ab,X}$ are symmetric with respect to the $a \leftrightarrow b$ permutation; this symmetry is ensured automatically if $\kappa_{b,r} \equiv \beta_{b,r}$ or

$$B_{ab,X} \approx \hat{B}_{ab,X} \equiv \sum_r^R \beta_{a,r} \beta_{b,r} \gamma_{X,r} \tag{10}$$

It is well known[57,58] that (aside from trivial examples) finding the exact CP rank R is hard, but there are efficient ways to construct such approximations for a fixed CP rank, R.[59−61]

Tensor factorization of Coulomb interaction eq 4 that utilizes CP topology has been long employed in electronic structure. This is due to the natural connection between CP factorization and quadrature approximation for an integral over a product of three or more factors. Most relevant for our purposes is Friesner's pioneering use of a pseudospectral (PS) method (PS methods are also known as discrete variable representation [DVR] methods) to solve the Hartree−Fock equations for electrons.[6] His work led to the pseudospectral family of methods,[6−9,11−15] which approximate Coulomb integrals using a numerical quadrature over one electron. This quadrature approximation is also employed in the COSX method[44−46,48−50,62] and in the approximation of many-electron integrals in explicitly correlated F12 methods.[63]

Computing $g_{ab,cd}$ using numerical quadrature involves replacing the integration over a single electron, for example electron 1, with a sum over a set of quadrature points

$$g_{ab,cd} \overset{\text{PS}}{\approx} \sum_g w_g \phi_a^*(\mathbf{r}_g) \phi_b(\mathbf{r}_g) \int g(\mathbf{r}_g, \mathbf{r}_2) \phi_c^*(\mathbf{r}_2) \phi_d(\mathbf{r}_2) \, d\mathbf{r}_2 \tag{11}$$

Introducing

$$X_{a,g} \equiv \sqrt{w_g} \, \phi_a(\mathbf{r}_g) \tag{12}$$

$$Y_{g,cd} \equiv \int g(\mathbf{r}_g, \mathbf{r}_2) \phi_c^*(\mathbf{r}_2) \phi_d(\mathbf{r}_2) \, d\mathbf{r}_2 \tag{13}$$

leads to the algebraic form of the PS approximation

$$g_{ab,cd} \overset{\text{PS}}{\approx} \sum_g X_{a,g}^* X_{b,g} Y_{g,cd} \tag{14}$$

which makes the connection to CP factorization obvious; note that the summation over grid points g corresponds to the three-way hyperedge in the diagrammatic representation of eq 14 in Figure 1d. In practice, accurate implementation of the PS approximation is sensitive to the choice of the grid and requires various measures to reduce the error.[24,45,48,50,64] However, the algebraic form of the PS approximation can be viewed as an abstract tensor network approximation of $g_{ab,cd}$, with factors X and Y defined not by the particular choice of real-space quadrature in eq 11, but by arbitrary fitness conditions.

Inserting a quadrature once for every particle leads to, what Martínez and co-workers termed, the tensor hypercontraction[c] (THC) approximation[23−29,31,65] of $g_{ab,cd}$

$$g_{ab,cd} \overset{\text{THC}}{\approx} \sum_{g_1,g_2} w_{g_1} w_{g_2} \phi_a^*(\mathbf{r}_{g_1}) \phi_b(\mathbf{r}_{g_1}) g(\mathbf{r}_{g_1}, \mathbf{r}_{g_2}) \phi_c^*(\mathbf{r}_{g_2}) \phi_d(\mathbf{r}_{g_2}) \tag{15}$$

and its algebraic form

$$g_{ab,cd} \overset{\text{THC}}{\approx} \sum_{g_1} \sum_{g_2} X_{a,g_1}^* X_{b,g_1} Y_{g_1,g_2} X_{c,g_2}^* X_{d,g_2} \tag{16}$$

The diagrammatic representation of eq 16, shown in Figure 1e, includes two three-way hyperedges. Clearly, the same idea can be applied to a matrix element of any (local) n-body operator.[65] THC approximation was originally exploited in the algebraic form, using algebraic CP decomposition of three-center overlap integrals in the context of (nonrobust) overlap-metric DF to define factors X and Y in eq 16 ("PF-THC").[23] It was subsequently formulated using real-space quadrature to define factors X and least-squares fitting to determine factor Y in eq 16 ("LS-THC").[24,27,31] What these approaches have in common with each other and with other related factorizations[37] is the use of the tensor network topology of eq 16; how the factors are determined can differ widely between the methods.

Although our focus in this paper is on the three-way CP factorization (CP3) we should also note that the direct four-way algebraic CP factorization of Coulomb integrals (CP4) has been employed by Benedikt and co-workers.[34−36] Related four-way factorizations of Coulomb integrals have been considered by Peng and Kowalski, who proposed to compress the Cholesky factors of the Coulomb tensor by the singular vector decomposition (SVD); the use of factorized integrals has been explored in the CC method.[66] More recently, Motta and co-workers employed a similar multistep factorization to reduce the cost of auxiliary-field quantum Monte Carlo methods.[67] The similarity of these factorizations to the four-way CP decomposition is due to the appearance of the four-way hyperedge, whereas all of the factorizations considered in this work are limited to three-way hyperedges only.

To introduce the main result of our work, consider how to best introduce the CP3 approximation (eq 10) for the symmetric (CD/DF-like) factorization in eq 5. Using CP3 once produces a PS-like factorization, to which we will refer as CP-PS

$$g_{ab,cd} \overset{\text{CP−PS}}{\approx} \sum_X \sum_r^R \beta_{a,r} \beta_{b,r} \gamma_{X,r} B_{cd,X} = \sum_r^R \beta_{a,r} \beta_{b,r} (\gamma B)_{cd,r} \tag{17}$$

where we introduced

$$(\gamma B)_{cd,r} \equiv \sum_X \gamma_{X,r} B_{cd,X} \tag{18}$$

compare eq 17 to eq 14 to recognize the connection to the algebraic PS factorization. Using CP3 twice produces a THC-like factorization, to which we will refer as CP-DF

$$g_{ab,cd} \overset{\text{CP−DF}}{\approx} \sum_X \sum_r^R \beta_{a,r} \beta_{b,r} \gamma_{X,r} \sum_{r'}^R \beta_{c,r'} \beta_{d,r'} \gamma_{X,r'}$$
$$= \sum_r^R \beta_{a,r} \beta_{b,r} \sum_{r'}^R \beta_{c,r'} \beta_{d,r'} (\gamma \gamma^T)_{r,r'} \tag{19}$$

where we introduced $(\gamma \gamma^T)_{r,r'} \equiv \sum_X \gamma_{X,r} \gamma_{X,r'}$; compare eq 19 to eq 16 to recognize the connection to the algebraic THC factorization.

Clearly, both CP-PS and CP-DF approximations are linear in the error introduced by the CP3 approximation (eq 10). As discussed in Section 2.1, it is possible to eliminate the linear error using the robust form of CP-DF, to which we will refer as rCP-DF

$$g_{ab,cd} \overset{\text{rCP-DF}}{\approx} 2g_{ab,cd}^{\text{CP-PS}} - g_{ab,cd}^{\text{CP-DF}}$$

$$= \sum_r^R \beta_{a,r}\beta_{b,r}\left(2(\gamma B)_{cd,r} - \sum_{r'}^R \beta_{c,r}\beta_{d,r'}(\gamma\gamma^{\text{T}})_{r,r'}\right) \tag{20}$$

Although the rCP-DF approximant has a higher computational cost, than either CP-PS or CP-DF, computing the PPL diagram with the rCP-DF approximation has the same complexity ($O(N^5)$) as the aforementioned approaches. However, the systematic error cancellation unique to rCP-DF should, at equal CP rank, result in significantly smaller errors than either CP-PS or CP-DF and thus should be computationally superior to these simpler alternatives.

2.3. Application to the Particle−Particle Ladder Diagram. Our primary objective is to reduce the computational cost of the particle−particle ladder (PPL) diagram in CC and other many-body methods. It is well known that both PS[11,15] and THC factorizations[27,37] can reduce the computational complexity of the PPL term in the canonical molecular orbital (MO) basis from $O(N^6)$ to $O(N^5)$; hence, the same should be possible for the PPL term in the rCP-DF approximation. Indeed, plugging in eq 17 into the spin-free PPL expression (permutational symmetry is ignored for simplicity) yields

$$\sum_{bd} g_{ab,cd}t_{bdij} \overset{\text{CP-PS}}{\approx} \text{PPL}^{\text{CP-PS}}$$

$$\equiv \sum_r^R \beta_{a,r}\left(\sum_d (\gamma B)_{cd,r}\left(\sum_b \beta_{b,r}t_{bdij}\right)\right) \tag{21}$$

The order of evaluation that minimizes the operation count is shown by parentheses, with the result of each binary tensor product stored in an intermediate tensor. The inner-most product, $\sum_b \beta_{b,r}t_{bdij} \to (I_1)_{rdij}$, is covariant (i.e., it is a pure tensor contraction) and has an operation cost of $2o^2u^2R$, where o and u are the numbers of occupied and unoccupied MOs, respectively, and R is the CP rank. The second product is of general type (i.e., it cannot be mapped to a single matrix multiplication) and has the same cost as the first product. The last product is a pure contraction and has the same cost as the other two contractions. The total operation count of the CP-PS approximated PPL is thus $6o^2u^2R$ vs the $2o^2u^4$ cost of the naive approach; note that precomputing the (γB) intermediate (eq 18) is done once, outside of the CCSD solver loop, and has the negligible cost ($2u^2XR$, where X is the size of the DF fitting basis). We can expect computational savings from the use of CP-PS when $R < u^2/3$.[d]

The PPL term can be similarly reformulated with the $O(N^5)$ cost using the CP-DF approximation. One approach, utilized by Parrish et al.[27] and Hummel et al.,[37] uses the CP-PS route (eq 21) by recomputing the appropriate intermediates

$$\sum_{bd} g_{ab,cd}t_{bdij} \overset{\text{CP-DF}}{\approx} \text{PPL}^{\text{CP-DF}}$$

$$\equiv \sum_r^R \beta_{a,r}\left(\sum_d (\gamma\hat{B})_{cd,r}\left(\sum_b \beta_{b,r}t_{bdij}\right)\right) \tag{22}$$

where $(\gamma\hat{B})_{cd,r}$ is the CP-factorized intermediate $(\gamma B)_{cd,r}$, obtained by inserting eq 10 into eq 18[e]

$$(\gamma\hat{B})_{cd,r} \equiv \sum_X \gamma_{X,r}\left(\sum_{r'}^R \beta_{a,r'}\beta_{b,r'}\gamma_{X,r'}\right) \tag{23}$$

The operation count of this route is $6o^2u^2R$; hence, the crossover relative to the naive PPL evaluation occurs at the same CP rank as in the CP-PS route.

Another CP-DF route, utilized by Hummel et al.[37] and Mardirossian et al.,[68] introduces order-4 tensors with two CP indices

$$\sum_{bd} g_{ab,cd}t_{bdij} \overset{\text{CP-DF}}{\approx} \text{PPL}^{\text{CP-DF}}$$

$$\equiv \sum_r^R \beta_{a,r} \sum_{r'}^R \left(\beta_{c,r'}\left((\gamma\gamma^{\text{T}})_{r,r'}\left(\sum_b \beta_{b,r}\left(\sum_d \beta_{d,r'}t_{bdij}\right)\right)\right)\right) \tag{24}$$

Compared to three tensor products in the CP-PS approach, the CP-DF route has five products, with all but the third product of $(\gamma\gamma^{\text{T}})$ being pure contractions. The operation count is $4o^2u^2R + 4o^2uR^2 + o^2R^2$; since in practice $R \gg u$, the cost is expected to be dominated by the $4o^2uR^2$ contribution.

To reduce the operation count, relative to the conventional PPL, the route outlined above requires $R < \sqrt{u^3/2} = u^{3/2}/\sqrt{2}$ (compared to $R < u^2/3$ requirement of the CP-PS-based route). Clearly, the cost crossover occurs earlier in the CP-PS-based route. Furthermore, the low arithmetic intensity of the element-wise (Hadamard-like) third product in eq 24 lowers the computational efficiency of this approach. For these reasons, throughout our work we used the CP-PS-based approach, eq 22, to implement CP-DF PPL.

Clearly, the PPL term can be therefore approximated via rCP-DF with the $O(N^5)$ cost by naively combining the CP-PS and CP-DF approximations

$$\sum_{bd} g_{ab,cd}t_{bdij} \overset{\text{rCP-DF}}{\approx} 2 \times \text{PPL}^{\text{CP-PS}} - \text{PPL}^{\text{CP-DF}} \tag{25}$$

Plugging eqs 21 and 22 into eq 25 and refactoring leads to the following evaluation scheme with optimal operation count

$$\sum_{bd} g_{ab,cd}t_{bdij} \overset{\text{rCP-DF}}{\approx} \text{PPL}^{\text{rCP-DF}}$$

$$\equiv \sum_r^R \beta_{a,r}\left(\sum_d (\gamma\tilde{B})_{cd,r}\left(\sum_b \beta_{b,r}t_{bdij}\right)\right) \tag{26}$$

in which we introduced

$$(\gamma\tilde{B})_{cd,r} \equiv 2(\gamma B)_{cd,r} - (\gamma\hat{B})_{cd,r} \tag{27}$$

The total operation count of the rCP-DF PPL approximation is $6o^2u^2R$, which is identical to that of the CP-PS and CP-DF PPL approximations. Thus, rCP-DF is the preferred three-way CP approach in the context of the PPL evaluation.

3. COMPUTATIONAL DETAILS

CP approximations for order-3 tensors were computed using the standard alternating least squares (ALS) method.[69,70] Although ALS can be slow to converge and the quality of the solution can strongly depend on the initial guess,[71] we found that our solver converged robustly with an initial guess of vectors generated using quasi-random numbers taken from the uniform distribu-
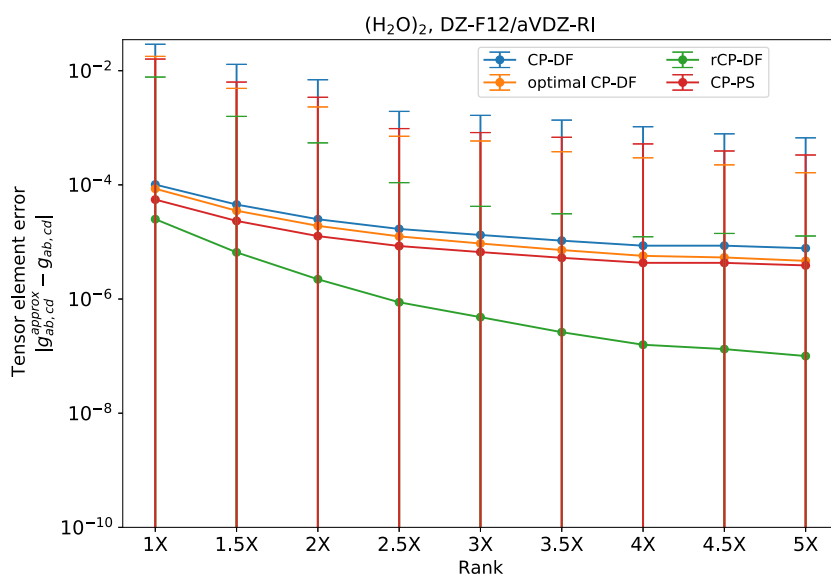
Figure 2. Absolute errors in matrix elements of $g_{ab,cd}$ for a water dimer with S66 configuration approximated by the CP-PS, CP-DF, and rCP-DF factorizations obtained with ALS precision of $\epsilon = 10^{-3}$. The error bars denote the max/min unsigned errors.

tion on $[-1,1]$. No consistent benefit was found from an initial guess scheme that generated factor matrices using the higher-order SVD (HOSVD)[72] padded with random vectors (where random vectors were generated as just described). Furthermore, no discernible benefit was found from the use of a regularized ALS (RALS) solver.[73] The use of nonlinear and gradient-based solvers[59,60] as an alternative to ALS will be investigated in future work.

Assessment of the CP-based Coulomb tensor factorizations utilized the full S66 benchmark set of weakly bound complexes[74] as well as a 12-system representative set of 12 complexes (S66/12);[f] some computations utilized a smaller seven-system subset of S66/12 (systems 1−4 and 10−12; dubbed S66/7). The S66 geometries were taken from the Benchmark Energy and Geometry Database (BEGDB).[75] Additional assessments utilized the HJO12 set of isogyric reaction energies,[76,77] the eight low-lying conformers of $(H_2O)_6$,[78] and a conformer of $(H_2O)_{20}$.[79,80] All of the above computations utilized the cc-pVDZ-F12 (abbreviated as DZ-F12) orbital basis set (OBS).[81] The two-electron interaction tensors were approximated using standard Coulomb-metric density fitting using the aug-cc-pVDZ-RI (abbreviated as aVDZ-RI) density-fitting basis set (DFBS).[82] Assessment of the basis set variation in the performance of rCP-DF used the following additional OBS/DFBS pairs: the aug-cc-pVDZ[83,84] (aVDZ) OBS paired with the aVDZ-RI DFBS, the aug-cc-pVTZ (aVTZ) OBS[83,84] paired with the aug-cc-pVTZ-RI[82] (aVTZ-RI) DFBS, and the cc-pVTZ-F12[81] (TZ-F12) OBS paired with the aVTZ-RI DFBS. The CP approximations of Coulomb integral tensors were utilized in only the PPL diagram of CCSD. Only valence electrons were correlated in all CCSD computations.

All computations were run on the Virginia Tech Advanced Research Computing's Cascades cluster which utilizes standard nodes that contain two Intel Xeon E5-2683 v4 CPUs, and high-memory nodes, each with four Intel Xeon E7-8867 v4 CPUs. Only the $(H_2O)_{20}$ computations utilized Cascades' high-memory nodes. In the following section, speedup is determined as

$$speedup = \frac{t_{DF-CCSD}}{t_{CP-PPL-DF-CCSD} + t_{CP-ALS}} \tag{28}$$

where $t_{DF\text{-}CCSD}$ and $t_{CP\text{-}PPL\text{-}DF\text{-}CCSD}$ are the total time it takes to compute the CCSD correlation energy with either the DF or CP approximation applied to the PPL diagram and $t_{CP\text{-}ALS}$ is the time it takes to compute the CP decomposition using the ALS method.

The CP-ALS decomposition was implemented in C++ in the open-source Basic Tensor Algebra Subroutine (BTAS) library.[85] The CP-DF, CP-PS, and rCP-DF approximations are implemented in a developmental version of the Massively Parallel Quantum Chemistry (MPQC) package.[86]

4. RESULTS

The discussion of computational experiments is organized as follows. In Section 4.1, we examine how the errors in the matrix elements of the Coulomb operator converge with respect to the CP rank. It turns out that the use of CP in the CP-PS and CP-DF approximations results in two types of errors: due to suboptimal factors in the tensor network and due to the deficient CP rank; the use of the robust approximation greatly reduces both types of errors. In Sections 4.2 and 4.3, we discuss the error in the CCSD energies introduced by and the cost reduction of the CP approximation of the PPL diagram, respectively. Note, to standardize CP rank across systems, we report the CP rank in the units of X (the size of the density-fitting basis), which grows proportionally to n.

4.1. Errors in Coulomb Matrix Elements: Effects of CP Factor Optimality, CP Rank, and Robustness. The most direct way to assess a particular factorization of the Coulomb interaction tensor is to examine the matrix elements themselves. Since the data varies little between systems, Figure 2 shows the absolute errors of the matrix elements of $g_{ab,cd}$ for a particular system, namely, the water dimer at the S66 geometry. The first observation is that both the average (solid circles) and the maximum (horizontal line) errors decrease in the CP-DF > CP-PS > rCP-DF series, with the CP-DF and CP-PS errors decaying with the CP rank at a similar rate, and much slower than the rCP-DF errors. This observation is easy to explain. Using the matrix
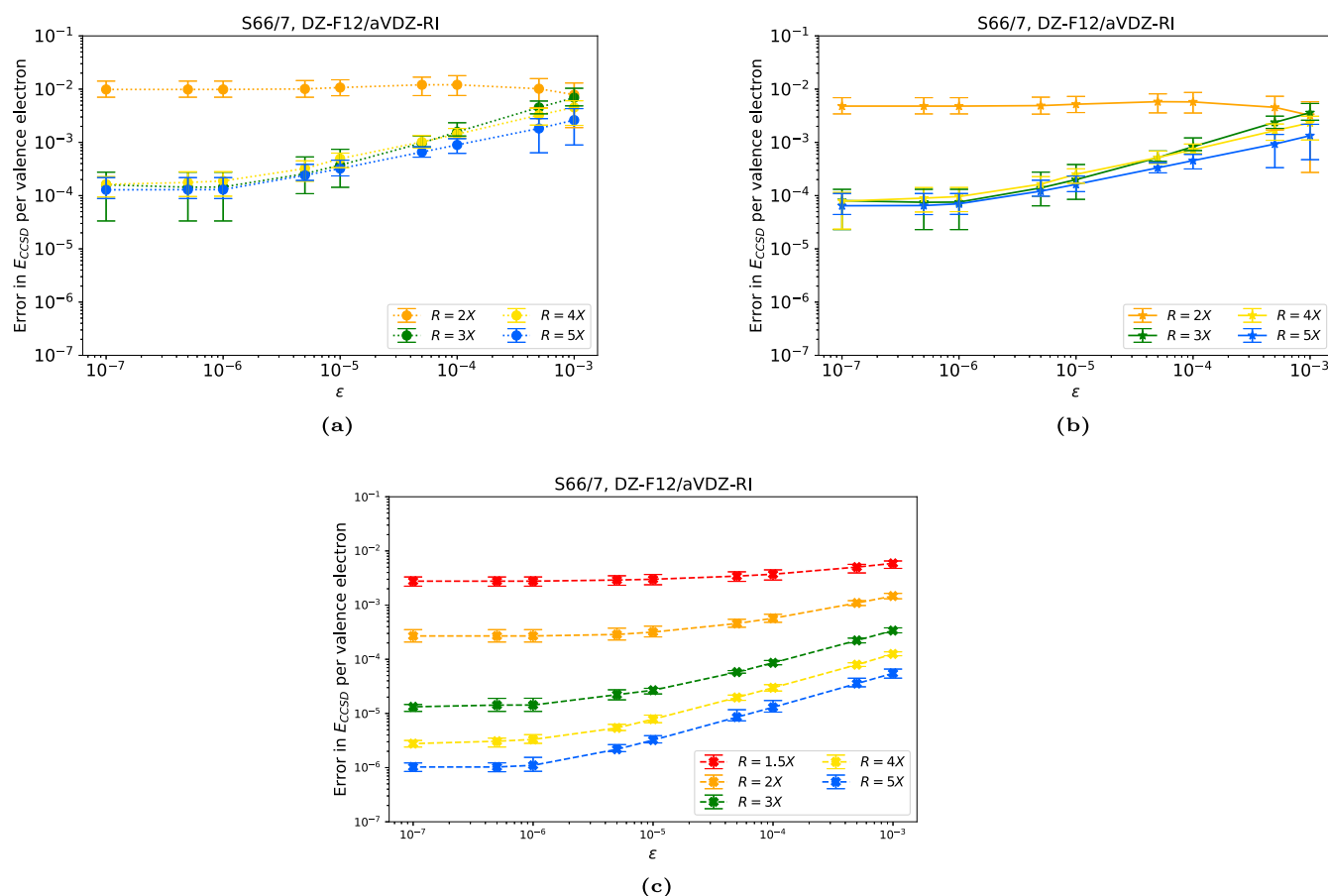
(a)

(b)

(c)

Figure 3. Mean unsigned errors in the per-electron CCSD correlation energies (kcal/mol) of molecules in the S66/12 data set, relative to canonical CCSD, induced by the (a) CP-DF, (b) CP-PS, or (c) rCP-DF approximations to PPL vs the ALS precision (ϵ). The error bars denote the max/min unsigned errors.

notation introduced in eq 6, it is clear that the leading-order error of the CP-DF-factorization should be roughly twice the error of CP-PS

$$\mathbf{g}^{DF} - \mathbf{g}^{CP-PS} = \mathbf{BB}^T - \frac{1}{2}(\hat{\mathbf{B}}\mathbf{B}^T + \mathbf{B}\hat{\mathbf{B}}^T)$$

$$= \frac{1}{2}(\boldsymbol{\delta}\mathbf{B}^T + \mathbf{B}\boldsymbol{\delta}^T) \tag{29}$$

$$\mathbf{g}^{DF} - \mathbf{g}^{CP-DF} = \mathbf{BB}^T - \hat{\mathbf{B}}\hat{\mathbf{B}}^T = \boldsymbol{\delta}\hat{\mathbf{B}}^T + \hat{\mathbf{B}}\boldsymbol{\delta}^T + \boldsymbol{\delta}\boldsymbol{\delta}^T$$

$$= 2(\mathbf{g}^{DF} - \mathbf{g}^{CP-PS}) + \boldsymbol{\delta}\boldsymbol{\delta}^T \tag{30}$$

where $\hat{\mathbf{B}}$ is the matricized form of the CP approximant in eq 10, and

$$\boldsymbol{\delta} \equiv \mathbf{B} - \hat{\mathbf{B}} \tag{31}$$

is the CP error tensor. Clearly, as the CP rank increases, the CP error $\boldsymbol{\delta}$ decreases but the CP-PS/CP-DF ratio of errors stays approximately 2. Since the rCP-DF is quadratic in $\boldsymbol{\delta}$, the rCP-DF error should decay with the CP rank faster than either that of CP-PS or CP-DF. The improvement of rCP-DF over CP-DF is approximately 1 order of magnitude for $R = 1.5X$ and approaches 2 orders of magnitude for $R = 5X$.

It is instructive to wonder whether it is possible to improve CP-PS and CP-DF approximations solely by relaxing the factors in the respective tensor networks approximating $g_{ab,cd}$. Indeed, it is important to recognize that CP-PS and CP-DF approximations utilize CP factorization of $\mathbf{B}$ that is optimal (in the

least-squares sense) for representing $\mathbf{B}$, not $\mathbf{g}$. It is therefore possible to optimize the factors in the tensor networks approximation of $\mathbf{g}$ directly. Partial relaxation of the factors in the CP-PS and CP-DF networks to minimize the error in $\mathbf{g}$ was already employed in some real-space-based THC developments by Parrish et al.,[24,27] and full relaxation of the CP-DF network cost was implemented by Schutski et al.[29] (e.g., see the discussion of their THC-ALS-RI solver). To investigate whether the suboptimality of the CP-DF network using the $\mathbf{B}$-optimized factors is significant, we implemented an ALS solver that minimizes the CP-DF error in $\mathbf{g}^{DF}$;[g] the operation complexity of such solver is identical to the $O(N^4)$ complexity of the ALS solver for the CP decomposition of $\mathbf{B}$, albeit the prefactor is somewhat larger. Only few iterations are needed to relax the CP-DF network fully with respect to $\mathbf{g}$ if we use, as the initial guess, the factors obtained by CP3 decomposing $\mathbf{B}$.

As the data in Sections 4.2 and 4.3 indicates, the tensor element errors obtained with the $\mathbf{g}$-optimized CP-DF network are moderately smaller than the errors of the reference CP-DF network, but still exceed the CP-PS errors and they are not competitive with the errors in the zero-cost robust CP-DF approximant. This observation suggests that the dominant source of error in the CP-DF (and CP-PS) approximants is the deficiency of the CP rank. The robust approximation is clearly able to greatly reduce both sources of error, due to the suboptimality (with respect to $\mathbf{g}^{DF}$) of the factors in the CP-DF network and due to the deficient CP rank.
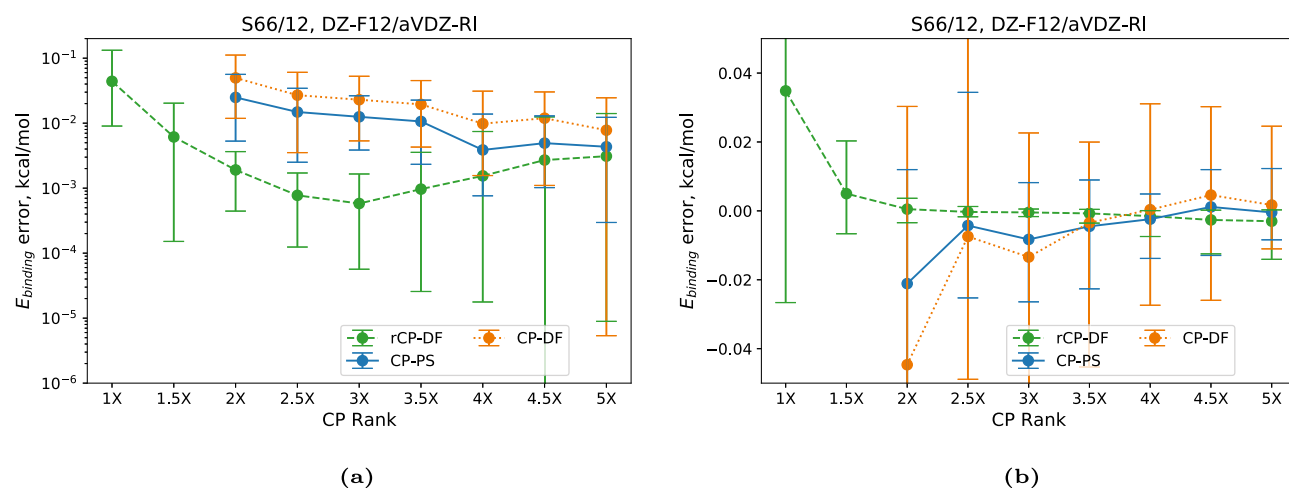
F

Figure 4. Mean unsigned (a) and signed (b) errors, respectively, in the CCSD binding energies (kcal/mol) of the S66/12 data set, relative to canonical CCSD, induced by the CP-DF, CP-PS, or rCP-DF approximations to PPL vs CP rank R (in units of the fitting basis, X). ALS precision fixed at $\epsilon = 10^{-3}$. The error bars denote the max/min errors.
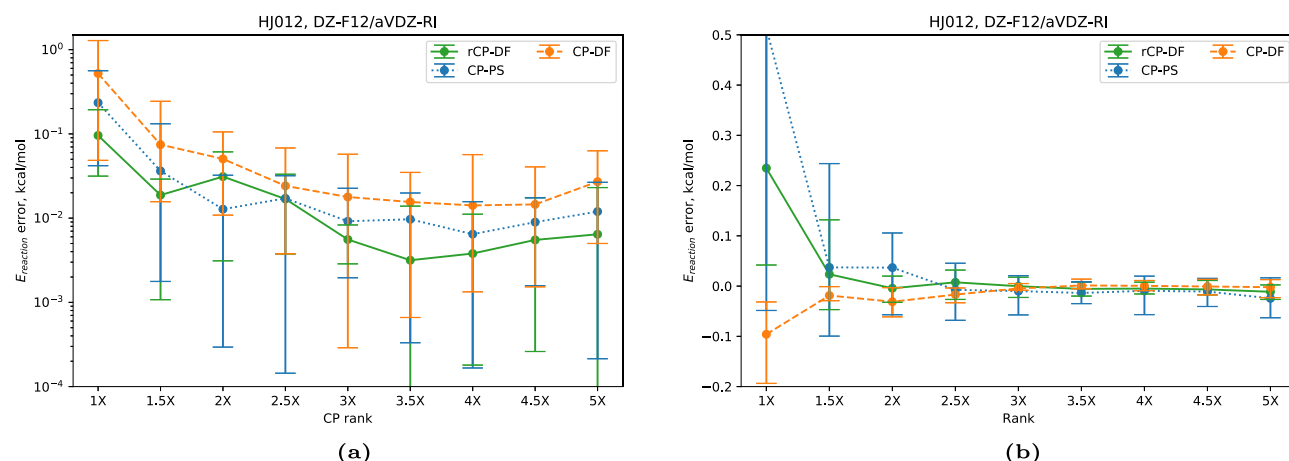

Figure 5. Mean unsigned (a) and signed (b) errors, respectively, in the CCSD reaction energies (kcal/mol) of the HJO12 data set, relative to canonical CCSD, induced by the CP-DF, CP-PS, or rCP-DF approximations to PPL vs CP rank R (in units of the fitting basis, X). ALS precision fixed at $\epsilon = 10^{-3}$. The error bars denote the max/min errors.

4.2. Errors in the CCSD Energies vs the CP Approximation Parameters. The error of the CP approximation is determined by the CP rank, R, and by the precision, ϵ, of the inexact CP solver (in our case, ALS); as already mentioned, we found negligible dependence of the ALS solution on the initial random guess. The ALS precision in this work is estimated by the difference between the current and previous iteration's decomposition "fit" Δ defined for eq 10 as

$$\Delta \equiv 1.0 - \frac{\left\| B_{ab,X} - \sum_r^R \beta_{a,r}\beta_{b,r}\gamma_{X,r} \right\|}{\|B_{ab,X}\|} = 1.0 - \frac{\|\delta\|}{\|B_{ab,X}\|} \tag{32}$$

where δ is the CP error tensor as defined in eq 31. Clearly, because ϵ depends on the change in the loss function, smaller values for ϵ do not necessarily lead to a smaller CP error. Thus, we first assessed how the error in E_{CCSD} due to the CP approximation depends on ϵ for a range of fixed CP ranks, R.

4.2.1. Variation of the CP Error with the ALS Solver Precision. Figure 3 reports the relationship between ϵ and the CP error in the valence CCSD correlation energy per electron for the S66/7 test set for CP ranks in the $X \leq R \leq 5X$ range. For

low CP ranks ($R \leq 2X$), the error varies little with ϵ. As CP rank increases, progressively smaller values of ϵ are required to obtain sufficiently converged ALS solutions. However, the effect of ϵ on the CCSD energy is significantly weaker than that of the CP rank R.

4.2.2. Variation of the CP Error with the CP Rank. Figure 3 indicates that increasing the CP rank R reduced the error in the CCSD energy monotonically. It also gives the first evidence of the performance advantage of rCP-DF over CP-DF and CP-PS. At $R = 1.5X$ (the red line in Figure 3c), rCP-DF is more accurate than both CP-DF and CP-PS with $R = 2X$ (the orange line in Figure 3a,3b). Furthermore, the error in CCSD energy is reduced at a fast rate, with respect to CP rank, for rCP-DF, which corroborates our discussion in Section 4.1. For each R and at converged ϵ, the rCP-DF approximation introduces an error which is at least an order of magnitude smaller than the error introduced by either CP-DF or CP-PS.

Then, we examined the influence of the CP rank on the errors in chemical energy differences, rather than in absolute correlation energies. The unsigned and signed errors in the weak noncovalent binding energies of the S66/12 test set and in the HJO12 isogyric reaction energies are reported in Figures 4
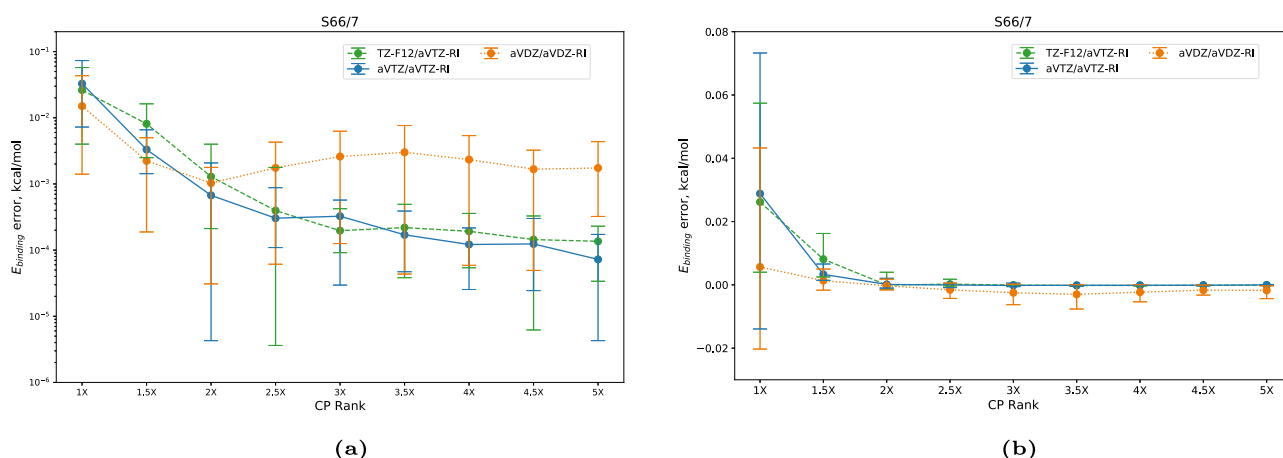
G

Figure 6. Mean unsigned (a) and signed (b) errors, respectively, in the CCSD binding energies (kcal/mol) for the S66/7 data set, relative to canonical CCSD, induced by the rCP-DF approximation to PPL vs CP rank R (in units of the fitting basis, X) using three different basis sets, aVDZ/aVDZ-RI, aVTZ/aVTZ-RI, and TZ-F12/aVTZ-RI. ALS precision fixed at $\epsilon = 10^{-3}$. The error bars denote the max/min errors.

and 5, respectively. Because, compared to R, ϵ has a relatively small influence on E_{CCSD}, we have limited this assessment to using relatively loose ALS tolerances of $\epsilon = 10^{-3}$.[h] The target level of performance, defined here stringently as the maximum error of less than 0.1 kcal/mol, is achieved with CP-DF and CP-PS when $R \geq 2X$. However, the use of rCP-DF allows us to attain the target accuracy with a much smaller CP rank, $R \geq X$. For all relevant CP ranks, rCP-DF is at least an order of magnitude more accurate than CP-DF and CP-PS. As expected, the CP-PS errors are roughly a factor of 2 smaller than those due to CP-DF.

The performance of the rCP-DF approximation to PPL is relatively insensitive to the basis set. Using the larger TZ-F12 OBS as well as the standard correlation-consistent aVD,TZ OBS does not appear to radically change the convergence trends, as illustrated in Figure 6.[i] The errors in binding energies are small (<0.1 kcal/mol even with $R = X$) and rapidly decrease when R is increased. The protracted convergence with the CP rank when using the aVDZ basis is somewhat puzzling, but is likely due to the need for tighter CP solver convergence for the smaller basis sets.

It is instructive to compare the rCP-DF approximation for the PPL diagram with the best THC-based approach for the same, namely the least-squares THC(DF) method [LS-THC(DF)] and its orbital-weighted extension [W-LS-THC(DF)] developed by Parrish et al.[27] Table 1 juxtaposes the maximum

Table 1. Maximum Absolute and Relative Errors in Valence TZ/TZ-RI DF-CCSD Correlation Energies (mE_h) of Eight Low-Lying $(H_2O)_6$ Clusters[78],[a]

	maximum absolute error	maximum relative error
rCP-DF	0.45	0.036
LS-THC(DF)[27]	2.13	0.18
W-LS-THC(DF)[27]	0.29	0.03

[a]For the rCP-DF approximation, CP rank and ALS precision were fixed at $R = 1.3X$ and $\epsilon = 10^{-3}$, respectively.

absolute and relative CCSD energy errors due to the rCP-DF and the THC PPL approximations for the eight low-lying $(H_2O)_6$ conformers. The same OBS/DFBS basis set pair, TZ/TZ-RI, was utilized for all computations. The rCP-DF approach used $R = 1.3X$, whereas the corresponding LS-THC grid size corresponds to $R \approx 4X$, i.e., roughly three times larger than used

by our method. Although the absolute energies are most accurate with the W-LS-THC(DF) method of Parrish et al., the relative energies of the clusters are nearly as accurate with our method, despite its much smaller CP rank. Most importantly, the rCP-DF approach greatly outperforms its true THC counterpart, LS-THC(DF), again despite the much smaller CP rank. It is clear that the errors of the rCP-DF approach can be reduced further in the context of the CC methods by combining it with the orbital-weighting idea of Parrish et al.[27]

4.3. Cost Reduction vs DF-CCSD. Then, we examined whether the stringent target errors in CCSD energies due to the rCP-DF PPL formulation can be attained along with demonstrated computational cost savings.

The observed speedups in the DF-CCSD computations due to the CP-based PPL reformulations are illustrated for the clusters in the S66/12 test set in Figure 7a. Just as in Section 4.2.2, only $\epsilon = 10^{-3}$ are reported in the paper, with the $\epsilon = 10^{-4}$ results available in the Supporting Information. Significantly smaller average speedups were observed with $\epsilon = 10^{-4}$ compared to $\epsilon = 10^{-3}$, for the same CP rank. This suggests that the cost of ALS CP solver can increase dramatically with ϵ, due to the increasing number of ALS iterations. To further illustrate this point, Figure 7b demonstrates the speedups obtained by excluding the cost of ALS. We see that ALS has the most dramatic effect on the cost when ϵ is tighter and R is larger.

Unsurprisingly, ALS optimization had the greatest impact on the smallest molecules. Figure 7c demonstrates that the speedup for the seven largest clusters in the S66/12 set is significantly greater than the average speedup over the entire set and for all values of R. Since we found the energies relatively insensitive to the choice of ϵ, we recommend the use of $\epsilon \approx 10^{-3}$ for all practical computations, unless extremely high target accuracy is sought.

We further assessed the performance of the rCP-DF PPL approximation for the S66/7 data set with three additional basis set pairs (Figure 8). As one might expect, for larger basis sets, like TZ-F12 or aVTZ, the PPL diagram contributes significantly more to the cost of CCSD; hence, even greater cost savings from rCP-DF are observed.

To further assess the performance of the rCP-DF PPL approximation, we computed the errors in CCSD binding energies for the entire S66 test set, using $R = 1.3X$ and $\epsilon = 10^{-3}$; the results are reported in Figures 9 and 10. For all systems, the
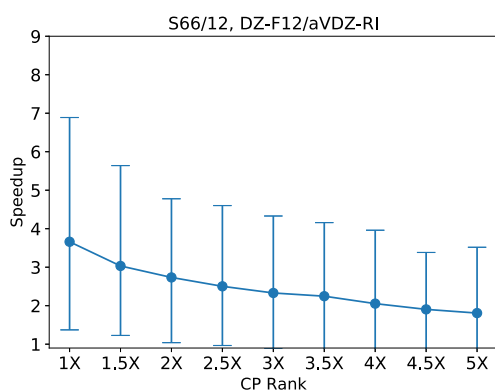
(a) Average speedup (Eq. (28)) of CCSD with rCP-DF-approximated PPL vs CP rank R (in units of the fitting basis, X) for the S66/12 dataset. ALS precision fixed at $\epsilon = 10^{-3}$. The error bars denote the max/min speedup.
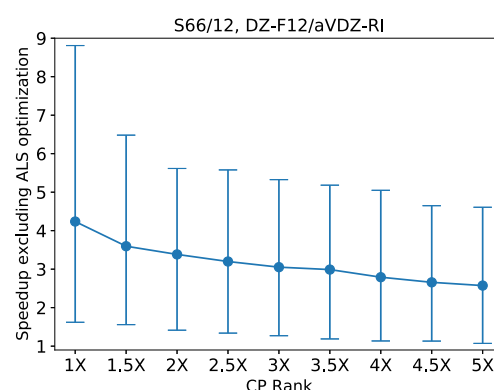
(b) Average speedup (Eq. (28), **excluding** the cost of CP-ALS) of CCSD with rCP-DF-approximated PPL vs CP rank R (in units of the fitting basis, X) for the S66/12 dataset. The error bars denote the max/min speedup.
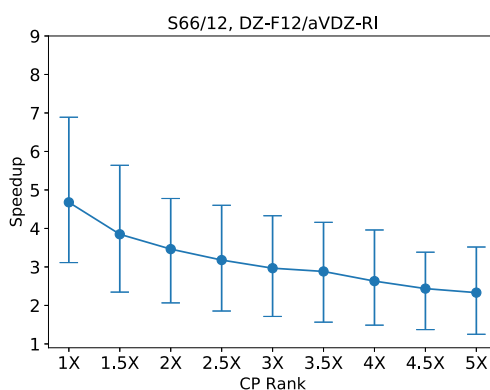
(c) Average speedup (Eq. (28)) of CCSD with rCP-DF-approximated PPL vs CP rank R (in units of the fitting basis, X) for the 7 largest clusters in the S66/12 dataset. The error bars denote the max/min speedup.

Figure 7. (a) Average speedup (eq 28) of CCSD with rCP-DF-approximated PPL vs CP rank R (in units of the fitting basis, X) for the S66/12 data set. ALS precision fixed at $\epsilon = 10^{-3}$. The error bars denote the max/min speedup. (b) Average speedup (eq 28, excluding the cost of CP-ALS) of CCSD with rCP-DF-approximated PPL vs CP rank R (in units of the fitting basis, X) for the S66/12 data set. The error bars denote the max/min speedup. (c) Average speedup (eq 28) of CCSD with rCP-DF-approximated PPL vs CP rank R (in units of the fitting basis, X) for the seven largest clusters in the S66/12 data set. The error bars denote the max/min speedup.
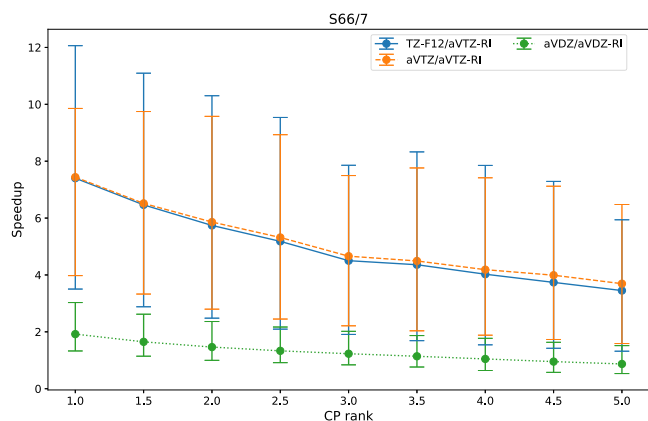

Figure 8. Average speedup (eq 28) of CCSD with rCP-DF-approximated PPL vs CP rank R (in units of the fitting basis, X) for the S66/7 data set. ALS precision fixed at $\epsilon = 10^{-3}$. The error bars denote the max/min speedup.

errors introduced by rCP-DF are significantly less than 0.1 kcal/mol, and the computational savings are realized for all systems, with the average speedups of 4. This figure shows a clear trend:

larger molecules benefit more from rCP-DF than smaller molecules. This trend is an artifact of the ALS optimization: as we increase the system size, the cost of CCSD increases faster than the cost of the ALS and, thus, computing the ALS takes up a smaller percentage of the total CCSD time, as illustrated in Figure 11. To note, although we only show speedup for the S66 cluster molecules, all of the dissociated cluster molecules also experienced a reduced cost over canonical DF-CCSD. The smallest dissociated molecule, a single water molecule, saw a cost reduction of a factor of 2. To demonstrate the performance of the DF-CCSD method with the rCP-DF-approximated PPL term for a larger system, we used it to compute the binding energy of $(H_2O)_{20}$, with results reported in Table 2. With the recommended values of R and ϵ, the cost of CCSD can be reduced by a factor of 3.8, with only a ~0.03 kcal/mol impact on the binding energy.

5. SUMMARY AND PERSPECTIVE

In this work, we considered how robust (in the Dunlap sense[53]) approximation of tensor networks, in which the leading-order error due to the approximation of the network constituents is explicitly canceled, can be used profitably to construct efficient
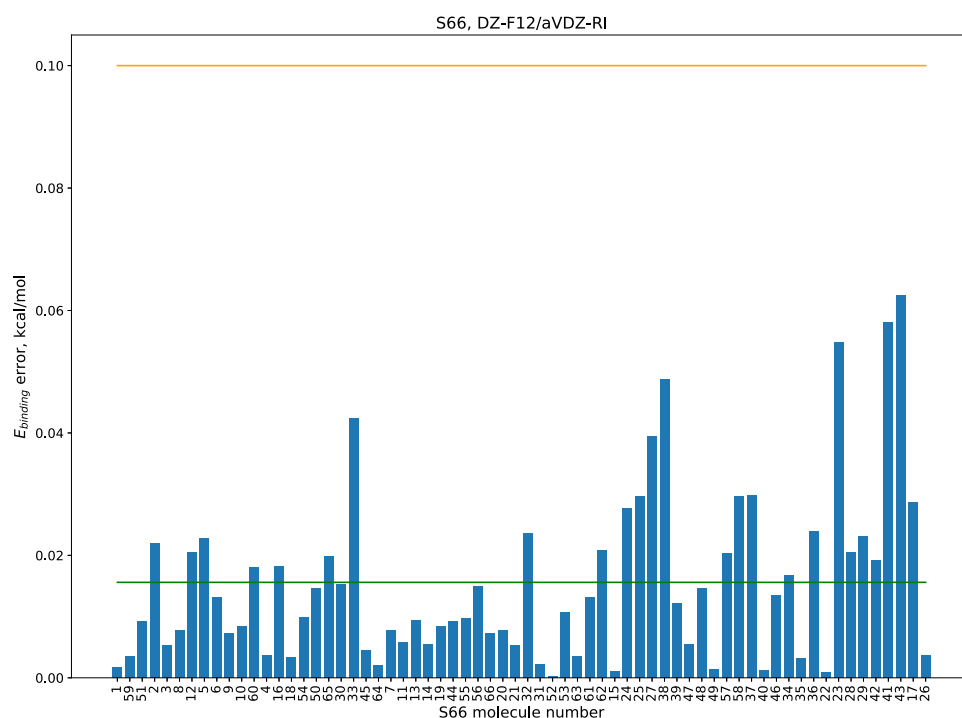
Figure 9. Unsigned errors in the S66 CCSD binding energies (kcal/mol), relative to canonical CCSD, induced by the rCP-DF approximation to PPL. CP rank and ALS precision are fixed at $R = 1.3X$ and $\epsilon = 10^{-3}$, respectively. Molecules ordered from the smallest to the largest number of occupied orbitals. The orange line is the target maximum error, 0.1 kcal/mol, and the green line is the average error of the set.
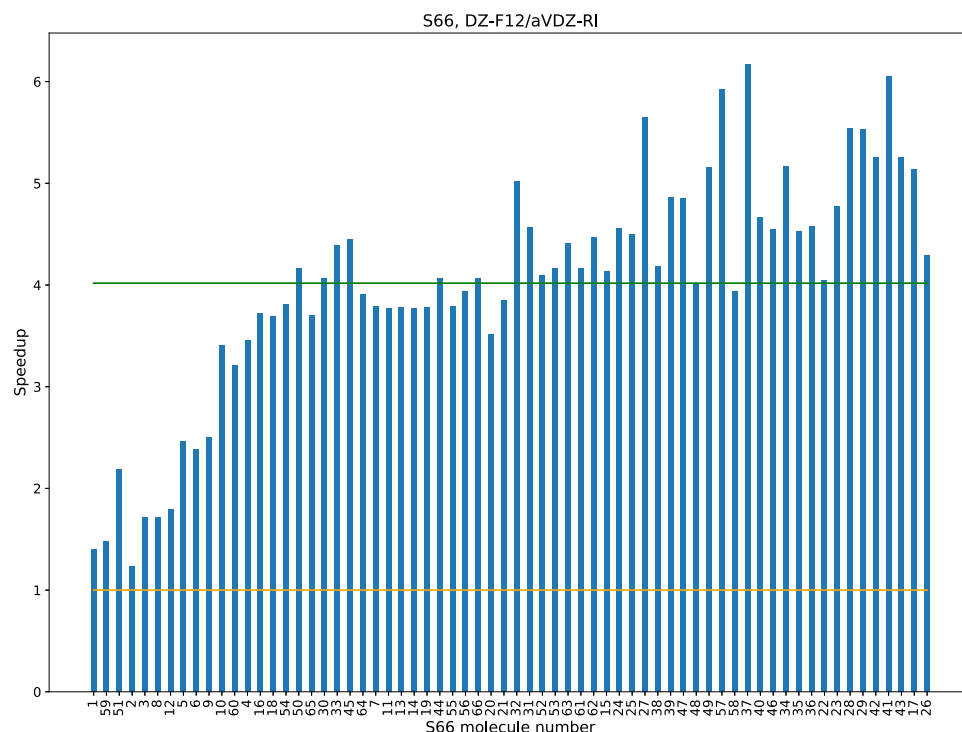

Figure 10. Speedup (eq 28) of CCSD with rCP-DF-approximated PPL for the entire S66 data set. CP rank and ALS precision are fixed at $R = 1.3X$ and $\epsilon = 10^{-3}$, respectively. Molecules are ordered according to the number of occupied orbitals, from smallest to largest. The orange line represents no speedup over CCSD and the green line is the average speedup of the set.

factorizations of the two-particle Coulomb interaction tensor. We specifically considered tensor networks utilizing CP decomposition of order-3 tensors that arise from generalized square-root factorizations of the Coulomb tensor, namely Cholesky and density fitting. The single use of the CP decomposition leads to a tensor network resembling the factorization in the well-known pseudospectral (PS) method, whereas double CP insertion leads to the tensor network topology of the tensor hypercontraction (THC) factorizations. Robust factorization combines these two base factorizations,
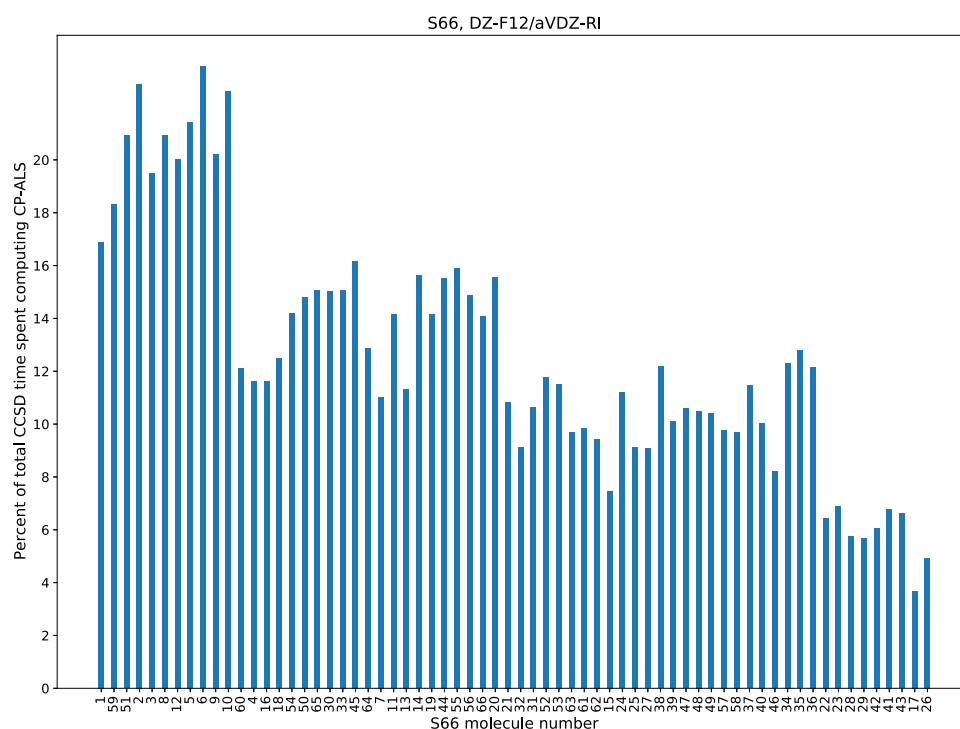
Figure 11. Percent of the total CCSD time spent in ALS for each cluster molecule in S66 data set using rCP-DF with CP rank $R = 1.3X$ and ALS precision of $\epsilon = 10^{-3}$. Molecules are ordered according to the number of occupied orbitals, from smallest to largest.

Table 2. Valence CCSD Correlation (E_{CCSD}, E_h) and Dissociation Energies (D_e, kcal/mol), the Average Per-Iteration Time Spent in CCSD (t_{CCSD}, s), and Its PPL Contribution (t_{PPL}, s) for the $(H_2O)_{20}$ Cluster[a]

	E_{CCSD}	D_e	t_{CCSD}	t_{PPL}	$t_{CP\text{-}ALS}$
DF	−5.02009	182.47	1.36×10^4	1.11×10^4	
CP	−5.02233	182.44	3.47×10^3	1.17×10^3	2.32×10^3
	error			speedup	
	-1.41×10^{-3}	2.81×10^{-2}		3.92	9.46

[a]The total time of the CP-ALS optimization is also reported ($t_{CP\text{-}ALS}$, s). CP rank and ALS precision are fixed at $R = 1.3X$ and $\epsilon = 10^{-3}$, respectively.

resulting in a 1−2 order reduction of the error over either naive substitution scheme. A deeper analysis of the errors in the Coulomb interaction tensor reveals that the novel factorization, dubbed rCP-DF, corrects both errors resulting from the suboptimality of the CP factors as well as the errors due to deficient CP rank.

As is also possible with the PS and THC factorizations, the rCP-DF-factorization lowers the operation complexity of the cost-dominant PPL diagram in pair theories from $O(N^6)$ to $O(N^5)$. Here, we demonstrated in practice that the rCP-DF-approximated PPL can lower the practical cost of DF-CCSD even for systems with as few as three atoms. We make this claim because sufficiently small (on the thermal energy scale) errors can be achieved with a CP rank approximately equal to the rank of the density-fitting basis itself; this hyperedge size requirement is substantially smaller than the requirements in previous PS and THC studies. For example, for the standard S66 and HJO12 benchmark sets of noncovalent interaction energetics and reaction energies, respectively, the use of such a low CP rank induces maximum errors of only ≈0.1 kcal/mol. For the larger

example of a 20-water cluster, the rCP-DF error in the dissociation energy was found to be only 0.03 kcal/mol.

Although the complexity reduction due to the use of rCP-DF is very modest, the use of rCP-DF-PPL in the context of divide-and-conquer reduced-scaling CC approaches like FMO,[87] CIM,[88] DEC,[89,90] and others[91] might be beneficial to reduce the cost of the fragment computation.

The proposed robust tensor factorization of the Coulomb interaction, clearly, can be improved further, as well as applied in other contexts. Some of the promising ideas are listed here:

- This particular robust CP-based factorization, which we consider here, utilized the density-fitting-based generalized square-root factorization of the Coulomb tensor, though it should be trivial to apply the factorization to other square-root factorizations, such as the (pivoted) Cholesky.

- Although we only considered algebraic CP decomposition of the square root factor, it should be possible to use the idea in the context of quadrature-based factorization, such as PS, COSX, and least-squares THC. For example, robust LS-THC should allow for the use of smaller grids than currently possible (the juxtaposition of the rCP-DF and LS-THC(DF) performance in Section 4.2.2, albeit limited, suggests that grid size reductions of a factor of 3 or more are realistic). Robust factorization should also simplify the formulation of analytic gradients.

- A combination with other ideas such as the use of orbital-biasing explored in LS-THC-based coupled-cluster[27] and the use of frozen natural orbitals should be beneficial.

- The efficiency of the CP solver can be greatly improved via the use of gradient-based techniques.

Work along some of these directions is underway.

■ ASSOCIATED CONTENT

ⓈⒾ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.0c01310.

Results with $\epsilon = 10^{-4}$ and the ALS algorithm for computing the optimal (for fixed rank) CP-DF approximation of the Coulomb tensor (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Edward F. Valeev − *Department of Chemistry, Virginia Tech, Blacksburg, Virginia 24061, United States;* ⓞ orcid.org/0000-0001-9923-6256; Email: efv@vt.edu

Authors

Karl Pierce − *Department of Chemistry, Virginia Tech, Blacksburg, Virginia 24061, United States*

Varun Rishi − *Department of Chemistry, Virginia Tech, Blacksburg, Virginia 24061, United States;* ⓞ orcid.org/0000-0003-0317-3815

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jctc.0c01310

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ ADDITIONAL NOTES

[a]In this work, the term "robust" mirrors its use in the discussion of fitting in quantum chemistry[53] rather than referring to the robust approximation of individual tensors.[54]

[b]In this work, we only consider Coulomb interactions using the Poisson kernel: $g(\mathbf{r}_1, \mathbf{r}_2) \equiv |\mathbf{r}_1 - \mathbf{r}_2|^{-1}$; extension to other multiplicative and nonmultiplicative kernels is straightforward.

[c]The term "hypercontraction" presumably refers to the appearance of hyperedges in the diagrammatic representation of CP-like tensor networks, e.g., Figure 1d.

[d]Note that the CP-PS approximation breaks particle equivalence symmetry and therefore, in practice, the result must be symmetrized with respect to the transpose of ia and jc index pairs.

[e]N.B. if $5X > 3R$, eq 23 can be reordered to compute $(\gamma\hat{B})$ more efficiently.

[f](1) Water⋯water, (2) water⋯MeOH, (3) water⋯MeNH$_2$, (4) MeNH$_2$⋯MeOH, (5) benzene⋯benzene $(\pi-\pi)$, (6) pyridine⋯pyridine $(\pi-\pi)$, (7) uracil⋯uracil $(\pi-\pi)$, (8) pentane⋯pentane, (9) benzene⋯benzene (TS), (10) benzene⋯ethyne (CH−π), (11) ethyne⋯water (CH−O), and (12) MeNH$_2$⋯pyridine.

[g]See the Supporting Information for the detailed algorithm description.

[h]The corresponding results for a tighter ALS tolerance, $\epsilon = 10^{-4}$, are reported in the Supporting Information.

[i]A note of caution to the readers not familiar with the D,TZ-F12 basis sets: they are actually quite a bit larger than their conventional counterparts and include even more diffuse Gaussians than the augmented correlation-consistent basis sets.

■ REFERENCES

(1) Whitten, J. L. Coulombic potential energy integrals and approximations. *J. Chem. Phys.* **1973**, *58*, 4496−4501.

(2) Vahtras, O.; Almlöf, J.; Feyereisen, M. W. Integral approximations for LCAO-SCF calculations. *Chem. Phys. Lett.* **1993**, *213*, 514−518.

(3) Scuseria, G. E.; Ayala, P. Y. Linear scaling coupled cluster and perturbation theories in the atomic orbital basis. *J. Chem. Phys.* **1999**, *111*, 8330−8343.

(4) Saebo, S.; Pulay, P. Local Treatment of Electron Correlation. *Annu. Rev. Phys. Chem.* **1993**, *44*, 213−236.

(5) Hampel, C.; Werner, H. Local treatment of electron correlation in coupled cluster theory. *J. Chem. Phys.* **1996**, *104*, 6286−6297.

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

(7) Friesner, R. A. Solution of the Hartree-Fock equations by a pseudospectral method: Application to diatomic molecules. *J. Chem. Phys.* **1986**, *85*, 1462−1468.

(8) Langlois, J.; Muller, R. P.; Coley, T. R.; Goddard, W. A.; Ringnalda, M. N.; Won, Y.; Friesner, R. A. Pseudospectral generalized valence-bond calculations: Application to methylene, ethylene, and silylene. *J. Chem. Phys.* **1990**, *92*, 7488−7497.

(9) Ringnalda, M. N.; Belhadj, M.; Friesner, R. A. Pseudospectral Hartree-Fock theory: Applications and algorithmic improvements. *J. Chem. Phys.* **1990**, *93*, 3397−3407.

(10) Friesner, R. A. New Methods For Electronic Structure Calculations on Large Molecules. *Annu. Rev. Phys. Chem.* **1991**, *42*, 341−367.

(11) Martinez, T. J.; Carter, E. A. Pseudospectral multireference single and double excitation configuration interaction. *J. Chem. Phys.* **1995**, *102*, 7564−7572.

(12) Martinez, T. J.; Mehta, A.; Carter, E. A. Pseudospectral full configuration interaction. *J. Chem. Phys.* **1992**, *97*, 1876−1880.

(13) Martinez, T. J.; Carter, E. A. Pseudospectral MøllerPlesset perturbation theory through third order. *J. Chem. Phys.* **1994**, *100*, 3631−3638.

(14) Ko, C.; Malick, D. K.; Braden, D. A.; Friesner, R. A.; Martínez, T. J. Pseudospectral time-dependent density functional theory. *J. Chem. Phys.* **2008**, *128*, No. 104103.

(15) Martinez, T. J.; Carter, E. A. Pseudospectral double excitation configuration interaction. *J. Chem. Phys.* **1993**, *98*, 7081−7085.

(16) Beebe, N. H. F.; Linderberg, J. Simplifications in the Two-Electron Integral Array in Molecular Calculations. *Int. J. Quantum Chem.* **1977**, *12*, 683−705.

(17) Löwdin, P.-O. Studies in Perturbation Theory. X. Lower Bounds to Energy Eigenvalues in Perturbation-Theory Ground State. *Phys. Rev.* **1965**, *139*, A357−A372.

(18) Löwdin, P.-O. Some properties of inner projections. *Int. J. Quantum Chem.* **2009**, *5*, 231−237.

(19) Folkestad, S. D.; Kjønstad, E. F.; Koch, H. An efficient algorithm for Cholesky decomposition of electron repulsion integrals. *J. Chem. Phys.* **2019**, *150*, No. 194112.

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

(21) Burant, J. C.; Strain, M. C.; Scuseria, G. E.; Frisch, M. J. Analytic energy gradients for the Gaussian very fast multipole method (GvFMM). *Chem. Phys. Lett.* **1996**, *248*, 43−49.

(22) Rudberg, E.; Sałek, P. Efficient implementation of the fast multipole method. *J. Chem. Phys.* **2006**, *125*, No. 084106.

(23) Hohenstein, E. G.; Parrish, R. M.; Martínez, T. J. Tensor hypercontraction density fitting. I. Quartic scaling second- and third-order Møller-Plesset perturbation theory. *J. Chem. Phys.* **2012**, *137*, No. 044103.

(24) Parrish, R. M.; Hohenstein, E. G.; Martínez, T. J.; Sherrill, C. D. Tensor hypercontraction. II. Least-squares renormalization. *J. Chem. Phys.* **2012**, No. 224106.

(25) Hohenstein, E. G.; Parrish, R. M.; Sherrill, C. D.; Martínez, T. J. Communication: Tensor hypercontraction. III. Least-squares tensor hypercontraction for the determination of correlated wavefunctions. *J. Chem. Phys.* **2012**, *137*, No. 221101.

(26) Hohenstein, E. G.; Kokkila, S. I. L.; Parrish, R. M.; Martínez, T. J. Quartic scaling second-order approximate coupled cluster singles and doubles via tensor hypercontraction: THC-CC2. *J. Chem. Phys.* **2013**, *138*, No. 124111.

(27) Parrish, R. M.; Sherrill, C. D.; Hohenstein, E. G.; Kokkila, S. I.; Martínez, T. J. Communication: Acceleration of coupled cluster singles and doubles via orbital-weighted least-squares tensor hypercontraction. *J. Chem. Phys.* **2014**, *140*, No. 181102.

(28) Shenvi, N.; Van Aggelen, H.; Yang, Y.; Yang, W.; Schwerdtfeger, C.; Mazziotti, D. Low rank factorization of the Coulomb integrals for periodic coupled cluster theory. *J. Chem. Phys.* **2013**, *139*, No. 054110.

(29) Schutski, R.; Zhao, J.; Henderson, T. M.; Scuseria, G. E. Tensor-structured coupled cluster theory. *J. Chem. Phys.* **2017**, *147*, No. 184113.

(30) Parrish, R. M.; Zhao, Y.; Hohenstein, E. G.; Martínez, T. J. Rank reduced coupled cluster theory. I. Ground state energies and wavefunctions. *J. Chem. Phys.* **2019**, *150*, No. 164118.

(31) Lee, J.; Lin, L.; Head-Gordon, M. Systematically Improvable Tensor Hypercontraction: Interpolative Separable Density-Fitting for Molecules Applied to Exact Exchange, Second- and Third-Order Møller-Plesset Perturbation Theory. *J. Chem. Theory Comput.* **2020**, *16*, 243−263.

(32) Carroll, J. D.; Chang, J. J. Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition. *Psychometrika* **1970**, *35*, 283−319.

(33) Harshman, R. In *Foundations of the PARAFAC Procedure: Models and Conditions for an "Explanatory" Multimodal Factor Analysis*, UCLA Working Papers in Phonetics, 1970; pp 1−84.

(34) Benedikt, U.; Auer, A. A.; Espig, M.; Hackbusch, W. Tensor decomposition in post-Hartree-Fock methods. I. Two-electron integrals and MP2. *J. Chem. Phys.* **2011**, *134*, No. 054118.

(35) Benedikt, U.; Böhm, K.-H.; Auer, A. A. Tensor decomposition in post-Hartree-Fock methods. II. CCD implementation. *J. Chem. Phys.* **2013**, *139*, No. 224101.

(36) Benedikt, U.; Auer, H.; Espig, M.; Hackbusch, W.; Auer, A. Tensor representation techniques in post-Hartree-Fock methods: matrix product state tensor format. *Mol. Phys.* **2013**, *111*, 2398−2413.

(37) Hummel, F.; Tsatsoulis, T.; Grüneis, A. Low rank factorization of the Coulomb integrals for periodic coupled cluster theory. *J. Chem. Phys.* **2017**, *146*, No. 124105.

(38) Böhm, K. H.; Auer, A. A.; Espig, M. Tensor representation techniques for full configuration interaction: A Fock space approach using the canonical product format. *J. Chem. Phys.* **2016**, *144*, No. 244102.

(39) Chinnamsetty, S. R.; Espig, M.; Khoromskij, B. N.; Hackbusch, W.; Flad, H.-J. Tensor product approximation with optimal rank in quantum chemistry. *J. Chem. Phys.* **2007**, *127*, No. 084110.

(40) Khoromskij, B.; Khoromskaia, V.; Chinnamsetty, S.; Flad, H.-J. Tensor decomposition in electronic structure calculations on 3D Cartesian grids. *J. Comput. Phys.* **2009**, *228*, 5749−5762.

(41) 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.

(42) Bischoff, F. A.; Valeev, E. F. Low-order tensor approximations for electronic wave functions: Hartree-Fock method with guaranteed precision. *J. Chem. Phys.* **2011**, *134*, No. 104104.

(43) Füsti-Molnár, L.; Pulay, P. The Fourier transform Coulomb method: Efficient and accurate calculation of the Coulomb operator in a Gaussian basis. *J. Chem. Phys.* **2002**, *117*, 7827−7835.

(44) Dutta, A. K.; Neese, F.; Izsák, R. Speeding up equation of motion coupled cluster theory with the chain of spheres approximation. *J. Chem. Phys.* **2016**, *144*, No. 034102.

(45) Izsák, R.; Neese, F.; Klopper, W. Robust fitting techniques in the chain of spheres approximation to the Fock exchange: The role of the complementary space. *J. Chem. Phys* **2013**, *139*, No. 094111.

(46) Izsák, R.; Hansen, A.; Neese, F. The resolution of identity and chain of spheres approximations for the LPNO-CCSD singles Fock term. *Mol. Phys.* **2012**, *110*, 2413−2417.

(47) Petrenko, T.; Kossmann, S.; Neese, F. Efficient time-dependent density functional theory approximations for hybrid density functionals: Analytical gradients and parallelization. *J. Chem. Phys.* **2011**, *134*, No. 054116.

(48) Izsák, R.; Neese, F. An overlap fitted chain of spheres exchange method. *J. Chem. Phys.* **2011**, *135*, No. 144105.

(49) Kossmann, S.; Neese, F. Efficient Structure Optimization with Second-Order Many-Body Perturbation Theory: The RIJCOSX-MP2 Method. *J. Chem. Theory Comput.* **2010**, *6*, 2325−2338.

(50) 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.

(51) Sun, X.; Pitsianis, N. P. A Matrix Version of the Fast Multipole Method. *SIAM Rev.* **2001**, *43*, 289−300.

(52) Börm, S.; Grasedyck, L.; Hackbusch, W. Introduction to hierarchical matrices with applications. *Eng. Anal. Boundary Elem.* **2003**, *27*, 405−422.

(53) Dunlap, B. I. Robust and variational fitting. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2113−2116.

(54) Goldfarb, D.; Qin, Z. T. Robust Low-Rank Tensor Recovery: Models and Algorithms. *SIAM J. Matrix Anal. Appl.* **2014**, *35*, 225−253.

(55) Reine, S.; Tellgren, E.; Krapp, A.; Kjærgaard, T.; Helgaker, T.; Jansik, B.; Høst, S.; Salek, P. Variational and robust density fitting of four-center two-electron integrals in local metrics. *J. Chem. Phys.* **2008**, *129*, No. 104101.

(56) Merlot, P.; Kjaergaard, 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.

(57) Håstad, J. Tensor rank is NP-complete. *J. Algorithms* **1990**, *11*, 644−654.

(58) Hillar, C. J.; Lim, L.-H. Most Tensor Problems Are NP-Hard. *J. ACM* **2013**, *60*, 1−39.

(59) Acar, E.; Dunlavy, D. M.; Kolda, T. G. A scalable optimization approach for fitting canonical tensor decompositions. *J. Chemom.* **2011**, *25*, 67−86.

(60) Sorber, L.; Van Barel, M.; De Lathauwer, L. Optimization-Based Algorithms for Tensor Decompositions: Canonical Polyadic Decomposition, Decomposition in Rank-$(L_r, L_r, 1)$ Terms, and a New Generalization. *SIAM J. Optim.* **2013**, *23*, 695−720.

(61) Phan, A. H.; Tichavský, P.; Cichocki, A. Fast alternating LS algorithms for high order CANDECOMP/PARAFAC tensor factorizations. *IEEE Trans. Signal Process* **2013**, *61*, 4834−4846.

(62) Kossmann, S.; Neese, F. Comparison of two efficient approximate Hartee-Fock approaches. *Chem. Phys. Lett.* **2009**, *481*, 240−243.

(63) Ten-no, S. Explicitly correlated second order perturbation theory: Introduction of a rational generator and numerical quadratures. *J. Chem. Phys.* **2004**, *121*, 117.

(64) Greeley, B. H.; Russo, T. V.; Mainz, D. T.; Friesner, R. A.; Langlois, J.-M.; Goddard, W. A.; Donnelly, R. E.; Ringnalda, M. N. New pseudospectral algorithms for electronic structure calculations: Length scale separation and analytical two-electron integral corrections. *J. Chem. Phys* **1994**, *101*, 4028.

(65) Parrish, R. M.; Hohenstein, E. G.; Schunck, N. F.; Sherrill, C. D.; Martínez, T. J. Exact Tensor Hypercontraction: AUniversal Technique for the Resolution of Matrix Elements of Local Finite-Range N-Body Potentials in Many-Body Quantum Problems. *Phys. Rev. Lett.* **2013**, *111*, No. 132505.

(66) Peng, B.; Kowalski, K. Highly Efficient and Scalable Compound Decomposition of Two-Electron Integral Tensor and Its Application in

Coupled Cluster Calculations. *J. Chem. Theory Comput.* **2017**, *13*, 4179−4192.

(67) 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.

(68) Mardirossian, N.; McClain, J. D.; Chan, G. K.-L. Lowering of the complexity of quantum chemistry methods by choice of representation. *J. Chem. Phys.* **2018**, *148*, No. 044106.

(69) Kroonenberg, P. M.; de Leeuw, J. Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika* **1980**, *45*, 69−97.

(70) Beylkin, G.; Mohlenkamp, M. M. J. Numerical operator calculus in higher dimensions. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 10246−10251.

(71) Uschmajew, A. Local Convergence of the Alternating Least Squares Algorithm for Canonical Tensor Approximation. *SIAM J. Matrix Anal. Appl.* **2012**, *33*, 639−652.

(72) Kolda, T. G.; Bader, B. W. Tensor Decompositions and Applications. *SIAM Rev.* **2009**, *51*, 455−500.

(73) Navasca, C.; De Lathauwer, L.; Kindermann, S. In *Swamp Reducing Technique for Tensor Decomposition*, 2008 16th European Signal Processing Conference, 2008; pp 1−5.

(74) Rezáč, J.; Riley, K. E.; Hobza, P. S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *J. Chem. Theory Comput.* **2011**, *7*, 2427−2438.

(75) Rezáč, J.; Jurečka, P.; Riley, K. E.; Černý, J.; Valdes, H.; Pluháčková, K.; Berka, K.; Rezáč, T.; Pitoňák, M.; Vondrášek, J.; Hobza, P. Quantum Chemical Benchmark Energy and Geometry Database for Molecular Clusters and Complex Molecular Systems (www.begdb.com): A Users Manual and Examples. *Collect. Czech. Chem. Commun.* **2008**, *73*, 1261−1270.

(76) Helgaker, T.; Jorgensen, P.; Olsen, J. *Molecular Electronic-Structure Theory*, 1st ed.; John Wiley & Sons, Ltd.: Chichester, U.K., 2000.

(77) Zhang, J.; Valeev, E. F. Prediction of Reaction Barriers and Thermochemical Properties with Explicitly Correlated Coupled-Cluster Methods: A Basis Set Assessment. *J. Chem. Theory Comput.* **2012**, *8*, 3175−3186.

(78) Bates, D. M.; Smith, J. R.; Tschumper, G. S. Efficient and Accurate Methods for the Geometry Optimization of Water Clusters: Application of Analytic Gradients for the Two-Body:Many-Body QM:QM Fragmentation Method to $(H_2O)_n$, $n = 3−10$. *J. Chem. Theory Comput.* **2011**, *7*, 2753−2760.

(79) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.

(80) Wales, D. J.; Hodges, M. P. Global minima of water clusters $(H_2O)_n$, $n \leq 21$, described by an empirical potential. *Chem. Phys. Lett.* **1998**, *286*, 65−72.

(81) Peterson, K. A.; Adler, T. B.; Werner, H.-J. Systematically convergent basis sets for explicitly correlated wavefunctions: The atoms H, He, B-Ne, and Al-Ar. *J. Chem. Phys.* **2008**, *128*, No. 084102.

(82) Weigend, F.; Köhn, A.; Hättig, C. Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations. *J. Chem. Phys.* **2002**, *116*, 3175−3183.

(83) 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.

(84) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96*, 6796−6806.

(85) BTAS Library. https://github.com/BTAS/BTAS (accessed April 9, 2009).

(86) Peng, C.; Lewis, C. A.; Wang, X.; Clement, M. C.; Pierce, K.; Rishi, V.; Pavošević, F.; Slattery, S.; Zhang, J.; Teke, N.; Kumar, A.; Masteran, C.; Asadchev, A.; Calvin, J. A.; Valeev, E. F. Massively Parallel Quantum Chemistry: A high-performance research platform for electronic structure. *J. Chem. Phys.* **2020**, *153*, No. 044120.

(87) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. Fragment molecular orbital method: an approximate computational method for large molecules. *Chem. Phys. Lett.* **1999**, *313*, 701−706.

(88) Li, W.; Piecuch, P.; Gour, J. R.; Li, S. Local correlation calculations using standard and renormalized coupled-cluster approaches. *J. Chem. Phys.* **2009**, *131*, No. 114109.

(89) Kristensen, K.; Ziółkowski, M.; Jansík, B.; Kjærgaard, T.; Jørgensen, P. A locality analysis of the divid-expand-consolidate coupled cluster amplitude equations. *J. Chem. Theory Comput.* **2011**, *7*, 1677−1694.

(90) Kjærgaard, T.; Baudin, P.; Bykov, D.; Kristensen, K.; Jørgensen, P. The divide-expand-consolidate coupled cluster scheme. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2017**, *9*, No. e1319.

(91) Friedrich, J.; Hanrath, M.; Dolg, M. Fully automated implementation of the incremental scheme: Application to CCSD energies for hydrocarbons and transition metal compounds. *J. Chem. Phys.* **2007**, *126*, No. 154110.