

Toward Using Matrix-free Tensor Decompositions to Systematically Improve Approximate Tensor-Networks

Karl Pierce*

Cite This: <https://doi.org/10.1021/acs.jctc.5c00413>

Read Online

ACCESS |



Metrics & More

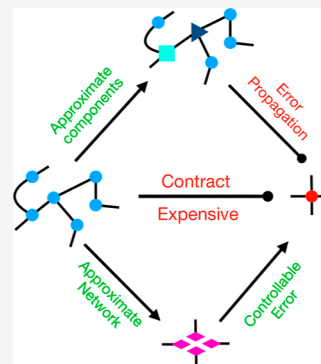


Article Recommendations



Supporting Information

ABSTRACT: We investigate a novel approach to approximate tensor-network contraction via the exact, matrix-free decomposition of full tensor-networks. We study this method as a means to eliminate the propagation of error in the approximation of tensor-networks. Importantly, this decomposition-based approach is generic, i.e., it does not depend on a specific tensor-network, the tensor index (physical) ordering, or the choice of tensor decomposition. Careful consideration should be made to determine the best decomposition strategy. Furthermore, this method does not rely on robust cancellation of errors (i.e., the Taylor expansion). As a means to study the effectiveness of the approach, we replace the exact contraction of the particle–particle ladder (PPL) tensor diagram in the popular coupled-cluster with single and double excitation (CCSD) method with a low-rank tensor decomposition, namely the canonical polyadic decomposition (CPD). With this approach, we replace an $O(N^6)$ tensor contractions with a potentially reduced-scaling $O(N^4R)$ optimization problem, where R is the CP rank, and we reduce the computational storage of the PPL tensor from $O(N^4)$ to $O(NR)$, although we do not take advantage of this compression in this study. To minimize the cost of the CPD optimization, we utilize the iterative structure of CCSD to efficiently initialize the CPD optimization. We show that accurate chemically relevant energy values can be computed with an error of less than 1 kcal/mol using a relatively low CP rank.



1. INTRODUCTION

Multi-index arrays, also known as tensors, are an integral tool in the translation of mathematical models into computer algorithms. Unfortunately, tensors are plagued by the so-called *curse of dimensionality* which dictates that the computational effort required to access and manipulate data in a tensor grows exponentially with the dimensionality of the tensor. Physics-based modeling algorithms, such as those found in electronic structure methods, are comprised of sets of tensor–tensor interactions commonly known as *tensor-networks*. In an effort to overcome the curse of dimensionality associated with the evaluation of these tensor-networks, researchers often find compressed or low-rank representations of tensor components in these networks. However, with each approximated component-tensor introduced into a tensor-network, a degree of error associated with that component is also introduced into the network. One can express the error in a component-wise approximated tensor-network as the error of each approximated component propagated through the tensor-network. The total accumulated error in an approximated tensor-network is therefore proportional to many variables such as total length of the tensor-network, total number of approximated tensor, error per approximated-tensor, etc. To mitigate this error propagation, researchers have found some success in using the Taylor series expansion to remove leading-order error in component-wise approximated tensor-networks.^{1,2} However,

this *robust* approximation approach cannot practically be applied to every approximated tensor-network.

In electronic structure methods there exists an incredibly large and diverse number of studied low-rank tensor approximations. These approximations can be organized into two categories: data-sparse methods and rank-sparse methods. Data-sparse methods (such as localization techniques like the pair-natural orbital approach,^{3–9} and fragmentation schemes^{10–15}) recast a higher-order tensor into a form which stores only nonzero valued elements of the tensor. Rank-sparse approaches (such as the density fitting approximation^{16–20}) recast higher-order tensors into compact tensor-network-based decompositions by revealing latent structures of the tensor. In principle, these techniques are integrated into electronic structure methods by first identifying high-cost tensors-networks and subsequently replacing single or multiple components of these networks with low-rank approximations. While there are many successful approximations introduced across the domain, here we investigate an alternative means to approximate a tensor-

Received: March 13, 2025

Revised: June 10, 2025

Accepted: June 11, 2025

network with directly controllable accuracy, as opposed to the indirect control associated with component-wise approximations.

The direction of this work is to extend the application of our previous work^{21,22} to approximate any high-cost tensor-network, not just the two-electron integral tensor. By construction, optimizing the low-rank decomposition of a tensor-network provides a direct means to control error. Importantly, with this approach we replace the cost of contracting a tensor-network with the cost of optimizing the low-rank decomposition and, furthermore, we replace the computational storage cost associated with the contracted tensor network with the cost associated with the chosen low-rank decomposition format. Although we should note, we do not take advantage of this storage benefit in this preliminary study. With these properties, we believe that this scheme is an innovative means to explore novel low-scaling formulations of electronic structure methods. To make our idea more concrete, in this work we construct the particle–particle ladder (PPL) diagram in the coupled cluster with single and double excitations (CCSD) method using a matrix-free canonical polyadic decomposition (CPD). We emphasize that this approach is a general technique that is not limited to this specific tensor-network, the coupled cluster method or the use of the CPD. It should also be noted that the focus of this study is to demonstrate the accurate application of matrix-free optimizations to electronic structure methods and we do not seek to reduce the wall-time of the CCSD method.

The rest of this manuscript is organized as follows. In Section 2 we introduce the coupled cluster method's particle–particle ladder (PPL) tensor-network, demonstrate how state of the art approximations to this network introduce unintended error and propose a reduced-scaling, matrix-free CPD approximation of the PPL tensor-network. In Section 3 we discuss the details of our computational experiments. In Section 4 we report on the capacity for a matrix-free CPD optimization to replace the exact PPL tensor-network. Finally, Section 5 summarizes our findings and provides future research directions for efficient application of this tensor-network approximation schemes to this and other threads of research.

2. FORMALISM

2.1. Coupled Cluster and the PPL Network. In this work we demonstrate how the decomposition of an entire tensor-network can replace the exact contraction of the tensor-network with controllable accuracy. With this approach we are provided a level of error control that is missing from today's standard methods for approximating tensor-networks. In this preliminary study, we apply this technique to a representative tensor-network from the popular CCSD excitation method. In this section we introduce the high-cost tensor-network from the CCSD method and demonstrate how the state-of-the-art tensor hyper-contraction approach introduces error into the approximation of the tensor-network. Please note, an in-depth description of the CC optimization will not be discussed in this work and readers can find extensive reviews on the topic, such as ref 23. In Section 2.3, we propose a method to approximate the tensor-network with reduced-scaling and controllable accuracy using a matrix-free tensor decomposition optimization strategy.

The goal of the coupled cluster method is to recover the 'dynamical correlation' that is missing from the Slater determinant picture of Hartree–Fock Theory.^{23–30} The

CCSD method achieves this goal by optimizing the weights of the single and double excitation amplitude tensors using a coupled set of nonlinear equations, i.e.

$$\begin{aligned} E_{\text{CC}} &= \langle \Phi_0 | \bar{H} | \Phi_0 \rangle \\ 0 &= \langle \Phi_i^a | \bar{H} | \Phi_0 \rangle \\ 0 &= \langle \Phi_{ij}^{ab} | \bar{H} | \Phi_0 \rangle \end{aligned} \quad (1)$$

where \bar{H} is the similarity transformed Hamiltonian and E_{CC} is the coupled cluster energy. The CCSD similarity transformed Hamiltonian can be expressed as $\bar{H} = e^{-\hat{T}_{\text{SD}}} H e^{\hat{T}_{\text{SD}}}$ where \hat{T}_{SD} is the excitation operator truncated at double excitations

$$\hat{T}_{\text{SD}} = \hat{T}_1 + \hat{T}_2 \quad (2)$$

The restricted, closed-shell single and double excitation operators can be expressed as

$$\begin{aligned} \hat{T}_1 &= t_i^a \hat{E}_a^i \\ \hat{T}_2 &= t_{ij}^{ab} \hat{E}_{ab}^{ij} \end{aligned} \quad (3)$$

where t_i^a is the singlet excitation amplitude tensor and $\hat{E}_a^i = a_{\alpha}^{\dagger} i_{\alpha} + a_{\beta}^{\dagger} i_{\beta}$ is the singlet excitation operator expressed as a sum of products of creation and annihilation operators. Similar definitions can be written for the t_{ij}^{ab} and \hat{E}_{ab}^{ij} . In eq 1, the labels i, j, k, \dots are representative of the set of O occupied orbital basis functions and the labels a, b, c, \dots are representative of the set of V unoccupied orbital basis functions. The weights, t_i^a and t_{ij}^{ab} , are then determined through the optimization of eq 1. Although the Coupled Cluster family of methods find great success in recovering dynamical correlation, the family is plagued by a restrictively large computational complexity.

The complexity of CCSD comes from the evaluation of tensor-networks in the residual equations of the nonlinear optimization of t_i^a and t_{ij}^{ab} . We introduce the most expensive term in the CCSD residual equations: the PPL tensor-network. In restricted, closed-shell format, the residual symmetrized PPL tensor-network can be written as

$$B_{ij}^{ab} = \sum_{cd} (g_{cd}^{ab} - 2 \sum_k g_{ck}^{ab} t_d^k) \tau_{ij}^{cd} \quad (4)$$

where $\tau_{ij}^{cd} = t_{ij}^{cd} + t_i^c t_j^d$ and g is the two-electron integral (TEI) tensor that will be discussed in Section 2.1.1. As an aside, eq 4 can be constructed using the intermediate \bar{g} such that

$$\bar{g}_{cd}^{ab} = g_{cd}^{ab} - 2 \sum_k g_{ck}^{ab} t_d^k \quad (5)$$

resulting in simplified expression

$$B_{ij}^{ab} = \sum_{cd} \bar{g}_{cd}^{ab} \tau_{ij}^{cd} \quad (6)$$

Figure 1 shows the diagrammatic expression of the PPL tensor-network (eq 6). In this figure, tensors are represented as nodes on a graph and tensor indices are edges. A tensor contraction (or summation over an index) is denoted as a solid line connecting two nodes and, later, a hyperedge (an index which connects more than two nodes) will be denoted as a dashed line. Also, the order of a tensor is defined as the number of indices or edges which belong to a tensor. From either of these representations, one can see that the computational complexity of computing the

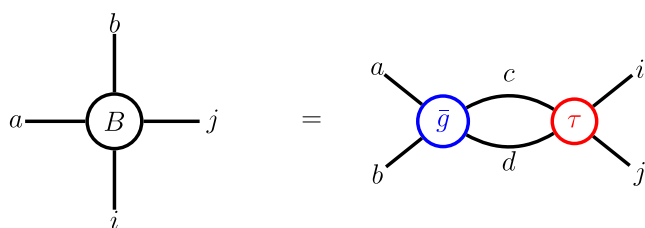


Figure 1. Diagrammatic representation of the PPL tensor-network.

PPL tensor-network is $O^2V^4 \approx O(N^6)$ where N is a measure of the system-size.

Although there are many formally $O(N^6)$ terms in the CCSD residual equation, the PPL tensor-network is the only term that carries an O^2V^4 cost of evaluation. Because of the PPL tensor-network's dominating cost and the prevalence of previous studies which have approximated the PPL tensor-network using component-wise decomposition schemes,^{2,31–35} we choose to apply our decomposition strategy to only this term in CCSD. Please note that our goal is not to eliminate the error associated with approximating the PPL tensor network, but to demonstrate that it is possible to replace a tensor-network contraction with a low-rank tensor decomposition. Therefore, error will still exist and propagate into the optimization of the wave function excitation tensors. Although, we believe that the analysis and mitigation of this error propagation should be considered in future studies. Below, we will introduce two canonical decomposition schemes and demonstrate how they can introduce a degree of error propagation in the evaluation of the PPL tensor-network.

2.1.1. The Density Fitting Approximation. The most standard approximation made in electronic structure methods is the density fitting (DF) approximation^{16–19} of the TEI tensor, g . The elements of the TEI tensor evaluated in a generic set of basis functions $\{\phi\}$ are

$$g_{st}^{pq} = \iint \phi_p^*(r_1) \phi_q^*(r_2) g(r_1, r_2) \phi_s(r_1) \phi_t(r_2) dr_1 dr_2 \quad (7)$$

In general, $g(r_1, r_2)$ may be any positive kernel and for this work we utilize the Coulomb interaction kernel, i.e., $g(r_1, r_2) \equiv |r_1 - r_2|^{-1}$. The DF approximation is a low-rank representation of the TEI tensor defined as

$$g_{st}^{pq} \stackrel{\text{DF}}{\approx} \sum_X D_s^{pX} D_t^{qX} \quad (8)$$

where the index X here represents an optimized, predetermined auxiliary set of basis functions which grows linearly with system size and D is denoted as the DF approximated TEI tensor. It should be noted that the DF approximation can alternatively be constructed analytically using the Cholesky decomposition^{36–39} or the related chain-of-spheres (COSX) method.^{40–46} One can make eq 8 exact by introducing a DF error correction term $(\Delta g)_{st}^{pq}$ into the equation, i.e.

$$g_{st}^{pq} = \sum_X D_s^{pX} D_t^{qX} + (\Delta g)_{st}^{pq} \quad (9)$$

Because of the significant compression gained from the DF approximation, the error associated with replacing the TEI tensor with its DF approximation is widely accepted and this work is no exception.^{47–49} By this we assume $(\Delta g)_{st}^{pq} \approx 0$ for all DF approximated TEI tensors and in CCSD, we will replace all TEI tensors with their DF approximations, accordingly.

In principle, the goal of introducing the DF approximation is to reduce the computational complexity of the tensor-networks in the CCSD method. In practice, the DF approximation fails to reduce the complexity of all tensor-networks in the CCSD method. Specifically, it is unable to reduce the complexity of the PPL tensor-network. By substituting the appropriate DF approximated TEI tensors into eq 4 we find the following expression

$$B_{ij}^{ab} = \sum_{cd} \sum_X (D_c^{aX} D_d^{bX} - 2 \sum_k D_c^{aX} D_k^{bX} t_{ij}^{ck}) \tau_{ij}^{cd} \quad (10)$$

To simplify this expression, it is possible to form an intermediate \bar{D} term which is similar to \bar{g} from eq 5

$$\bar{D}_d^{bX} = (D_d^{bX} - 2 \sum_k D_k^{bX} t_{ij}^{ck}) \quad (11)$$

Substituting \bar{D} into eq 10 simplifies the expression to be

$$B_{ij}^{ab} = \sum_{cd} \left(\sum_X D_c^{aX} \bar{D}_d^{bX} \right) \tau_{ij}^{cd} \quad (12)$$

We express the diagrammatic representation of this network in Figure 2 and this figure allows one to easily see that the indices

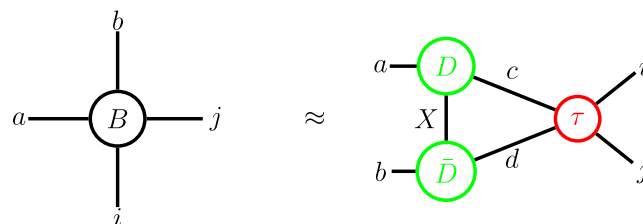


Figure 2. Diagrammatic representation of the DF approximated PPL tensor-network.

c and d occupy different order-3 D tensors. By construction, the DF approximation pins pairs of orbital indices of a single electronic particle to the same D tensor and, therefore, this observation means that c and d are indices of different particles. The DF approximated PPL tensor-network does not have a reduced complexity over the canonical PPL tensor-network.

Because the computation of the PPL tensor-network is a significant bottleneck in the CCSD optimization procedure, theorists have studied additional approximations that can effectively reduce the scaling of this tensor-network. These include the pseudospectral^{34,35,50–57} and the tensor hyper-contraction^{31–33,58–73} methods. In the following section, we will discuss how the popular tensor hyper-contraction is formulated and how the approximation introduces error that is propagated through the PPL tensor-network.

2.1.2. Propagation of Error in the THC Approximation. As shown in the previous section, the DF approximation cannot reduce the computational scaling of the expensive PPL tensor-network. It is well understood that the tensor hyper-contraction (THC) method^{32,61–63} can be deployed to reduce the complexity of this tensor-network. Therefore, in this section we introduce the THC approximation and highlight the error it contributes to the construction of the PPL tensor-network. Please note, we choose this example because its behavior has been studied and documented² and not because we believe that the THC method does a poor job approximating the PPL tensor-network or because this error propagation issue is unique to the THC approximation.

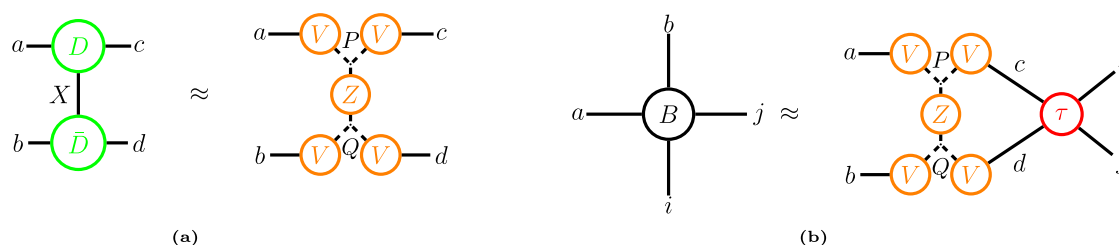


Figure 3. (a) Diagrammatic representation of the tensor hyper-contraction approximation, (b) diagrammatic representation of the THC approximation of the PPL tensor-network.

The THC is a low-rank decomposition of the order-4 TEI tensor such that

$$g_{cd}^{ab} \approx \sum_{PQ} V_P^a V_c^P Z_Q^P V_Q^b V_d^Q \quad (13)$$

The THC representation of the DF approximated TEI can be visualized using Figure 3a. In principle, the THC can be formed by replacing both order-3 DF approximated TEI tensors with the low-rank approximation

$$D_q^{pX} \approx \sum_P V_P^p V_q^P V_X^P \quad (14)$$

where the THC rank, P , can be determined analytically using a rank revealing CPD or using a predefined grid basis set. In eq 14, the V tensors are denoted as THC factor matrices. In this example, the THC core matrix is defined as $Z_Q^P = \sum_X V_X^P V_Q^X$. However, Z can also be computed using alternative methods, such as a low-cost least-squares optimization⁶⁰ which will not be discussed in this work.

We can make the approximation in eq 14 exact by introducing a THC approximation error term ΔD

$$D_q^{pX} = \sum_P V_P^p V_q^P V_X^P + (\Delta D)_q^{pX} \quad (15)$$

And for simplicity in the following analysis we will define

$$\tilde{D}_q^{pX} = \sum_P V_P^p V_q^P V_X^P \quad (16)$$

where a tilde indicates the THC approximation of a given tensor; therefore

$$D_q^{pX} = \tilde{D}_q^{pX} + (\Delta D)_q^{pX} \quad (17)$$

If we substitute eq 17 into the definition of the DF approximation of the TEI tensor, eq 8, we find the following

$$g_{cd}^{ab} = \sum_X (\tilde{D} + \Delta D)_c^{aX} (\tilde{D} + \Delta D)_d^{bX} \quad (18)$$

Equation 18 can be expanded into

$$g_{cd}^{ab} = \sum_X \tilde{D}_c^{aX} \tilde{D}_d^{bX} + (\Delta D)_c^{aX} \tilde{D}_d^{bX} + \tilde{D}_c^{aX} (\Delta D)_d^{bX} + (\Delta D)_c^{aX} (\Delta D)_d^{bX} \quad (19)$$

and, as one can see from this expression, the accuracy of the THC approximated TEI tensor depends on the error associated with the THC approximation propagated through eq 8. Similar to eq 18, we can define the error in the THC approximation of g as

$$(\Delta g)_{cd}^{ab} = \sum_X (\Delta D)_c^{aX} \tilde{D}_d^{bX} + (\Delta D)_c^{aX} \tilde{D}_d^{bX} + \sum_X (\Delta D)_c^{aX} (\Delta D)_d^{bX} \quad (20)$$

This error term simplifies eq 19 into the following expression

$$g_{cd}^{ab} = \tilde{g}_{cd}^{ab} + (\Delta g)_{cd}^{ab} \quad (21)$$

where $\tilde{g}_{cd}^{ab} \equiv \sum_{PQ} V_P^a V_c^P Z_Q^P V_Q^b V_d^Q$ is the THC approximation of g .

Introducing eq 21 into the PPL tensor-network results in the following expression

$$B_{ij}^{ab} = \sum_{cd} (\tilde{g}_{cd}^{ab} + (\Delta g)_{cd}^{ab}) \tau_{ij}^{cd} = \sum_{cd} \tilde{g}_{cd}^{ab} \tau_{ij}^{cd} + (\Delta B)_{ij}^{ab} \quad (22)$$

where $(\Delta B)_{ij}^{ab}$ is the error in the THC approximated PPL tensor-network. As an aside, the THC approximated PPL tensor-network is represented in Figure 3b and with this figure, it is easy to see how the THC approximation can be used to reduce the complexity of the PPL tensor-network. The total, accumulated error in the THC approximated PPL tensor-network can be expanded as

$$\Delta B_{ij}^{ab} = \sum_{cd} (\Delta g)_{cd}^{ab} \tau_{ij}^{cd} = \sum_{Xcd} (\Delta D)_c^{aX} \tilde{D}_d^{bX} \tau_{ij}^{cd} + \sum_{Xcd} \tilde{D}_c^{aX} (\Delta D)_d^{bX} \tau_{ij}^{cd} + \sum_{Xcd} (\Delta D)_c^{aX} (\Delta D)_d^{bX} \tau_{ij}^{cd} \quad (23)$$

This equation shows that the error in the PPL tensor-network is a sum of error tensors propagated through the PPL tensor network. In our previous work,² we introduced the robust approximation of the THC approximated TEI tensor. This approximation uses a Taylor series expansion to identify terms which can be added to a component-wise approximated tensor network to cancel the leading-order error propagation. The error in the robust THC approximated PPL tensor-network can be written as

$$\Delta B_{ij}^{ab} \stackrel{rCP}{=} \sum_{Xcd} (\Delta D)_c^{aX} (\Delta D)_d^{bX} \tau_{ij}^{cd} \quad (24)$$

Although significantly suppressed, this equation shows that the error in the robust component-wise approximated PPL tensor-network is still a result of error-propagation, i.e., it cannot be determined by the error in the THC approximation alone. Furthermore, although the robust approximation was effective at

mitigating error propagation for this network, it is not easily applicable to all tensor-networks.

In the remainder of this work, we introduce a method to approximately contract the PPL tensor-network in a way that makes the error in the contraction a directly controllable component of our approximation. To accomplish this task, we compute the canonical polyadic decomposition of the entire DF approximated PPL tensor-network. In the following section, we remind the readers on the details of the canonical polyadic decomposition before using it to approximate the PPL tensor-network.

2.2. The Canonical Polyadic Decomposition. The CANDECOMP/PARAFAC, or, simply, the canonical polyadic decomposition (CPD)^{74,75} is a data-sparse representation of higher-order tensors that has been used with various success in computational modeling to reduce the cost and/or complexity of the tensor algebra which accompanies physical simulation. We discuss various applications of the CPD to electronic structure methods in Section 2.5. The CPD of an order- n tensor can be understood as a compression of a tensor of n modes to a tensor-network of n order-2 tensors connected by a single hyperedge. For example, given the tensor $T \in \mathbb{R}^{I_p \times I_q \times I_s}$, the CPD can be expressed as

$$T_{pqs}^{\text{CPD}} \approx \tilde{T}_{pqs} = \sum_r \lambda_r M_{pr} N_{qr} O_{sr} \quad (25)$$

where $[M, N, O]$, are the set of CP factor matrices and R_{CP} is the CP rank of the decomposition. We enforce that the factor matrices be column-wise normalized and store the column-wise scaling factors in λ . Without loss of generality, it is possible to factorize lambda into any one or many of the factor matrices and, therefore, we will ignore it in future CP-based equations. Notice, too, that we denote the CP approximation of a tensor with a double tilde. The diagrammatic representation of the CPD can be found in Figure 4 where the dashed line represents the CP hyperedge index.

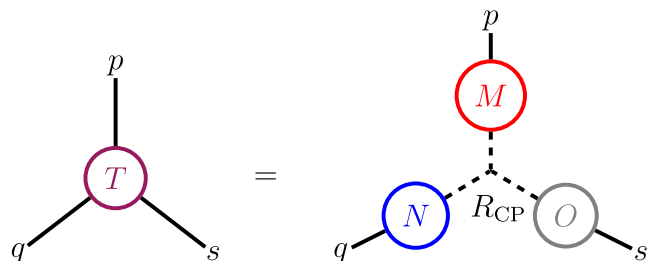


Figure 4. Diagrammatic representation of the canonical polyadic decomposition of a tensor T .

As one can see, the topology of the CPD is relatively simple and the effectiveness of the CP compression depends explicitly on the extent of the CP hyperedge (CP rank).^{76,77} Unfortunately, there exists no simple formula that can determine the CP rank. Therefore, the value is revealed by constructing multiple rank- R CPD approximations and choosing the value of R which satisfies the problem's predetermined accuracy thresholds. Therefore, a more accurate representation of the CPD would be

$$T_{pqs} = \sum_r M_{pr} N_{qr} O_{sr} + (\Delta T)_{pqs} \quad (26)$$

where ΔT is the error in the CPD approximation. In the following sections we demonstrate how we utilize the CPD to approximate the PPL tensor-network in CCSD.

2.3. The CPD Approximation of the PPL Tensor-Network. Our goal is to develop a method to approximate tensor-networks such that the computational complexity of the tensor-network is reduced and the error in the network's approximation is exactly controllable through the optimization process. In this example, we achieve this goal by representing the DF approximated PPL tensor-network (eq 12) in a low-rank CPD format

$$B_{ij}^{ab} \stackrel{\text{CP-PPL}}{\approx} \sum_r \alpha_r^a \beta_r^b \gamma_r^i \rho_r^j \quad (27)$$

in which we denote the CP-PPL approximation. In this equation R is the rank of the decomposition and α, β, γ and ρ , are the CP factor matrices. A diagrammatic representation of the CP-PPL approximation is shown in Figure 5.

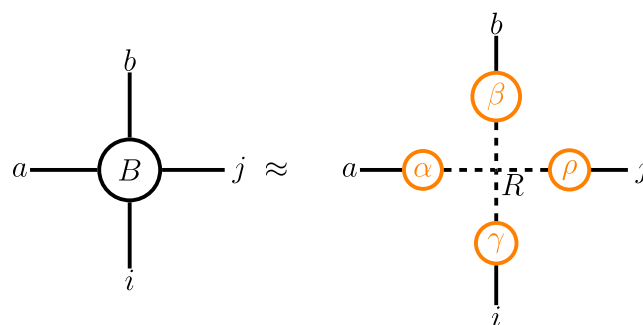


Figure 5. Diagrammatic representation of CPD approximation of the PPL tensor.

Similar to the previous approximations, the CP-PPL can be made exact by introducing the CP-PPL error tensor ΔB . For convenience we define the CPD approximation

$$\tilde{B}_{ij}^{ab} = \sum_r \alpha_r^a \beta_r^b \gamma_r^i \rho_r^j \quad (28)$$

and the exact decomposition as

$$B_{ij}^{ab} = \tilde{B}_{ij}^{ab} + \Delta B_{ij}^{ab} \quad (29)$$

With this approximation the error in the CP-PPL is correlated with the compressibility of the tensor B and the ability to find accurate CP factor matrices for a given rank.

We construct an accurate rank R CP-PPL approximation in a *matrix-free* way by minimizing the following loss function

$$f(\tilde{B}) = \frac{1}{2} \left\| \sum_{cd} \left(\sum_X D_c^{aX} \tilde{D}_d^{bX} \right) \tau_{ij}^{cd} - \tilde{B}_{ij}^{ab} \right\|_2^2 \quad (30)$$

The error in the approximated CP-PPL tensor-network is then exactly the evaluation of the function $f(\tilde{B})$.

In principle this does not mean that a given fixed rank CP-PPL decomposition will be a better approximation than a given component-wise approximation of the network, just that the error in the approximate contraction of the tensor-network is controllable through the minimization of f . The optimization of \tilde{B} is called *matrix-free* because we effectively “load” the output of the PPL network into the CPD factor matrices without a-priori

computing the tensor B . To follow, we attempt to replace the exact PPL tensor-network with the CP-PPL optimization, eq 30.

It should be noted that, in this study, \tilde{B} is not preserved in the CPD form after a rank R optimization is completed. Instead the tensor is reconstructed into its canonical form for the remainder of the CCSD optimization algorithm. We also point out that we do not explicitly enforce symmetry in our minimization of eq 30, i.e., $\tilde{B}_{ij}^{ab} \neq \tilde{B}_{ji}^{ba}$. But, because we do not preserve the low-rank structure of B , we recover this symmetry in the post tensor-reconstruction by defining

$$\tilde{B}_{ij}^{ab} = \frac{1}{2}(\tilde{B}_{ij}^{ab} + \tilde{B}_{ji}^{ba}) \quad (31)$$

For a given CP rank we minimize eq 30 using a simple alternating least-squares (ALS) optimization^{78,79} and in the following section we evaluate the computational scaling of each step in the CP-PPL ALS optimization.

2.4. Optimizing the CP-PPL Tensor-Network. In this section, we walk through the computational cost associated with optimizing eq 30 using an ALS algorithm. In doing so, we demonstrate that such an optimization has a reduced-scaling compared to the canonical evaluation of the PPL tensor-network. In general, computing a CP-ALS update involves four steps. First, one chooses a CPD factor matrix in the cost function, for example the alpha from the n th iteration (α_n), and fixes the other factors. Second, one takes the gradient of the CP loss function, here eq 30, with respect to the chosen factor matrix. The gradient of $f(\tilde{B}_n)$ with respect to α_n is

$$\begin{aligned} \frac{\partial f(\tilde{B}_n)}{\partial \alpha_n} &= \frac{1}{2} \frac{\partial}{\partial \alpha_n} (B^2 - B\tilde{B}_n - \tilde{B}_n B + \tilde{B}_n^2) \\ &= -B \frac{\partial \tilde{B}_n}{\partial \alpha_n} + \tilde{B}_n \frac{\partial \tilde{B}_n}{\partial \alpha_n} \end{aligned} \quad (32)$$

where \tilde{B}_n is the CPD approximation of B during at the n th ALS iteration and

$$\left(\frac{\partial \tilde{B}_n}{\partial \alpha_n} \right)_{ij}^{rb} = (\beta_n^b)_r (\gamma_n^r)_i (\rho_n^r)_j \quad (33)$$

Next, one sets the gradient expression from the previous step equal to zero and solves the developed linear problem to determine a new optimal factor matrix, α_{n+1} . This process can be repeated for each of the problem's other factor matrices and sweeps of factor matrix updates can be repeated as many times as necessary. Finally, after a round of updating each factor matrix a stopping condition is assessed. We will discuss appropriate stopping conditions in Section 2.4.1. Next, we evaluate the cost of computing the CP-PPL gradient and finding an updated CP factor matrix.

To exemplify the computational cost of the optimization, we walk through the optimization of the α_{n+1} factor matrix. Please note, in this analysis we assume that the CP rank grows linearly with system size based on findings compiled in Section 4 and, for simplicity, we drop the ALS iteration label (n). The α_{n+1} solution to the ALS update for the CP-PPL cost function is

$$[\alpha_{n+1}]_{r'}^a = \sum_{cd} \left(\sum_{cbijr} D_c^{aX} \bar{D}_d^{bX} \right) t_{ij}^{cd} \beta_r^b \gamma_i^r \rho_j^r \left[\left(\sum_b \beta_r^b \beta_{r'}^b \right) \left(\sum_i \gamma_i^r \gamma_{i'}^{r'} \right) \left(\sum_j \rho_j^r \rho_{j'}^{r'} \right) \right]^{-1} \quad (34)$$

In the literature, it is common to group all of the terms in the square bracket into a term denoted as $W_{r'}^r$.

$$W_{r'}^r = \left(\sum_b \beta_r^b \beta_{r'}^b \right) \left(\sum_i \gamma_i^r \gamma_{i'}^{r'} \right) \left(\sum_j \rho_j^r \rho_{j'}^{r'} \right) \quad (35)$$

Computation of W can be done efficiently with a cost of $3NR^2 + 2R^2 \approx O(N^3)$ and its inversion can also be done conveniently using a symmetric, square linear solve algorithm such as the Cholesky decomposition. The inversion of W has a cost of $R^3 \approx O(N^3)$.

Substituting W into eq 34 simplifies the equation to

$$[\alpha_{n+1}]_{r'}^a = \sum_r \left[\sum_{Xc} D_c^{aX} \left(\sum_d \left(\sum_b \bar{D}_d^{bX} \beta_r^b \right) \left(\sum_{ij} \gamma_i^r \rho_j^r \tau_{ij}^{cd} \right) \right) \right] [W^{-1}]_{r'}^r \quad (36)$$

Next, one can compute the two innermost contractions $\tau_r^{cd} = \sum_{ij} \gamma_i^r \rho_j^r \tau_{ij}^{cd}$ and $\bar{D}_d^{rX} = \sum_b \bar{D}_d^{bX} \beta_r^b$. These contractions have a cost of $N^4 R \approx O(N^5)$ and $N^3 R \approx O(N^4)$, respectively. The intermediates produced from these contractions can be cached to help minimize the cost of repeated gradient evaluations. After these contractions the following equation remains

$$[\alpha_{n+1}]_{r'}^a = \sum_r \left[\sum_{Xc} D_c^{aX} \left(\sum_d \bar{D}_d^{rX} \tau_r^{cd} \right) \right] [W^{-1}]_{r'}^r \quad (37)$$

The terms in the left set of square brackets can be efficiently computed for each value of the rank as

$$\sum_{Xc} D_c^{aX} \left(\sum_d \bar{D}_d^{rX} \tau_r^{cd} \right) \quad (38)$$

where \bar{D}_d^{rX} and τ_r^{cd} are matrices associated with a specific value of r . In eq 38, the sum over d can be computed efficiently using a standard level-3 BLAS *gemm* kernel and the sum over the compound index $\{Xc\}$ can be computed using a standard level-2 BLAS *gemv* kernel. The cost of both of these operations is N^3 for each value of r , leading to an overall scaling of $N^3 R \approx O(N^4)$. It is obvious that contractions in eqs 37 and 38 could be distributed over blocks of r to improve computational performance. Although, we plan to exploit this parallelism in future work as our current algorithm is implemented in a serial tensor algebra library. Finally one is left with

$$[\alpha_{n+1}]_{r'}^a = \sum_r D_r^a [W^{-1}]_{r'}^r \quad (39)$$

which can be computed with a cost of $NR^2 \approx O(N^3)$. As one can now see, the computational complexity of an ALS update is $O(N^5)$, which is a reduction over the canonical complexity of the PPL tensor-network contraction. Readers can please refer to Table 1 for a consolidated representation of the algorithmic cost. In Section 4 we will show that an approximately linear-scaling and modest value of R is necessary to compute chemically accurate energy differences.

We do concede that CPD optimization does have a large computational prefactor due in part to its hyperedge structure

Table 1. Cost and Scaling of Each Contraction in the CP-PPL Optimization

contraction	cost	scaling
$\sum_{ij} \rho_i^a \rho_j^b \tau_{ij}^{cd}$	O^2V^2R	N^5
$\sum_b \bar{D}_d^{bx} \rho_r^b$	V^2XR	N^4
$\sum_d \bar{D}_d^{Xc} \tau_d^{rc}$	V^2XR	N^4
$\sum_{xc} D_c^{ax} \tau_{xc}^r$	V^2XR	N^4
$\sum_b \beta_r^b \beta_{r'}^b$	VR^2	N^3
W^{-1}	R^3	N^3
$\sum_r D_r^a [W^{-1}]_{r'}$	VR^2	N^3

and in part to the necessary repeated evaluations of the gradient. In this work, we do our best to minimize the prefactor by caching intermediate tensors, implementing efficient kernels and developing an effective initialization, discussed in Section 2.4.2. However, ultimately this work is a starting point to a more advanced study and application of matrix-free optimization to electronic structure methods. In future work, we plan to integrate localization and randomized linear algebra techniques to reduce computational prefactors and introduce optimization acceleration techniques such as the DIIS line-search method.⁸⁰ As this is a preliminary study into the application of matrix-free decompositions to replace tensor-network contractions, we focus less on the reduced-scaling impacts of this work and more on the effects our method has on the CCSD residual equation and subsequent optimization accuracy. In this next section we develop a new and reduced-scaling metric to terminate the matrix-free CP-PPL ALS minimization.

2.4.1. Low Cost CPD Stopping Condition. The canonical stopping condition for the CPD is based on the change in the fit of approximation. The fit of the approximated PPL tensor is

$$\Delta_n = 1.0 - \frac{\|B - \tilde{B}_n\|}{\|B\|} \quad (40)$$

and, therefore, the canonical stopping condition is

$$\Delta_n - \Delta_{n-1} < \varepsilon \quad (41)$$

where ε is the chosen threshold value. For the CP-PPL, computing the fit is prohibitively expensive (with a complexity of $O(N^6)$). We will break down the cost of the canonical stopping condition.

The numerator of eq 40, also defined as the residual, can be expanded as

$$\|B - \tilde{B}_n\| = \sqrt{B^2 - 2\tilde{B}_n B \tilde{B}_n + \tilde{B}_n^2} \quad (42)$$

The second and third terms under the square root can be computed with relatively low-cost; the term $\tilde{B}_n B \tilde{B}_n$ is partially formed during the computation of the CPD gradient and formally has a complexity of $O(N^5)$, as demonstrated previously.

While, the term \tilde{B}_n^2 can be computed by leveraging the CPD format

$$\tilde{B}_n \tilde{B}_n = \sum_{rr'} [((\alpha_n) T \alpha_n)^* ((\beta_n) T \beta_n)^* ((\gamma_n) T \gamma_n)^* ((\rho_n) T \rho_n)^r] \quad (43)$$

and has a complexity of $O(N^3)$. Unfortunately, any attempt to construct B or B^2 results in a $O(N^6)$ computational complexity.

In principle, it would be possible to forego a check completely and simply require a certain number of ALS iterations be completed. However, it would be preferred to have some metric to effectively stop the optimization in as few steps as possible. Therefore, in this work we introduce the new metric

$$\eta = \frac{\|\tilde{B}_n - \tilde{B}_{n-1}\|}{\|\tilde{B}_{n-1}\|} \quad (44)$$

The numerator of η can be expanded as

$$\eta_{\text{num}} = \sqrt{\tilde{B}_n^2 - 2\tilde{B}_n \tilde{B}_{n-1} + \tilde{B}_{n-1}^2} \quad (45)$$

and all terms in this residual can be efficiently computed in $O(N^3)$ time using the same mechanism found in eq 43. The CPD is then terminated when $\eta < \varepsilon$ for some chosen ε and an appropriate value of ε will be discussed in Section 4. As with most optimization problems, the quality and duration of the CPD optimization can strongly depend on the initial guess. Therefore, in the following section we will propose an improved initial guess strategy to minimize the required number of ALS iterations and therefore the cost of computing the CPD.

2.4.2. An Improved Initial Guess Strategy. In general, nonlinear optimizations require an unknown (potentially large) number of iterations and the number of iterations can depend strongly on the quality of an initial guess. In this application we require the CPD optimization be computed every iteration of the CCSD optimization. It is, therefore, in our best interest to find a high-quality initial guess strategy that minimizes the number of CPD optimization iterations.

In our previous work,² we have found that a reasonable starting guess is a collection of random numbers chosen from a uniform distribution between $[-1, 1]$. In this work, too, we see that this guess puts the CPD solver in a good position to converge to a relatively accurate optimization in as few as 13 iterations. However, consider the following, if we choose a random guess for each CP-PPL optimization during a CCSD optimization, we are effectively treating each PPL tensor-network as an independent CPD problem. We hypothesize that, for a single CCSD calculation, the optimized CP-PPL factor matrices during the n th CCSD iteration are close in optimization space to the CP-PPL factor matrices of the $m > n$ CCSD iteration. In this way, we are positing that there is some relationship between optimized CPD approximations of a given tensor-network, which are components of an iterative gradient optimization method, such as CCSD. Therefore, our proposed strategy is to find and cache a set of optimized CP-PPL factor matrices during the n th CCSD iteration to use as an initial guess for a later CP-PPL optimization. A more complete description of our initial guess procedure will be presented and analyzed in Section 4. Next, we attempt to put our approach in perspective of previous studies that apply low-rank tensor decompositions to the CC family of methods.

2.5. Literature Perspective. In electronic structure methods, there exists only a small number of studies that utilize analytically constructed CPD tensors^{72,73,81–85} and thus, the

utility of the analytical CPD in the field is still an open question. In their previous work, Auer et al. replaced practically every high-order tensor with a CP approximation. Although these authors formally achieve a reduced-scaling implementation, Auer concedes that these CP-based approaches are slowed significantly by the spontaneous growth of the CP rank which results in the need for periodic reoptimization of CP approximated tensors. Recently, Madsen et al.⁸⁵ use a similar idea in the vibrational coupled cluster (VCC) method. While their work, again, requires the periodic reoptimization of CP factor matrices, these optimizations can be computed relatively efficiently and the authors also take advantage of the Tucker decomposition to further reduce the cost of their CPD optimization.⁷² Overall Madsen finds a significant benefit with the introduction of the CPD to the VCC method. In this work, we explore a new application of the analytic CPD to directly approximate high-scaling tensor-networks without approximating components of the networks. In a sense, this manuscript serves to extend Auer and Madsen's work by preserving the structure of a tensor-network using a fast-update scheme that simultaneously contracts the tensor-network and optimizes a low-rank decomposition. It should be noted that in this preliminary study, we do not preserve the decomposed tensor-network format and choose instead to reconstruct the result of our decomposition.

The core idea of this manuscript parallels the idea presented by Schutski et al.⁶³ In their work, the authors propose a means to optimize a THC representation of the two-particle excitation tensor while simultaneously minimizing the CCSD equations, eq 1. In a similar manner to the proposed method in this work, the authors first create the loss function

$$f(\hat{T}_2) = \frac{1}{2} \|T_2 - \tilde{T}_2\|^2 \quad (46)$$

where \tilde{T}_2 is a THC representation of the two-particle excitation tensor. Schutski then replaces the exact T_2 tensor with a tensor-network representation obtained from the canonical update of the tensor, i.e.

$$t_{ij}^{ab} = \frac{\langle \Phi_{ij}^{ab} | \bar{H} | \Phi_0 \rangle}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j} \quad (47)$$

By replacing the exact T_2 tensor with this tensor-network approximation the authors develop a matrix-free optimization of \hat{T}_2 and, like we demonstrate in this work, are able to reduce the complexity of tensor-network contractions in the residual equation. While this matrix-free optimization is functionally equivalent, Schutski's optimization process is made complicated by the sheer number of tensor-networks which must be evaluated to form the least-squares solutions to eq 46 and the fact that the CCSD residual tensor, $R_{ij}^{ab} = \langle \Phi_{ij}^{ab} | \bar{H} | \Phi_0 \rangle$, is itself a function of T_2 . In that work, the authors choose to replace the T_2 in each iteration with the optimized approximation \hat{T}_2 from the previous CCSD iteration. In this work, we effectively reduce the scope of Schutski's work to understand the role of error propagation in approximated tensor-networks and to demonstrate that individual components of the CCSD residual equation are low-rank tensors. In principle, we could reformulate the optimization of T_2 presented by Schutski and replace the THC approximation with a CPD; however, practically, this would also lead to an equally large and

complicated set of tensor-network intermediates. Alternatively, we are currently investigating efficient means to combine these studies to accurately accelerate the CC family of methods.

The work by Schutski motivated the development of the rank-reduced coupled cluster method.^{64,67–69,86} This approximation also represents the excitation amplitude tensors in a low-rank THC decomposition. However, unlike Schutski's tensor-structured decomposition, the rank-reduced approximation only determines updates to the THC core tensor. By doing so, this rank-reduced approximation effectively reduces the number of parameters in the gradient optimization of CCSD from N^4 to N^2 and in coupled cluster with single, double and triple excitations (CCSDT) from N^6 to N^3 using a projection-based approach. This idea creates relatively compact CC algorithms compared to Schutski's tensor-structured implementation. Next, we describe the computational experiments used to validate this work.

3. COMPUTATIONAL DETAILS

All CCSD calculations have been computed using a developmental version of the massively parallel quantum chemistry (MPQC) software package^{87,88} and computations were run on the Flatiron Institute Scientific Computing Core's (SCC) Rusty cluster using the standard Rome nodes that are equipped with 2 AMD EPYC 7742 processors that each have 64 CPU cores and 512 GB of memory. Each CC calculation was run on a 16 core partition of a single Rome CPU with each partition containing 128 GB of memory. The CP approximation of the order-4 PPL tensor-network was computed using a standard alternating least-squares (ALS) method,^{78,79} as discussed previously. Using the initialization schemes discussed in Section 2.4.2, we see rapid convergence to high accuracy approximation and the full extent of these optimizations will be discussed in Section 4. All CPD optimizations that utilize the canonical CP stopping condition, eq 40, use a convergence criteria of $\epsilon = 10^{-3}$, which was found to be sufficiently accurate in our previous published results.^{2,21} The focus of our study is small water clusters with between 1 to 6 molecules in the TIP4P optimized geometry^{89,90} from the Web site <http://www-wales.ch.cam.ac.uk/~wales/CCD/TIP4P-water.html> and small alkane chains with between 1 and 6 carbon atoms. We analyze the different initial-guess and convergence strategies and wall-time performance of the CPD using the water clusters in the cc-pVDZ-F12 (DZ-F12) orbital basis set (OBS), accompanied by the corresponding aug-cc-pVDZ-RI (aVDZ-RI) DF basis set (DFBS).^{47,91} We choose this basis for its intermediate size between the double and triple- ζ basis. We also analyze the approach using the cc-pVTZ (TZ)^{92,93} OBS accompanied by the corresponding cc-pVTZ-RI (TZ-RI)⁹⁴ DFBS to determine the scaling of the CP rank with system size. In the following section, we present our findings from these experiments.

4. RESULTS

In this section, we demonstrate that it is possible to accurately replace the exact computation of the CCSD PPL tensor-network with a low-rank, matrix-free tensor factorization. Further, we establish an understanding of the following CP-PPL parameters: initial guess strategy, CP rank and CPD stopping condition, and we consider the computational limitations of this approach.

4.1. Role of the Initial Guess. First, we investigate the role of the CP-PPL initial guess and study its effects on the CCSD

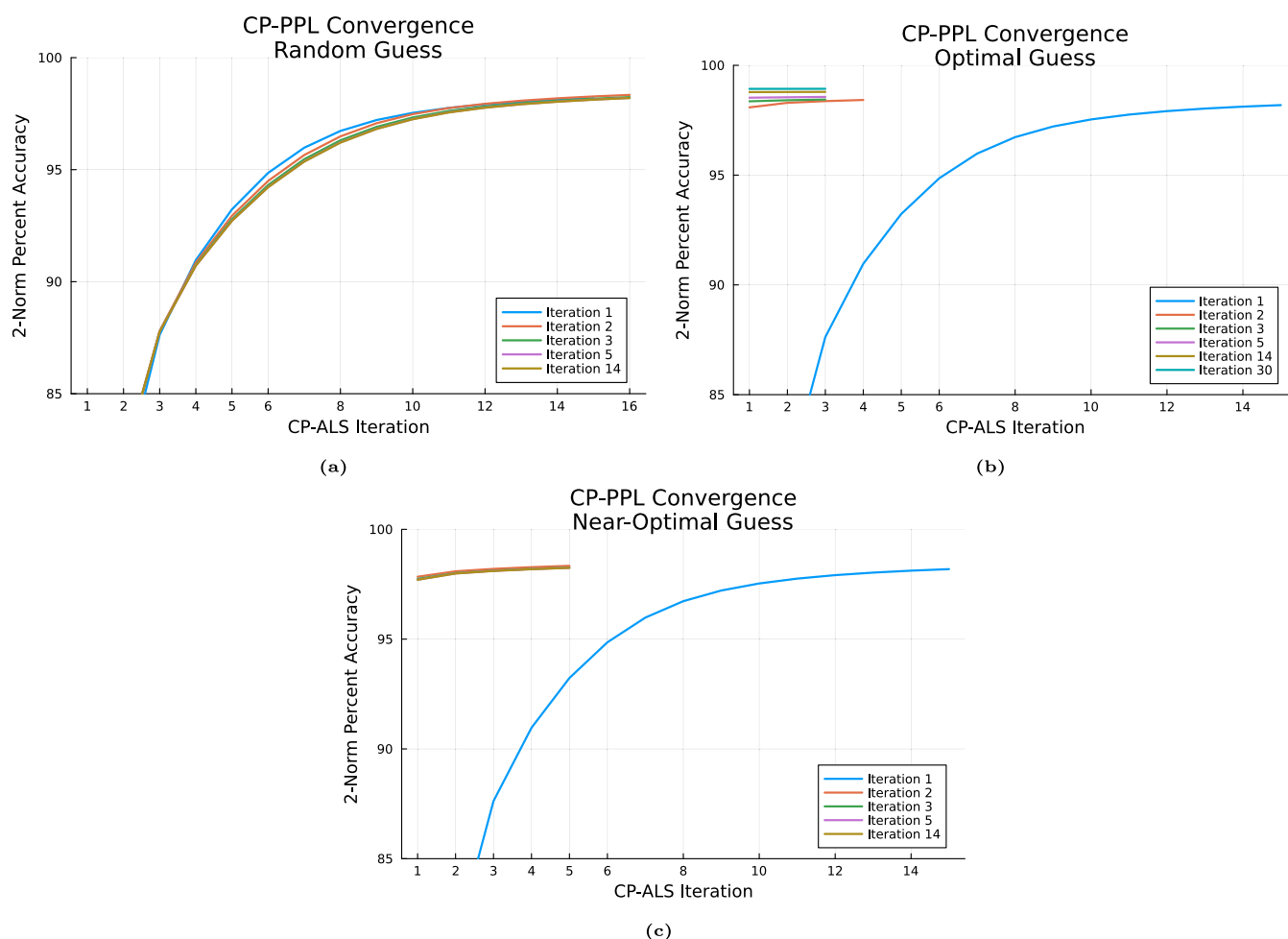


Figure 6. Convergence of the CP-PPL ALS optimization using the (a) “Random guess”, (b) “Optimal guess” and (c) “Near-Optimal guess” for a 2 water molecule cluster in DZ-F12/aVDZ-RI with a fixed CP rank of 3.5X during different CCSD iterations.

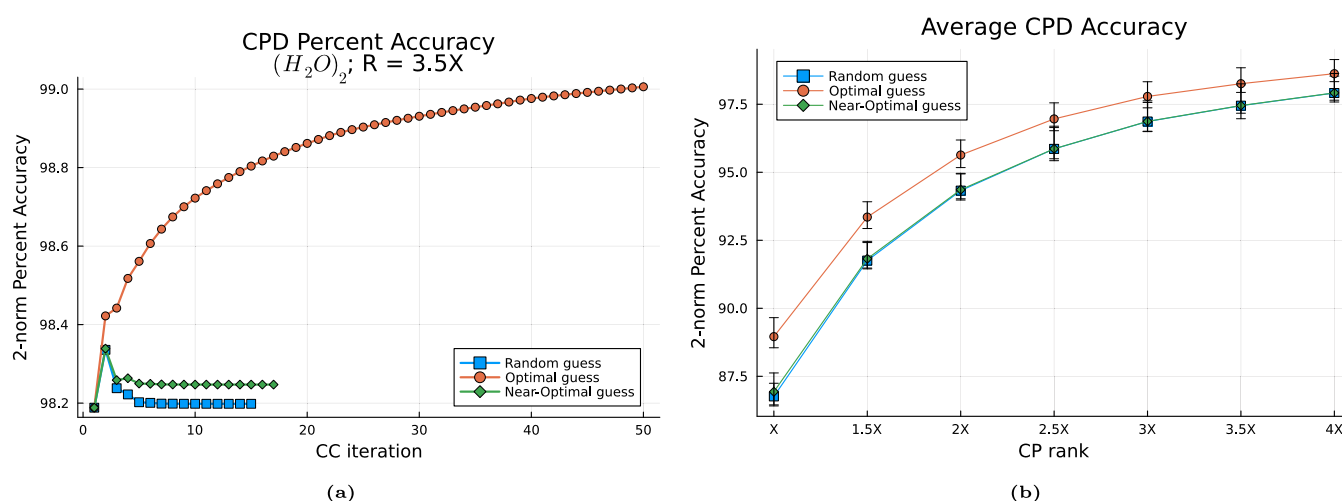


Figure 7. (a) CP-PPL percent accuracy across a CCSD optimization for a two water molecule cluster in DZ-F12/aVDZ-RI at fixed CP rank of 3.5X using the three CPD initialization schemes. (b) CP-PPL percent accuracy for clusters with between 1 and 5 water molecules in DZ-F12/aVDZ-RI using three CPD initialization schemes. Error bars represent maximum and minimum percent accuracy across water clusters.

optimization. In this work we use three different initial guess strategies.

- with “Random guess” we initialize each CPD with a set of randomly initialized factor matrices;
- with “Optimal guess” we initialize every CPD with the output of the previous CPD optimization; and
- with “Near-Optimal guess” we initialize the current CPD with a cached set of factor matrices from a previous optimization. The caching procedure is as follows: if the

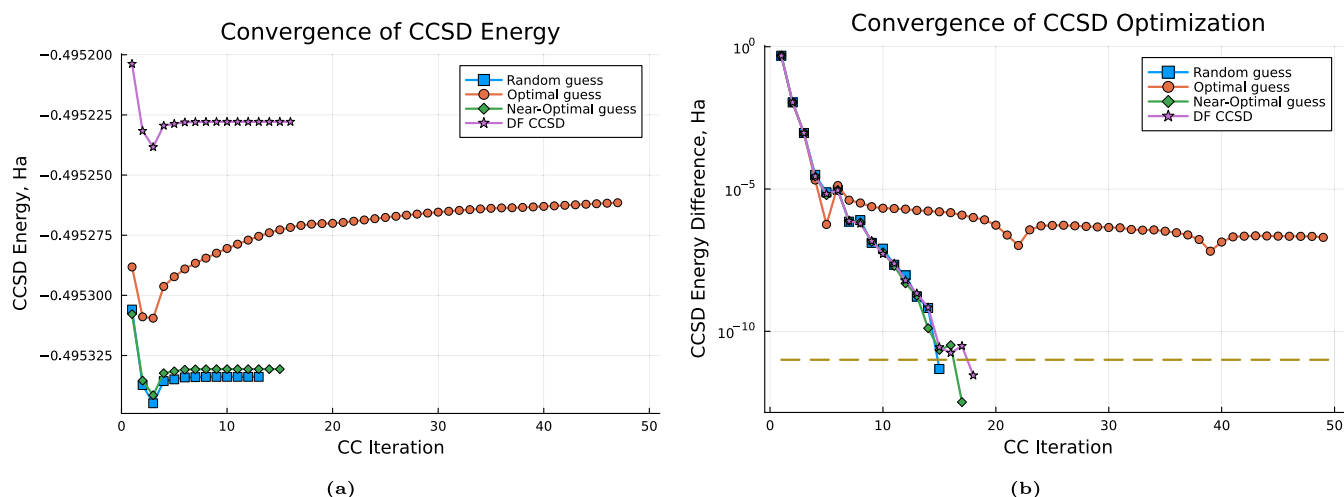


Figure 8. Comparing the CCSD energy convergence using (a) absolute energy and (b) per iteration energy differences for a two water molecule cluster in DZ-F12/aVDZ-RI and CP-PPL calculations using a fixed CP rank of 3.5X. The dashed line in (b) represents the CCSD energy convergence criteria.

current CPD optimization requires $n > 5$ iterations, then the factor matrices from the $(n - 5)$ th iteration are cached. Else, the current optimized factor matrices are not cached for later use and a previously cached set is used in the subsequent optimization.

For all of these schemes, the first CPD optimization of a network always uses the “Random guess.” For the “Near-Optimal guess” scheme the value 5 was chosen to be close enough to the end of the optimization for future optimizations to converge quickly and far enough away to consistently optimize the CP-PPL to nearly the same 2-norm accuracy, as one can see in Figures 6c and 7a.

In Figure 6, we illustrate the convergence of the CP-PPL ALS optimization at different iterations of CCSD using the three different initial guess strategies. As mentioned previously, we define convergence as when the change in fit, eq 41, is less than 10^{-3} . In Figure 6 we plot the 2-norm percent accuracy (which is computed as 100 times the CP fit, eq 40) for each CP-ALS iteration in lieu of the change in CP fit. Therefore, in this figure, convergence correlates with the slope between consecutive points. As discussed, all three methods use the “Random guess” for the first iteration and these optimizations converge consistently in about 13 iterations. With Figure 6a, one can see that this “Random guess” consistently converges in roughly 13 iterations for each iteration of the CCSD algorithm and the accuracy of the CP-PPL is consistent across iterations. With Figure 6c, one can see that the “Near-Optimal guess” significantly reduces the number of ALS iterations from 13, using the “Random guess,” to 5. And, like the “Random guess,” the accuracy of the CP-PPL using the “Near-Optimal guess” is consistent every iteration of CCSD, unlike the “Optimal guess” strategy.

With Figure 6b one can see that the “Optimal guess” significantly reduces the number of ALS iterations to 4 or fewer. Interestingly, with the “Optimal guess” the accuracy of the CP-PPL improves every CC iteration. A possible explanation for this phenomenon is that from iteration 3 and beyond, the PPL tensor does not dramatically change. Therefore, the “Optimal guess” strategy continues to improve the CP-PPL during each CCSD iteration, converging the approximation far beyond the prescribed ALS convergence criteria. If this is the case, then this result demonstrates that there is room for improvement in

the fixed-rank CP-PPL optimization. One route to improve the convergence of the ALS solver is using line-search methods such as the Direct Inversion of the Iterative Subspace method;⁸⁰ alternatively we could employ more efficient optimization strategies like the accelerated gradient descent technique.^{81,95–98}

We illustrate the 2-norm percent accuracy of the CP-PPL across a CCSD optimization of a 2 water molecule cluster using the three initial guess strategies with a fixed CP rank of 3.5X in Figure 7a. The results in Figure 7a corroborate the results from Figure 6 in that the 2-norm accuracy of the CP-PPL is consistent across a CCSD optimization using the “Random guess” and “Near-Optimal guess” strategies but improves consistently using the “Optimal guess” strategy. In Figure 7b, we plot the averaged 2-norm based percent accuracy of the CP-PPL across water clusters with between 1 and 5 molecules and averaged across all CCSD iterations. To compile the data in Figure 7b we first average the CPD accuracy over the number of CCSD iterations for each molecule. Then we average the averaged CPD accuracies over the set of molecules tested. This figure demonstrates that, as expected, increasing the CP rank directly improves the accuracy of the approximation of B . Similar to the previous results, Figure 7b shows that the “Random” and “Near-Optimal” guesses produce roughly the same CP-PPL approximation conversely, the “Optimal guess” initial guess scheme produces a better CPD optimization on average than the other two initial guess schemes. Also in Figure 7b, we see that the averaged CP-PPL accuracy is rather consistent with molecular system size.

One should notice that in Figure 7a the “Optimal guess” strategy is computed for 50 CC iterations, whereas the other strategies require fewer CC iterations. We investigate this further in Figure 8. In our calculations, CCSD convergence is determined when the change in CCSD energy falls below 10^{-11} Ha. In this figure, one can see that the “Random guess” and “Near-Optimal guess” strategies allow CCSD to converge quickly to this threshold, requiring roughly the same number of CCSD iterations as the canonical calculation, here denoted “DF CCSD.” One can see that with the “Optimal guess” scheme, because the CP-PPL improves consistently throughout the course of the CCSD optimization, the CCSD energy also improves consistently throughout the optimization. This behavior is not exactly desirable as it causes the CC optimization

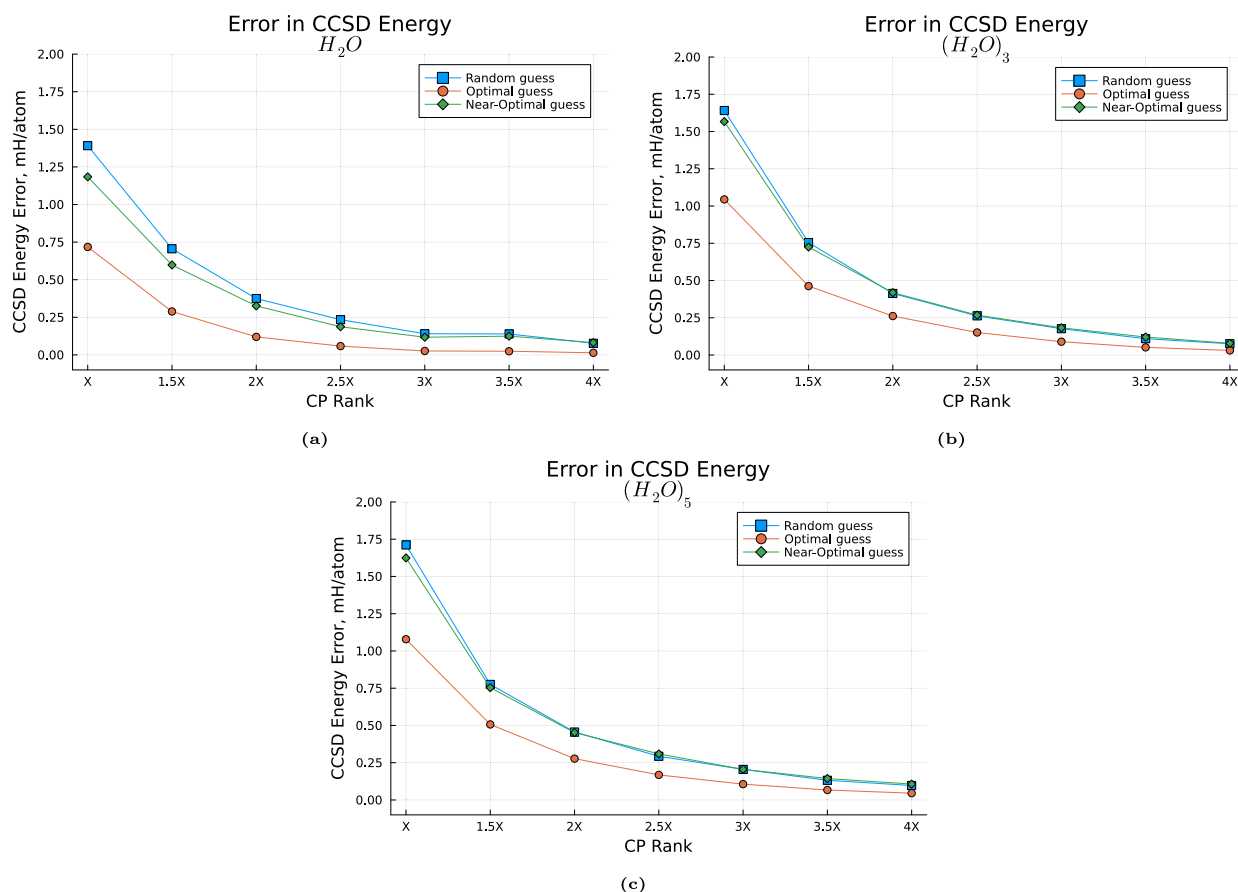


Figure 9. Absolute CCSD energy error per non-hydrogenic atom for clusters of (a) 1, (b) 3 and (c) 5 water molecules using the DZ-F12/aVDZ-RI basis versus CP rank using the various initial guess schemes. The “Random guess” and “Near-Optimal guess” schemes were converged to a CCSD energy difference of 10^{-11} Ha, however all “Optimal guess” computations did not converge in 50 CCSD iterations.

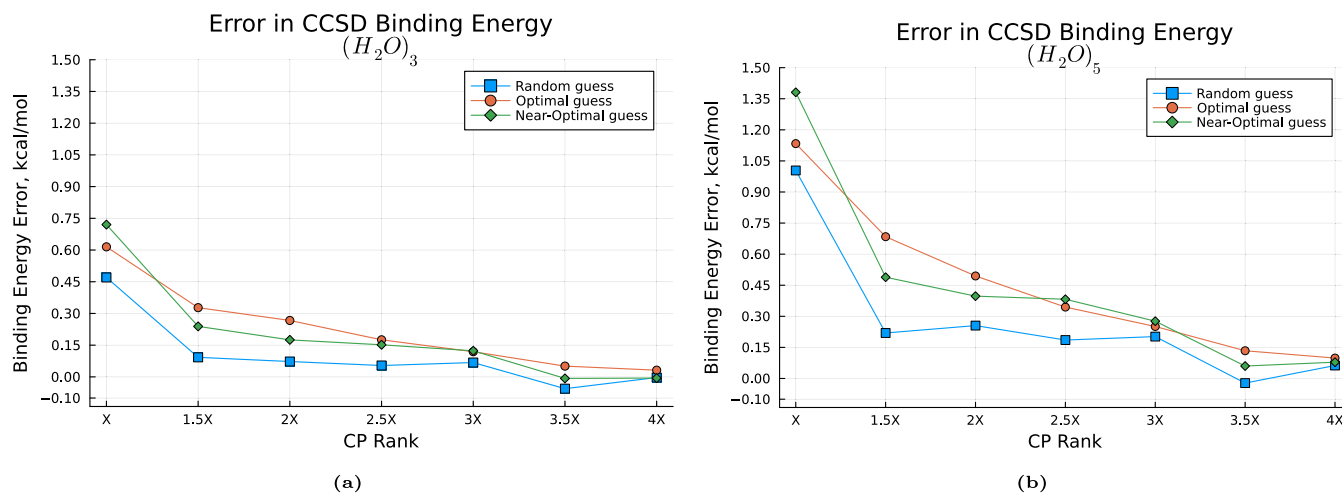


Figure 10. Signed binding energy error in kcal/mol versus CP rank for clusters of (a) 3 and (b) 5 water molecules in the DZ-F12/aVDZ-RI basis.

to “swamp” and never reach the target energy convergence of 10^{-11} Ha. In no single CP-PPL based CCSD optimization computed with the “Optimal guess” scheme does the CCSD loop converge to the requested energy difference cutoff in 50 iterations. Perhaps this suggests an issue with the robustness of the standard Jacobi-DIIS CCSD solver and that such a solver only performs well with consistently approximated terms in its gradient expression. Alternatively, this could be fixed using a tighter CP-PPL ALS convergence criteria or by stopping the CP-

PPL optimization after a number of CCSD iterations. In future work, we hope to garner a better understanding of the interplay between the CPD optimization/approximation of the PPL ladder diagram and the CC optimization landscape to resolve this swamping issue. Next, we analyze the accuracy of the CP-PPL approximated CCSD method using these three initialization strategies.

4.2. Analyzing the Accuracy of the CPD-PPL. Here we show that sufficiently accurate and systematically improvable

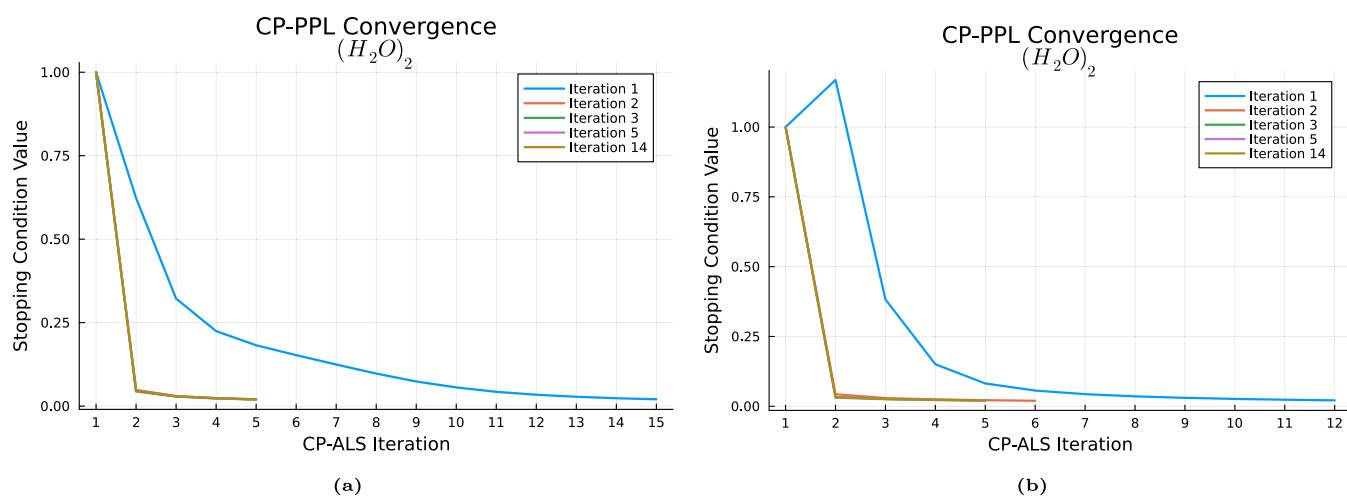


Figure 11. CP-PPL convergence via the evaluation of the reduced-scaling N^3 stopping condition for clusters of (a) 2 and (b) 5 water molecules in the DZ-F12/aVDZ-RI basis.

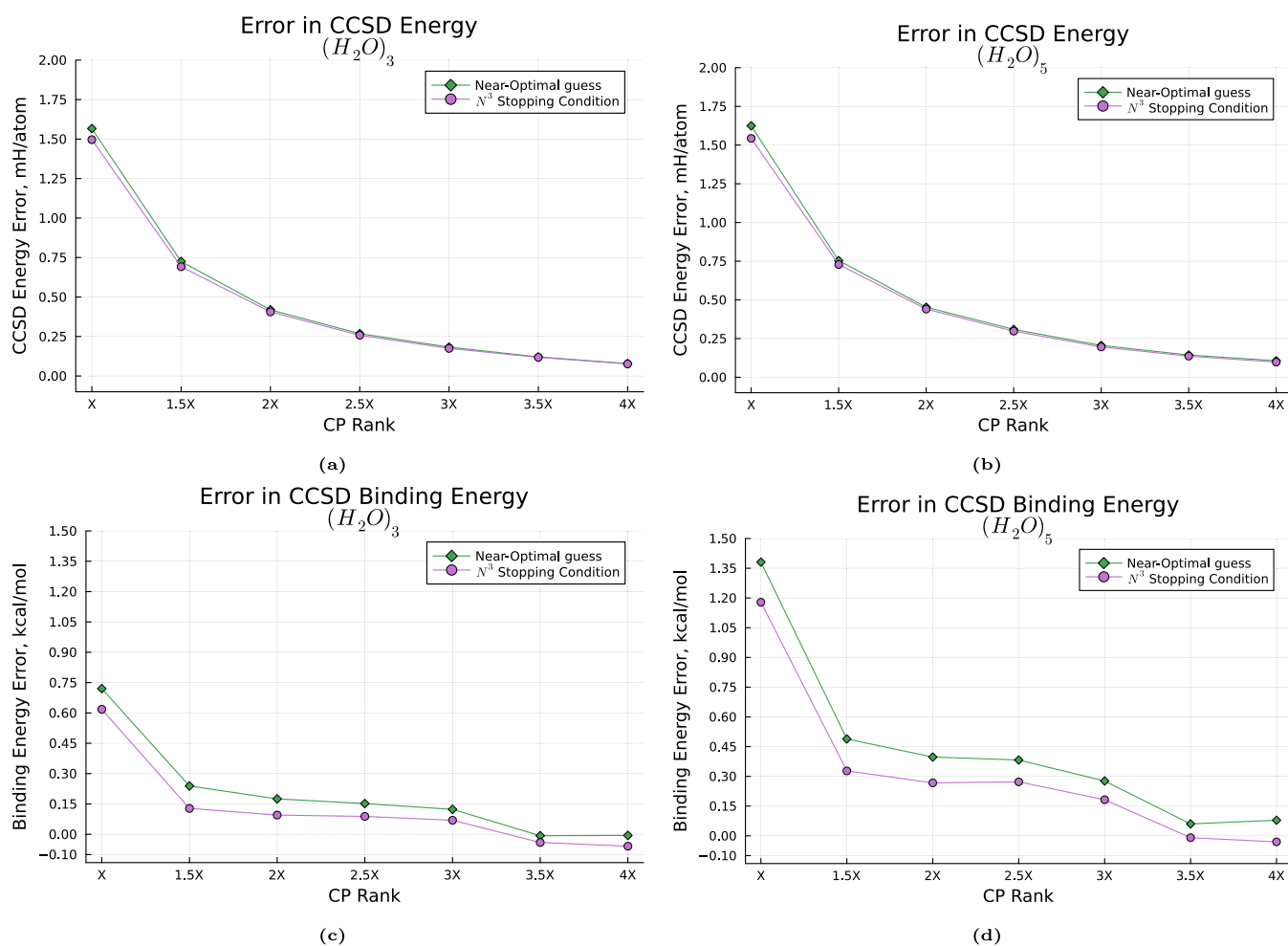


Figure 12. Comparing the absolute CCSD energy error for (a) 3 and (b) 5 water molecule cluster using the canonical and reduced-scaling stopping CP-PPL condition. Comparing the binding energy error for (c) 3 and (d) 5 water molecule cluster using the canonical and reduced-scaling CP-PPL stopping condition.

energy values can be computed by replacing the PPL tensor-network contraction with the CP-PPL approximation. First, Figure 9 shows the error in the CP-PPL approximated CCSD energy compared to the canonical DF approximated CCSD energy for various water molecule clusters in the DZ-F12/

aVDZ-RI basis using the three different initial guess strategies described in the previous section. As the CP-PPL approximation improves, all initial guess strategies find increasingly accurate energy values. The “Random guess” and “Near-Optimal guess” approaches both find nearly identical energy values while the

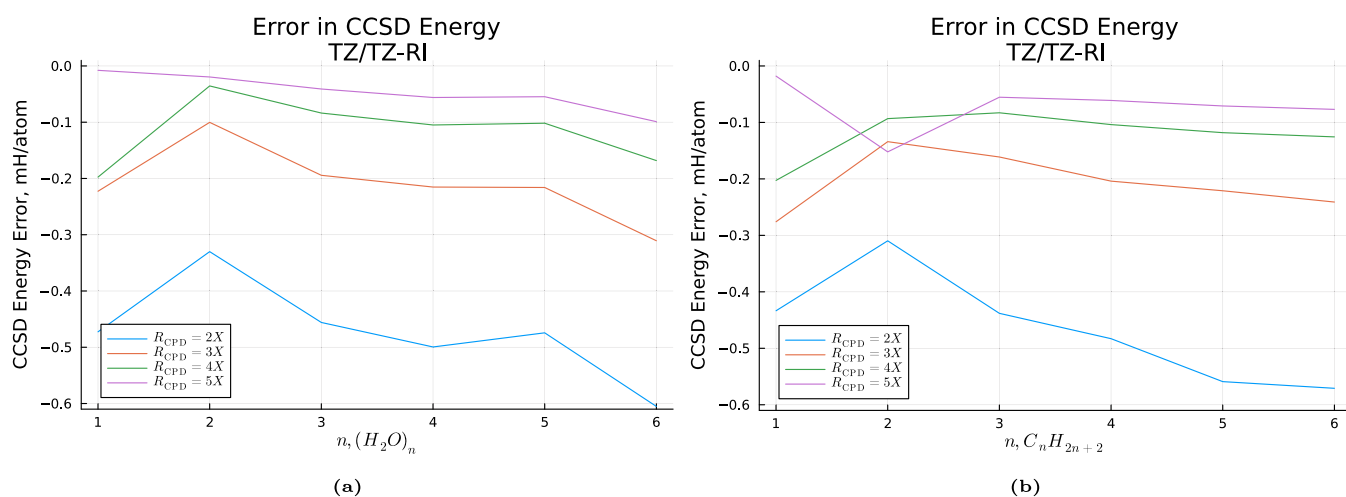


Figure 13. Absolute error in CP-PPL CCSD energy per non-hydrogenic atom for (a) water clusters in the TIP4P geometry using the TZ/TZ-RI basis sets and (b) for alkane chains in the TZ/TZ-RI basis.

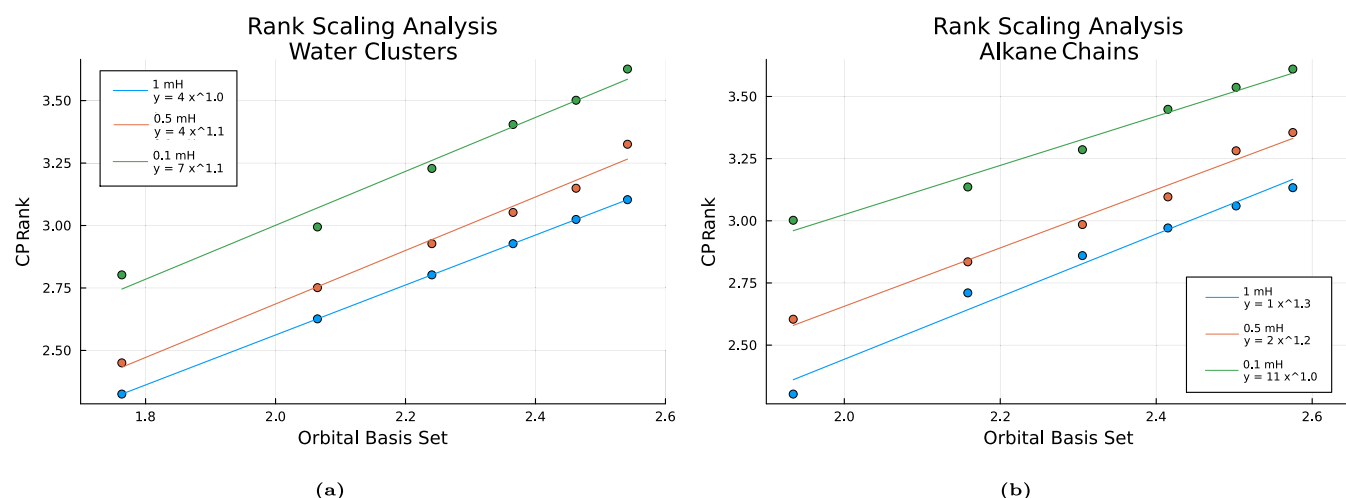


Figure 14. Modeling the growth of the CP rank with system size for (a) water molecule clusters and (b) alkane chains in the TZ/TZ-RI basis using three different error thresholds.

“Optimal guess” scheme finds significantly more accurate CCSD energy values for every molecule and at every rank. Also, there does not seem to be a dramatic difference in CCSD energy error for these different water molecule clusters.

Figure 10 shows that, although the “Optimal guess” finds more accurate CCSD energies, this does not categorically translate to more accurate energy differences. In fact, in this example, the “Optimal guess” scheme always finds higher errors in the binding energy of water when compared to the “Random guess” scheme. Most likely, the lower accuracy binding energies are related to the fact that the “Optimal guess” does not converge the CCSD energies beyond 10^{-6} Ha. However, the other approaches may more systematically approximate the CCSD energy, resulting in a fortuitous cancellation of error. From Figure 10, one can see that only a modest CP rank of 1.5X is required to compute the binding energy with an accuracy beyond chemical accuracy, 1 kcal/mol. But, we recommend a rank of 3.5X or greater to find an accuracy beyond 0.1 kcal/mol. In the Supporting Information, readers can find additional results comparing the accuracy of the CP-PPL to the component-wise approximation of the PPL tensor-network via

the order-4 CPD of the TEI tensor. In the next section, we introduce our reduced-scaling convergence CPD criteria.

4.3. Analyzing the Reduced-Scaling ALS Convergence Criteria. Here we introduce our reduced-scaling CPD convergence criteria. Thus, far, our results have used the canonical stopping condition, eq 40. As discussed previously, this stopping condition depends on the norm of the PPL tensor, which has an irreducible $O(N^6)$ scaling. In the figures to follow, we use the label “ N^3 Stopping Condition” to denote the results using this new, reduced-scaling stopping condition from eq 44. It should be noted that all calculations that use the reduced-scaling stopping condition also use the “Near-Optimal guess” scheme. In general, the choice of stopping condition threshold, ϵ , is an arbitrary decision and was chosen to replicate the balance between CP-PPL accuracy and number of least-squares iterations found using the canonical stopping condition. Here we do not perform a deep evaluation of the new stopping condition and choose the stopping condition of $\epsilon = 0.02418$ based on a simple heuristic: we choose a value of ϵ such that the water dimer molecule with a CP rank of 2X completes its CP-PPL optimization in 16 ALS iterations using the “Random Guess” initialization scheme. With Figure 11, we show the

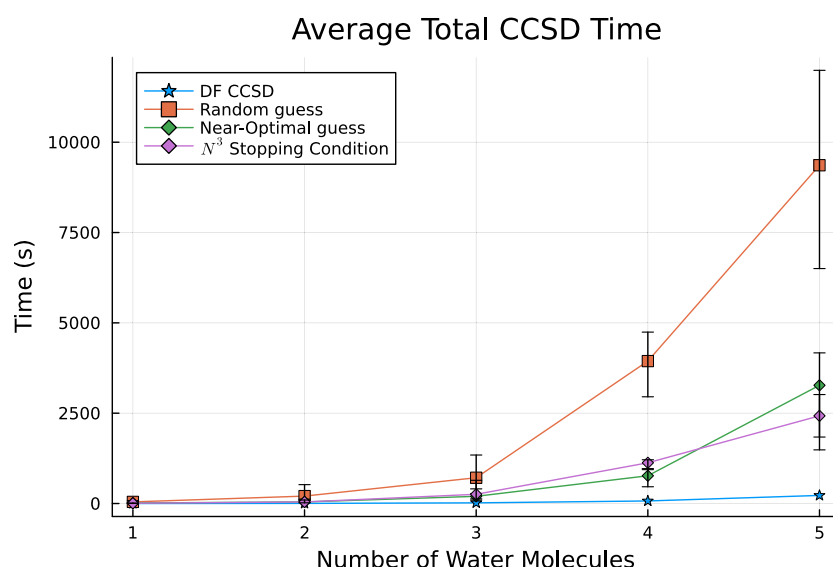


Figure 15. Total time to compute the CCSD energy for water clusters in the DZ-F12/aVDZ-RI basis. CP-PPL approximated CCSD calculations were averaged over all CP rank values and the error bars represent the maximum and minimum recorded times.

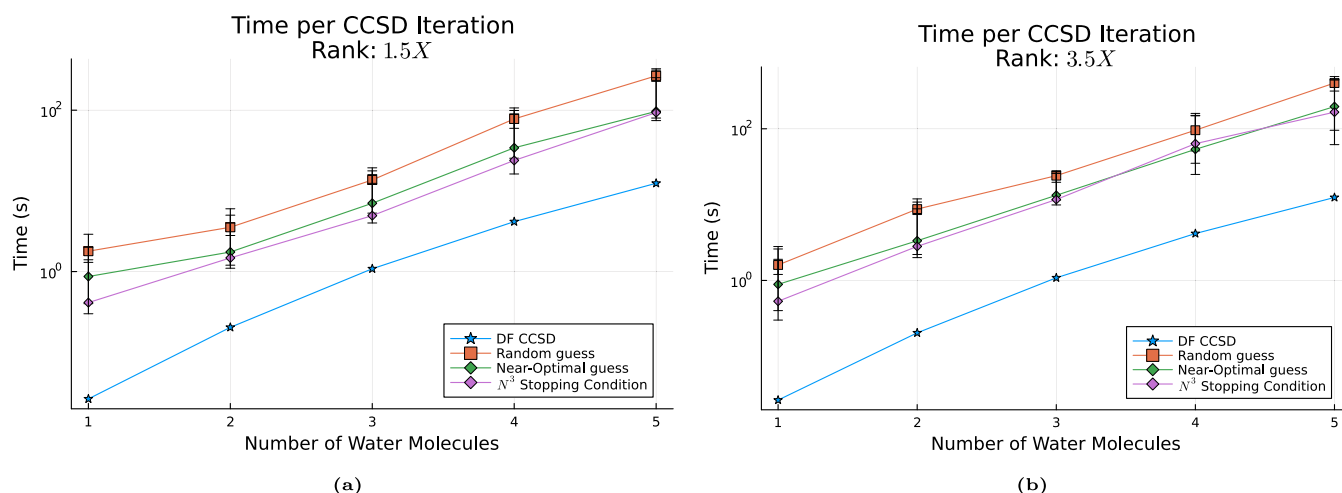


Figure 16. CCSD per iteration cost versus number of water molecules in the DZ-F12/aVDZ-RI basis. Results were compiled using a CP rank of (a) 1.5X and (b) 3.5X. Error bars represent maximum and minimum per iteration time.

convergence behavior of the CP-PPL using the reduced-scaling stopping condition. One can see that the behavior of this convergence is nearly identical to the convergence of the “Near-Optimal guess” with the canonical stopping condition, i.e., Figure 6c. Finally, we analyze the accuracy of the CCSD results using this new stopping condition in Figure 12. This figure illustrates that the reduced-scaling initial guess scheme realizes nearly identical results for absolute and relative energy quantities compared to the canonical stopping condition. These results demonstrate that our choice of ϵ is reasonable and that the new convergence criteria can be effectively used by the CP-PPL optimization scheme. Next we analyze the scaling of the rank of the CP-PPL approximation.

4.4. CP-PPL Rank Scaling Models. Here we analyze the accuracy of the CP-PPL approximated CCSD calculation against systematically increasing system size to determine the scaling of the CP rank for this tensor-network. In Figure 13 we plot the absolute error in the CCSD energy using fixed values of CP rank across a set of water clusters and alkane chains in the TZ/TZ-RI basis. One can see that, for the most part, increasing

the CP rank systematically improves the accuracy of the CP-PPL approximated CCSD method. One exception to this trend is the C_2H_8 molecule with a CP rank of 5X. We can occasionally find this kind of inconsistency when decomposing small problems to large CP rank; it is most likely related to the optimization of an overparameterized alternating least-squares problem.

Furthermore, there appears to be a slight increase in CCSD energy error, with increasing system size. Although, because the tested systems are rather small, it is hard to determine if this trend will continue for larger system sizes. Because this work is a preliminary study, we leave that investigation to future works on this matter. Using the data from Figure 13, we plot the CP rank versus system size for various fixed error thresholds in Figure 14. The models computed from the data in Figure 14 illustrate that the rank of the CP-PPL approximation is nearly linear, falling between N and $N^{1.3}$. This scaling is consistent with other works studying low-rank decompositions of tensors in CCSD such as Benedikt et al.⁸² and Schutksi et al.⁶³

4.5. CP-PPL Approximated CCSD Performance. Finally, because the CP-PPL is technically a reduced scaling approach,

we show the wall-time cost of the CP-PPL approximated CCSD compared to a canonical DF CCSD optimization in Figures 15 and 16. As discussed earlier, the point of this work is not to realize a wall-time improvement over the canonical contraction of the PPL diagram although, for completeness, we discuss the theoretical crossover. This matrix-free CPD optimization has a leading cost of (V^2O^2R) per CPD ALS update and an additional (V^2O^2R) cost to reconstruct the PPL tensor after the CPD optimization. The canonical PPL has a computational scaling of (V^4O^2). Theoretically there should be a crossover in cost when $V^2 > R(N_{\text{ALS}} + 1)$ where N_{ALS} is the number of ALS iterations. In this work, we have shown that a CP rank of 4X is sufficiently accurate and an acceptable value of X is on the order of 2V. Therefore, this approach should find advantage over the canonical PPL when $V > 8(N_{\text{ALS}} + 1)$. The number of ALS iterations depends on the ALS stopping condition and initial guess but is typically between 8 and 30, which would put the theoretical crossover point between [65, 241] unoccupied orbitals. However, because the CPD has a relatively high prefactor, as discussed in Section 2.3, it is difficult to realize this crossover. In this direction, we seek to improve the timing of the code by distributing the optimization algorithm and utilizing randomized/interpolation based methods to further reduce the prefactor of the CPD of tensor-networks. Also, we are investigating the application of this decomposition-based tensor-network contraction technique to networks that can be computed more efficiently.

5. CONCLUSIONS

In this study, we demonstrate that it is possible to approximately contract tensor-networks through the optimization of a low-rank tensor decomposition and, by doing so, make the error in the contraction a controllable parameter of the decomposition. This idea is not limited to a specific tensor-network or decomposition topology. Here, as a case study, we investigate the approximation of the DF-approximated PPL tensor-network in the CCSD method using the CPD. We find that computing a CPD approximation of the PPL tensor-network can be done accurately with a relatively small CP rank that grows nearly linearly with chemical system size. These results complement our previous results from the investigation into the order-4 CPD of the coulomb integral tensor-network.^{21,22} Furthermore, we utilize the iterative nature of CCSD to construct an effective CPD initial guess strategy that minimizes the cost of the CPD optimization. Unfortunately, we find that the approximation of the PPL ladder with this new scheme can have a significant effect on the optimization of the CCSD wave function. To overcome this problem in a standard Jacobi-DIIS CCSD optimization loop, measures must be taken to ensure a consistent CPD approximation. We believe that these results warrant an investigation into CCSD solvers with improved stability.

In future works, we will study methods to improve the approximation of tensor-networks computed within iterative optimization schemes like coupled cluster. For example, we are currently investigating an approach to reduce the computational storage complexity of the CCSD residual tensor by computing its approximation in a low-rank tensor decomposition format. We also seek to replace more complex and higher-order tensor-networks such as the PPL diagram found in the rank-reduced coupled cluster with single, double and triple excitations (CCSDT) method.⁹⁹ Finally, we look to use more elaborate strategies to better and more quickly decompose tensor-networks. These strategies include using sparsity constraints,

utilizing low-cost nonlinear optimization schemes, and using random/interpolation methods to more quickly approximate the gradient of tensor-networks.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.5c00413>.

xyz files for the alkane chains (ZIP)

Additional result figures for CP-PPL convergence and CP rank scaling (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Karl Pierce – Center for Computational Quantum Physics,
Flatiron Institute, New York, New York 10010, United States;
orcid.org/0009-0008-5135-9239; Email: kpierce@flatironinstitute.org

Complete contact information is available at:
<https://pubs.acs.org/doi/10.1021/acs.jctc.5c00413>

Notes

The author declares no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the Flatiron Institute and the Simons Foundation. We also acknowledge the Scientific Computing Core (SCC) at Flatiron Institute (<https://www.simonsfoundation.org/flatiron/scientific-computing-core/>) for providing computational resources and technical support that have contributed to the results reported within this paper. Furthermore, it is important to acknowledge Dr. J Kaye and Dr. A. Dawid at the Flatiron Institute for their valuable discussions and the author thanks Mina Mandic for contributing in the early stages of the project.

■ REFERENCES

- (1) Dunlap, B. I. Robust and Variational Fitting. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2113–2116.
- (2) Pierce, K.; Rishi, V.; Valeev, E. F. Robust Approximation of Tensor Networks: Application to Grid-Free Tensor Factorization of the Coulomb Interaction. *J. Chem. Theory Comput.* **2021**, *17*, 2217–2230.
- (3) Pulay, P. Localizability of Dynamic Electron Correlation. *Chem. Phys. Lett.* **1983**, *100*, 151–154.
- (4) Pulay, P.; Saebo, S.; Meyer, W. An Efficient Reformulation of the Closed-Shell Self-Consistent Electron Pair Theory. *J. Chem. Phys.* **1984**, *81*, 1901–1905.
- (5) Ahlrichs, R.; Keil, F.; Lischka, H.; Kutzelnigg, W.; Staemmler, V. P. N. O. –C. I. PNO–CI (pair natural-orbital configuration interaction) and CEPA–PNO (coupled electron pair approximation with pair natural orbitals) calculations of molecular systems. III. The molecules MgH₂, AlH₃, SiH₄, PH₃ (planar and pyramidal), H₂S, HCl, and the Ar atom. *J. Chem. Phys.* **1975**, *63*, 455–463.
- (6) Neese, F.; Wennmohs, F.; Hansen, A. Efficient and Accurate Local Approximations to Coupled-Electron Pair Approaches: An Attempt to Revive the Pair Natural Orbital Method. *J. Chem. Phys.* **2009**, *130*, 114108.
- (7) Neese, F.; Valeev, E. F. Revisiting the Atomic Natural Orbital Approach for Basis Sets: Robust Systematic Basis Sets for Explicitly Correlated and Conventional Correlated *Ab Initio* Methods? *J. Chem. Theory Comput.* **2011**, *7*, 33–43.
- (8) Rolik, Z.; Kallay, M. A general-order local coupled-cluster method based on the cluster-in-molecule approach. *J. Chem. Phys.* **2011**, *135*, 104111.

- (9) Rolik, Z.; Szegedy, L.; Ladjanszki, I.; Ladoczki, B.; Kallay, M. An efficient linear-scaling CCSD(T) method based on local natural orbitals. *J. Chem. Phys.* **2013**, *139*, 094105.
- (10) 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.
- (11) Kristensen, K.; Ziolkowski, M.; Jansik, B.; Kjaergaard, T.; Jorgensen, P. A locality analysis of the divide–expand–consolidate coupled cluster amplitude equations. *J. Chem. Theory Comput.* **2011**, *7*, 1677–1694.
- (12) Li, S.; Ma, J.; Jiang, Y. Linear scaling local correlation approach for solving the coupled cluster equations of large systems. *J. Comput. Chem.* **2002**, *23*, 237–244.
- (13) Li, S.; Shen, J.; Li, W.; Jiang, Y. An efficient implementation of the “cluster-in-molecule” approach for local electron correlation calculations. *J. Chem. Phys.* **2006**, *125*, 074109.
- (14) Li, W.; Piecuch, P.; Gour, J. R.; Li, S. Local correlation calculations using standard and renormalized coupled-cluster approaches. *J. Chem. Phys.* **2009**, *131*, 114109.
- (15) Guo, Y.; Becker, U.; Neese, F. Comparison and Combination of “Direct” and Fragment Based Local Correlation Methods: Cluster in Molecules and Domain Based Local Pair Natural Orbital Perturbation and Coupled Cluster Theories. *J. Chem. Phys.* **2018**, *148*, 124117.
- (16) Whitten, J. L. Coulombic Potential Energy Integrals and Approximations. *J. Chem. Phys.* **1973**, *58*, 4496–4501.
- (17) Dunlap, B. I.; Connolly, J. W. D.; Sabin, J. R. On First-Row Diatomic Molecules and Local Density Models. *J. Chem. Phys.* **1979**, *71*, 4993.
- (18) Vahtras, O.; Almlöf, J.; Feyereisen, M. W. Integral approximations for LCAO-SCF calculations. *Chem. Phys. Lett.* **1993**, *213*, 514–518.
- (19) Jung, Y.; Sodt, A.; Gill, P. M. W.; Head-Gordon, M. Chemical Theory and Computation Special Feature: Auxiliary Basis Expansions for Large-Scale Electronic Structure Calculations. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6692–6697.
- (20) Mintmire, J.; Dunlap, B. Fitting the Coulomb Potential Variationally in Linear-Combination-of-Atomic-Orbitals Density-Functional Calculations. *Phys. Rev. A* **1982**, *25*, 88–95.
- (21) Pierce, K.; Valeev, E. F. Efficient Construction of Canonical Polyadic Approximations of Tensor Networks. *J. Chem. Theory Comput.* **2023**, *19*, 71–81.
- (22) Pierce, K.; Morales, M. Using Matrix-Free Tensor-Network Optimizations To Construct a Reduced-Scaling and Robust Second-Order Møller-Plesset Theory. *J. Chem. Theory Comput.* **2025**, *21*, 5952–5964.
- (23) Crawford, T. D.; Schaefer, H. F. *Reviews in Computational Chemistry*; John Wiley, 2000; Vol. 14.
- (24) Čížek, J. On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods. *J. Chem. Phys.* **1966**, *45*, 4256–4266.
- (25) Cizek, J. *Adv. Chem. Phys.*; John Wiley and Sons, Inc.: Hoboken, NJ, USA, 2007; pp 35–89.
- (26) Cizek, J.; Paldus, J. Correlation problems in atomic and molecular systems III. Rederivation of the coupled-pair many-electron theory using the traditional quantum chemical methods. *Int. J. Quantum Chem.* **1971**, *5*, 359–379.
- (27) Tajti, A.; Szalay, P. G.; Császár, A. G.; Kállay, M.; Gauss, J.; Valeev, E. F.; Flowers, B. A.; Vázquez, J.; Stanton, J. F. HEAT: High Accuracy Extrapolated Ab Initio Thermochemistry. *J. Chem. Phys.* **2004**, *121*, 11599–11613.
- (28) Harding, M. E.; Vázquez, J.; Ruscic, B.; Wilson, A. K.; Gauss, J.; Stanton, J. F. High-Accuracy Extrapolated Ab Initio Thermochemistry. III. Additional Improvements and Overview. *J. Chem. Phys.* **2008**, *128*, 114111.
- (29) Thorpe, J. H.; Lopez, C. A.; Nguyen, T. L.; Baraban, J. H.; Bross, D. H.; Ruscic, B.; Stanton, J. F. High-accuracy extrapolated ab initio thermochemistry. IV. A modified recipe for computational efficiency. *J. Chem. Phys.* **2019**, *150*, 224102.
- (30) Purvis, G. D.; Bartlett, R. J. A Full Coupled-Cluster Singles and Doubles Model: The Inclusion of Disconnected Triples. *J. Chem. Phys.* **1982**, *76*, 1910.
- (31) Hummel, F.; Tsatsoulis, T.; Grüneis, A. Low Rank Factorization of the Coulomb Integrals for Periodic Coupled Cluster Theory. *J. Chem. Phys.* **2017**, *146*, 124105.
- (32) Parrish, R. M.; Sherrill, C. D.; Hohenstein, E. G.; Kokkila, S. I. L.; Martínez, T. J. Communication: Acceleration of Coupled Cluster Singles and Doubles via Orbital-Weighted Least-Squares Tensor Hypercontraction. *J. Chem. Phys.* **2014**, *140*, 181102.
- (33) Datar, A.; Matthews, D. A. Robust Tensor Hypercontraction of the Particle–Particle Ladder Term in Equation-of-Motion Coupled Cluster Theory. *J. Chem. Theory Comput.* **2024**, *20*, 708–720.
- (34) Martínez, T. J.; Carter, E. A. Pseudospectral multireference single and double excitation configuration interaction. *J. Chem. Phys.* **1995**, *102*, 7564–7572.
- (35) Martínez, T. J.; Carter, E. A. Pseudospectral double excitation configuration interaction. *J. Chem. Phys.* **1993**, *98*, 7081–7085.
- (36) 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.
- (37) Löwdin, P.-O. Studies in Perturbation Theory. IX. Connection between Various Approaches in the Recent Development—Evaluation of Upper Bounds to Energy Eigenvalues in Schrödinger’s Perturbation Theory. *J. Math. Phys.* **1965**, *6*, 1341–1353.
- (38) Löwdin, P.-O. Some properties of inner projections. *Int. J. Quantum Chem.* **1970**, *5*, 231–237.
- (39) Folkestad, S. D.; Kjønsdal, E. F.; Koch, H. An efficient algorithm for Cholesky decomposition of electron repulsion integrals. *J. Chem. Phys.* **2019**, *150*, 194112.
- (40) Izsák, R.; Neese, F. An Overlap Fitted Chain of Spheres Exchange Method. *J. Chem. Phys.* **2011**, *135*, 144105.
- (41) 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.
- (42) Dutta, A. K.; Neese, F.; Izsák, R. Towards a Pair Natural Orbital Coupled Cluster Method for Excited States. *J. Chem. Phys.* **2016**, *145*, 034102.
- (43) 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*, 094111.
- (44) 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.
- (45) Kossmann, S.; Neese, F. Efficient Structure Optimization with Second-Order Many-Body Perturbation Theory: The RIJCOSX-MP2Method. *J. Chem. Theory Comput.* **2010**, *6*, 2325–2338.
- (46) Kossmann, S.; Neese, F. Comparison of two efficient approximate Hartree–Fock approaches. *Chem. Phys. Lett.* **2009**, *481*, 240–243.
- (47) 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.
- (48) Hättig, C. Optimization of auxiliary basis sets for RI-MP2 and RI-CC2 calculations: Core-valence and quintuple- ζ basis sets for H to Ar and QZVPP basis sets for Li to Kr. *Phys. Chem. Chem. Phys.* **2005**, *7*, 59–66.
- (49) DePrince, A. E.; Sherrill, C. D. Accuracy and Efficiency of Coupled-Cluster Theory Using Density Fitting/Cholesky Decomposition, Frozen Natural Orbitals, and a t 1 -Transformed Hamiltonian. *J. Chem. Theory Comput.* **2013**, *9*, 2687–2696.
- (50) Friesner, R. A. Solution of Self-Consistent Field Electronic Structure Equations by a Pseudospectral Method. *Chem. Phys. Lett.* **1985**, *116*, 39–43.
- (51) Friesner, R. A. Solution of the Hartree–Fock equations by a pseudospectral method: Application to diatomic molecules. *J. Chem. Phys.* **1986**, *85*, 1462–1468.

- (52) 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.
- (53) Ringnalda, M. N.; Belhadj, M.; Friesner, R. A. Pseudospectral Hartree–Fock Theory: Applications and Algorithmic Improvements. *J. Chem. Phys.* **1990**, *93*, 3397–3407.
- (54) Friesner, R. A. New Methods For Electronic Structure Calculations on Large Molecules. *Annu. Rev. Phys. Chem.* **1991**, *42*, 341–367.
- (55) Martinez, T. J.; Mehta, A.; Carter, E. A. Pseudospectral Full Configuration Interaction. *J. Chem. Phys.* **1992**, *97*, 1876–1880.
- (56) Martinez, T. J.; Carter, E. A. Pseudospectral Møller–Plesset perturbation theory through third order. *J. Chem. Phys.* **1994**, *100*, 3631–3638.
- (57) 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*, 104103.
- (58) 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*, 044103.
- (59) 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*, 221101.
- (60) Parrish, R. M.; Hohenstein, E. G.; Martínez, T. J.; Sherrill, C. D. Tensor Hypercontraction. II. Least-squares Renormalization. *J. Chem. Phys.* **2012**, *137*, 224106.
- (61) Hohenstein, E. G.; Kokkila, S. I. L.; Parrish, R. M.; Martinez, T. J. Quartic scaling second-order approximate coupled cluster singles and doubles via tensor hypercontraction: THC-CC2. *J. Chem. Phys.* **2013**, *138*, 124111.
- (62) Shenvi, N.; Van Aggelen, H.; Yang, Y.; Yang, W.; Schwerdtfeger, C.; Mazziotti, D. The tensor hypercontracted parametric reduced density matrix algorithm: Coupled-cluster accuracy with $O(r_4)$ scaling. *J. Chem. Phys.* **2013**, *139*, 54110.
- (63) Schutski, R.; Zhao, J.; Henderson, T. M.; Scuseria, G. E. Tensor-structured coupled cluster theory. *J. Chem. Phys.* **2017**, *147*, 184113.
- (64) Parrish, R. M.; Zhao, Y.; Hohenstein, E. G.; Martinez, T. J. Rank reduced coupled cluster theory. I. Ground state energies and wavefunctions. *J. Chem. Phys.* **2019**, *150*, 164118.
- (65) 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.
- (66) Song, C.; Martínez, T. J. Analytical Gradients for Tensor Hypercontracted MP2 and SOS-MP2 on Graphical Processing Units. *J. Chem. Phys.* **2017**, *147*, 161723.
- (67) Hohenstein, E. G.; Zhao, Y.; Parrish, R. M.; Martínez, T. J. Rank reduced coupled cluster theory. II. Equation-of-motion coupled-cluster singles and doubles. *J. Chem. Phys.* **2019**, *151*, 164121.
- (68) Hohenstein, E. G.; Martínez, T. J. GPU acceleration of rank-reduced coupled-cluster singles and doubles. *J. Chem. Phys.* **2021**, *155*, 184110.
- (69) Hohenstein, E. G.; Fales, B. S.; Parrish, R. M.; Martínez, T. J. Rank-reduced coupled-cluster. III. Tensor hypercontraction of the doubles amplitudes. *J. Chem. Phys.* **2022**, *156*, 054102.
- (70) Jiang, A.; Turney, J. M.; Schaefer, H. F. Tensor Hypercontraction Form of the Perturbative Triples Energy in Coupled-Cluster Theory. *J. Chem. Theory Comput.* **2023**, *19*, 1476–1486.
- (71) Zhao, T.; Simons, M.; Matthews, D. A. Open-Shell Tensor Hypercontraction. *J. Chem. Theory Comput.* **2023**, *19*, 3996–4010.
- (72) Schmitz, G.; Madsen, N. K.; Christiansen, O. Atomic-batched tensor decomposed two-electron repulsion integrals. *J. Chem. Phys.* **2017**, *146*, 161723.
- (73) Khoromskaia, V.; Khoromskij, B. N. Tensor numerical methods in quantum chemistry: from Hartree–Fock to excitation energies. *Phys. Chem. Chem. Phys.* **2015**, *17*, 31491–31509.
- (74) 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.
- (75) Harshman, R. A. Foundations of the PARAFAC Procedure: Models and Conditions for an “Explanatory” Multimodal Factor Analysis. *UCLA Work. Pap. Phon.* **1970**, *16*, 1–84.
- (76) Hastad, J. Tensor rank is NP-complete. *Algorithms* **1990**, *11*, 644–654.
- (77) Hillar, C. J.; Lim, L.-H. Most Tensor Problems Are NP-Hard. *J. ACM* **2013**, *60*, 1–39.
- (78) 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.
- (79) Beylkin, G.; Mohlenkamp, M. J. Numerical Operator Calculus in Higher Dimensions. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 10246–10251.
- (80) Yu, H.; Larsen, K. G.; Christiansen, O. Optimization methods for tensor decomposition: A comparison of new algorithms for fitting the CP(CANDECOMP/PARAFAC) model. *Chemom. Intell. Lab. Syst.* **2025**, *257*, 105290.
- (81) 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*, 054118.
- (82) Benedikt, U.; Böhm, K.-H.; Auer, A. A. Tensor decomposition in post-Hartree–Fock methods. II. CCD implementation. *J. Chem. Phys.* **2013**, *139*, 224101.
- (83) 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.
- (84) 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*, 244102.
- (85) Madsen, N. K.; Godtliebsen, I. H.; Losilla, S. A.; Christiansen, O. Tensor-decomposed vibrational coupled-cluster theory: Enabling large-scale, highly accurate vibrational-structure calculations. *J. Chem. Phys.* **2018**, *148*, 024103.
- (86) Lesiuk, M. Quintic-scaling rank-reduced coupled cluster theory with single and double excitations. *J. Chem. Phys.* **2022**, *156*, 064103.
- (87) Calvin, J. A.; Peng, C.; Rishi, V.; Kumar, A.; Valeev, E. F. Many-Body Quantum Chemistry on Massively Parallel Computers. *Chem. Rev.* **2021**, *121*, 1203–1231.
- (88) 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*, 044120.
- (89) Joergensen, P.; Simons, J. Ab Initio Analytical Molecular Gradients and Hessians. *J. Chem. Phys.* **1983**, *79*, 334–357.
- (90) Wales, D. J.; Hodges, M. P. Global minima of water clusters (H₂O)_n, $n \leq 21$, described by an empirical potential. *Chem. Phys. Lett.* **1998**, *286*, 65–72.
- (91) 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*, 084102.
- (92) 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.
- (93) 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.
- (94) 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.
- (95) Acar, E.; Dunlavy, D. M.; Kolda, T. G. A scalable optimization approach for fitting canonical tensor decompositions. *J. Chemom.* **2011**, *25*, 67–86.

- (96) Phan, A.-H.; Tichavský, P.; Cichocki, A. Low Complexity Damped Gauss–Newton Algorithms for CANDECOMP/PARAFAC. *SIAM J. Matrix Anal. Appl.* **2013**, *34*, 126–147.
- (97) Tichavsky, P.; Phan, A. H.; Cichocki, A. A further improvement of a fast damped Gauss-Newton algorithm for candecomp-parafac tensor decomposition *2013 IEEE International Conference Acoust. Speech Signal Process*, 2013, pp 5964–5968.
- (98) Espig, M.; Hackbusch, W.; Rohwedder, T.; Schneider, R. Variational calculus with sums of elementary tensors of fixed rank. *Numer. Math.* **2012**, *122*, 469–488.
- (99) 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.