Breaking the curse of dimensionality in electronic structure methods: towards optimal utilization of the canonical polyadic decomposition

Karl Martin Pierce

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Chemistry

Eduard F. Valeyev, Chair

Daniel Crawford

Diego Troya

Nicholas Mayhall

December 10, 2021

Blacksburg, Virginia

Keywords: Electronic Structure, Canonical Polyadic Decomposition, Reduced-Scaling, Coupled Cluster Theory

Breaking the curse of dimensionality in electronic structure methods: towards optimal utilization of the canonical polyadic decomposition

Karl Martin Pierce

(ABSTRACT)

Despite the fact that higher-order tensors (HOTs) plague electronic structure methods and severely limits the modeling of interesting chemistry problems, introduction and application of higher-order tensor (HOT) decompositions, specifically the canonical polyadic (CP) decomposition, is fairly limited. The CP decomposition is an incredibly useful sparse tensor factorization that has the ability to disentangle all correlated modes of a tensor. However the complexities associated with CP decomposition have made its application in electronic structure methods difficult.

Some of the major issues related to CP decomposition are a product of the mathematics of computing the decomposition: determining the exact CP rank is a non-polynomially hard problem, finding stationary points for rank-R approximations require non-linear optimization techniques, and inexact CP approximations can introduce a large degree of error into tensor networks. While other issues are a result of the construction of computer architectures. For example, computer processing units (CPUs) are organized in a way to maximize the efficiency of dense linear algebra and, thus, the performance of routine tensor algebra kernels, like the Khatri-Rao product, is limited. In this work, we seek to reduce the complexities associated with the CP decomposition and create a route for others to develop reduced-scaling electronic structure theory methods using the CP decomposition.

In Chapter 2, we introduce the robust tensor network approximation. This approximation is a way to, in general, eliminate the leading-order error associated with approximated tensors in a network. We utilize the robust network approximation to significantly increase the accuracy of approximating density fitting (DF) integral tensors using rank-deficient CP decompositions in the particle-particle ladder (PPL) diagram of the coupled cluster method with single and double substitutions (CCSD). We show that one can produce results with negligible error in chemically relevant energy differences using a CP rank roughly the same size as the DF fitting basis; which is a significantly smaller rank requirement than found using either a nonrobust approximation or similar grid initialized CP approximations (the pseudospectral (PS) and tensor hypercontraction (THC) approximations). Introduction of the CP approximation, formally, reduces the complexity of the PPL diagram from $\mathcal{O}(N^6)$ to $\mathcal{O}(N^5)$ and, using the robust approximation, we are able to observe a cost reduction in CCSD calculations for systems as small as a single water molecule.

In Chapter 3, we further demonstrate the utility of the robust network approximation and, in addition, we construct a scheme to optimize a grid-free CP decomposition of the order-four Coulomb integral tensor in $\mathcal{O}(N^4)$ time. Using these ideas, we reduce the complexity of ten bottleneck contractions from $\mathcal{O}(N^6)$ to $\mathcal{O}(N^5)$ in the Laplace transform (LT) formulation of the perturbative triple, (T), correction to CCSD. We show that introducing CP into the LT (T) method with a CP rank roughly the size of the DF fitting basis reduces the cost of computing medium size molecules by a factor of about 2.5 and introduces negligible error into chemically relevant energy differences. Furthermore, we implement these low-cost algorithms using newly developed, optimized tensor algebra kernels in the massively-parallel,

block-sparse TiledArray [Calvin, et. al *Chemical Reviews* **2021** 121 (3), 1203-1231] tensor framework.

This work was supported in part by the U.S. National Science Foundation and the Virginia Tech Institute for Critical Technology and Applied Science

Breaking the curse of dimensionality in electronic structure methods: towards optimal utilization of the canonical polyadic decomposition

Karl Martin Pierce

(GENERAL AUDIENCE ABSTRACT)

Electronic structure methods and accurate modeling of quantum chemistry have developed alongside the advancements in computer infrastructures. Increasingly large and efficient computers have allowed researchers to model remarkably large chemical systems. Sadly, for as fast as computer infrastructures grow (Moores law predicts that the number of transistors in a computer will double every 18 months) the cost of electronic structure methods grows more quickly. One of the least expensive electronic structure methods, Hartree Fock (HF), grows quartically with molecular size; this means that doubling the size of a molecule increase the number of computer operations by a factor of 16. However, it is known that when chemical systems become sufficiently large, the amount of physical information added to the system grows linearly with system size.[Goedecker, et. al. *Comput. Sci. Eng.*, **2003**, 5, (4), 14-21] Unfortunately, standard implementations of electronic structure methods will never achieve linear scaling; the disparity between actual cost and physical scaling of molecules is a result of storing and manipulating data using dense tensors and is known as the curse of dimensionality.[Bellman, *Adaptive Control Processes*, **1961**, 2045, 276]

Electronic structure theorists, in their desire to apply accurate methods to increasingly large systems, have known for some time that the cost of conventional algorithms is unreasonably high. These theorists have found that one can reveal sparsity and develop reduced-complexity

algorithms using matrix decomposition techniques. However, higher-order tensors (HOTs), tensors with more than two modes, are routinely necessary in algorithm formulations. Matrix decompositions applied to HOTs are not necessarily straight-forward and can have no effect on the limiting behavior of an algorithm. For example, because of the positive definiteness of the Coulomb integral tensor, it is possible to perform a Cholesky decomposition (CD) to reduce the complexity of tensor from an order-4 tensor to a product of order-3 tensors.[Beebe, et. al. *Int. J. Quantum Chem.*, **1977**, 12, 683-705] However, using the CD approximated Coulomb integral tensors it is not possible to reduce the complexity of popular methods such as Hartree-Fock or coupled cluster theory.

We believe that the next step to reducing the complexity of electronic structure methods is through the accurate application of HOT decompositions. In this work, we only consider a single HOT decomposition: the canonical polyadic (CP) decomposition which represents a tensor as a polyadic sum of products. The CP decomposition disentangles all modes of a tensor by representing an order-N tensor as N order-2 tensors. In this work, we construct the CP decomposition of tensors using algebraic optimization. Our goal, here, is to tackle one of the biggest issues associated with the CP decomposition: accurately approximating tensors and tensor networks. In Chapter 2, we develop a robust formulation to approximate tensor networks, a formulation which removes the leading-order error associated with tensor approximations in a network.[Pierce, et. al. *J. Chem. Theory Comput.*, **2021** 17 (4), 2217-2230] We apply a robust CP approximation to the coupled cluster method with single and double substitutions (CCSD) to reduce the overall cost of the approach. Using this robust CP approximation we can compute CCSD, on average, 2.5-3 times faster and introduce negligibly

small error in chemically relevant energy values. Furthermore in Chapter 3, we again use the robust CP network approximation in conjunction with a novel, low cost approach to compute order-four CP decompositions, to reduce the cost of 10 high cost computations in the the perturbative triple, (T), correction to CCSD. By removing these computations, we are able to reduce the cost of (T) by a factor of about 2.5 while introducing significantly small error.

To my parents, Robert and Kelly

Acknowledgments

This work would not exist without the following:

- My parents and family for their unconditional love and support;
- Dr. Edward Valeev for guiding my endeavors into complex and uncharted mathematical territories;
- Dr. Chong Peng for his patience in teaching me best programming practices;
- Dr. Xiao Wang for his time and tea;
- The current and past members of the Valeev research group as friends and colleagues;
- Hanna Anderson for all of her support, especially during my preliminary exam;
- Kalyani, Ben, Ruhee, Shannon, Nakul, Vibin Marjory, and Kirk, for allowing me to annoy them with talk of mathematics and science;
- My devoted friends: James Kittleman, Zelia Rosas, Julia Sunderland, Mckenzie Johnson, Sarah and Jedd Blackowski, Eleftheria Agioutanti, and Alyssa Gatto; and
- Janey Dike for her love and help through the end of this arduous journey.

Finally, I would like to thank everyone who has showed me their love and support during my graduate school career.

Attribution

Chapter 2 was co-authored by Varun Rishi and Edward F. Valeev. Valeev helped develop the research idea and served as the editor to the manuscript. I implemented the CP decompositions, the tensor algebra, and the reduced cost CP implementations, collected all of the data and contributed to writing the manuscript.

Chapter 3 was co-authored by Fabijan Pavosevic, Yue Bao and Edward F. Valeev. Pavosevic implemented the initial (T) code, computed the quadrature point error and wrote a first draft of the manuscript. Bao rewrote the (T) code to work with future studies. Valeev, again, helped develop the research idea and served as the editor to the manuscript. I implemented the CP decompositions, tensor algebra and reduced cost CP (T) implementations, collected the remaining data and contributed to writing the manuscript

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The complexity of electronic structure theory

Modern electronic structure theory methods are plagued by the *curse of dimensionality*,[3] a term coined by Bellman in the field of dynamic programming, is a computational phenomenon where the number of operations required to estimate a function increases exponentially with the number of inputs. For example, consider the exact Schrödinger equation,[6] the central equation to quantum mechanics, which describes the physical phenomena of chemistry. To solve the Schrödinger equation requires determining the form and then optimizing a wavefunction (Ψ) of N particles where each particle's position and behavior is correlated with every other particle. A naïve choice for the form of the Ψ is a single function of N parameters which can be represented as an order-N tensor. Every mode of this wavefunction tensor describes the kinetic energy of a single particle i and particle i's interaction with all other $N-1$ particles; thus, the dimension of each mode must grow linearly with

the size of the molecule. The curse of dimensionality tells us that storing and optimizing such a wavefunction has a computational cost on the order of N^N.* For example, consider a two-water cluster which contains 20 electrons. The exact wavefunction for this system would contain $20^{20} \approx 1.0 \times 10^{26}$ elements and any operation applied to a tensor of this size would take a standard laptop equipped with two Quad-Core Intel i7 processor about 9 million years to complete.

Obviously, it is not reasonable to define Ψ as a single function of N parameters. Therefore, routinely the wavefunction is approximated as a separable, antisymmetric product of N one-particle functions. Unfortunately, even from this simplified approximation, it is still impractical to determine exact solutions to the Schrödinger equation, as that requires recovering N-body correlation. Algorithms which incorporate the complete picture of many-body correlation still have a restrictively large cost, for example the full correlation-interaction (CI) approach applied to a system of N electrons and K basis-function orbitals requires a factorial large number, $\mathcal{O}(\binom{K}{N})$, of operations.[7] It is, however, computational reasonable to introduce a portion of the low-order correlation. To determine an accurate, low-order correlated, approximate wavefunction solution to the Schrödinger equation, electronic structure theorists have developed many ab-initio approximation algorithms like the Hartree-Fock (HF),[8, 9] Moller-Plesset (MP),[10, 11] and coupled cluster (CC)[12–14] methods. Unfortunately, these approximations too suffer from the curse of dimensionality, though to a lesser degree, with HF scaling as $\mathcal{O}(N^4)$, the popular second-order MP (MP2) approach scaling as $\mathcal{O}(N^5)$ and the family of CC methods scaling as $\mathcal{O}(N^{2k+2})$, where k is the number of cor-

*The operational cost of methods in this work will be referred to using big $\mathcal{O}$ scaling notation, which describes the limiting behavior of an algorithm.

related particles. The computational scaling of these methods is not a result of the physics of electrons but of modern computer architectures which store information in tensors and are optimized to use dense matrix algebra. It is obvious that the curse of dimensionality is actually an issue of data representation. Fortunately, modern mathematics provides us with ways to disentangle computer tensor structures using tensor decompositions and factorizations. These factorizations allow for the development of reduced scaling and low cost algorithms. Thus, much research has gone into the development of tensor decomposition and sparse tensor representation in electronic structure methods. This work seeks to push this field of research by constructing error-minimized approximated tensor network models and reduced-cost method implementations.

1.2 Electronic structure theory methods

1.2.1 First-order approximation to the Schrödinger equation

The central equation to non-relativistic electronic structure theory is the time-independent Schrödinger equation[6]

$$\hat{H}\Psi = E\Psi \tag{1.1}$$

which describes the electronic and nuclear mechanics of a system. In this equation the Hamiltonian operator, $\hat{H}$, acts on a wavefunction Ψ and produces the corresponding energy of the system E. In general, Ψ is a function of electronic and nuclear position however,

the exact form of Ψ is only known for the hydrogen atom. To reduce the complexity of wavefunction, two assumptions are made about the form of Ψ. First, it is assumed that nuclear positions are fixed via the Born-Oppenheimer approximation,[15] and thus, we seek to optimize an electronic wavefunction, ψ, subject to these fixed nuclei. Second, it is assumed that the electronic wavefunction can be written as a product single-electron functions also known as spin-orbitals, $\chi(\mathbf{x}) = \phi(\mathbf{r})\alpha(\omega)$; where the function $\phi(\mathbf{r})$ describes the spatial position of an electron and $\alpha(\omega)$ describes the electronic spin.[16] Note, in the following work we will only consider the spatial orbital portion and thus $\chi(\mathbf{x}) = \phi(\mathbf{r})$ One way to construct a wavefunction ansatz is by using a Slater determinant:[17]

$$\psi\left(\mathbf{r}_1, \ldots \mathbf{r}_N\right) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1\left(\mathbf{r}_1\right) & \phi_2\left(\mathbf{r}_1\right) & \ldots & \phi_N\left(\mathbf{r}_1\right) \\ \phi_1\left(\mathbf{r}_2\right) & \phi_2\left(\mathbf{r}_2\right) & \ldots & \phi_N\left(\mathbf{r}_2\right) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1\left(\mathbf{r}_N\right) & \phi_2\left(\mathbf{r}_N\right) & \ldots & \phi_N\left(\mathbf{r}_N\right) \end{vmatrix}. \tag{1.2}$$

as a determinant naturally introduces antisymmetry into the wavefunction, i.e. swapping any two columns of a determinant introduces a negative sign into the wavefunction. Furthermore, a determinant based wavefunction introduces Fermi correlation, i.e. it prevents the formation of a wavefunction with two electrons to occupy the same spatial position and spin-states. The set of one-electron functions used to describe a single particle is called the atomic orbital (AO) basis and is denoted by the indices $\mu, \nu, \sigma, \lambda \ldots$. The Hartree-Fock method, discussed in the next section, uses variational minimization to find the best set of AOs which minimizes

the electronic energy expression

$$E_0 = \langle \psi | H_{\text{elec}} | \psi \rangle \tag{1.3}$$

where H_{elec} is defined in Eq. (1.5)

1.2.2 The Hartree-Fock approximation

The Hartree-Fock (HF) method[8, 9] seeks to determine the best ground-state, single-reference Slater determinant via variational minimization of the electronic wavefunction

$$E_0(\psi) = \langle \psi | \hat{H}_{\text{elec}} | \psi \rangle. \tag{1.4}$$

for a system of N electrons and M nuclei, where

$$\hat{H}_{\text{elec}} = -\sum_{i}^{N} \frac{1}{2} \nabla_i^2 - \sum_{i}^{N} \sum_{A}^{M} \frac{Z_A}{r_{iA}} + \sum_{i}^{N} \sum_{j>i}^{N} \frac{1}{r_{ij}} \tag{1.5}$$

where ∇_i describes the kinetic energy of the ith electron, $\frac{Z_A}{r_i A}$ describes the Coulombic interaction between the ith electron with the Ath nucleus, and $\frac{1}{r_{ij}}$ describes the Coulombic interaction between the ith and jth electron. Algebraic solutions to HF can found for molecules using a set of molecular orbitals (MOs) which are, by definition, linear combinations of the finite set of atomic orbital (AO) basis functions: [18]

$$|\psi_i\rangle = \sum_{\mu}^{K} C_{\mu i} \phi_\mu \tag{1.6}$$

where $C_{\mu i}$ are MO expansion coefficients. The molecular HF problem reduces to determining optimized MO expansion coefficients which can be done using the matrix Roothaan equations[16]

$$FC = SC\epsilon \tag{1.7}$$

where F is the Fock operator, S is a matrix describing the AO overlap and ϵ is the MO orbital eigenvalues. The variational HF minimization, in the limit of an infinite AO basis, determines the best single-determinant, self-consistent, mean-field approximation to the ground-state electronic wavefunction.

The MOs computed using Eq. (1.7) are called the canonical HF molecular orbitals and can be partitioned into N occupied orbitals, which represent the lowest N eigenvalue labeled $i, j, \ldots$ and the remaining $K - N$ occupied orbitals labeled $a, b, \ldots$, where K is the number of AO basis functions. The N occupied MO orbitals can be used to form the ground-state single determinant HF wavefunction.[18] The MO subspaces are invariant to unitary rotations, provided the rotations do not mix the occupied and unoccupied subspaces. This is otherwise known as Brillouin's theorem.[19, 20]

The HF method has, formally, an asymptotic complexity of $\mathcal{O}(N^4)$, where N is a measure of the system size. The most cost intensive portion of HF is the construction of the Fock

tensor, specifically, the computation of the two electron integral tensor (TEI)[†]:

$$g^{\mu\nu}_{\rho\sigma} = \langle \mu\nu|\rho\sigma\rangle = (\mu\rho|\nu\sigma) = \iint \frac{\phi^*_\mu(\mathbf{r}_1)\phi_\rho(\mathbf{r}_1)\phi^*_\nu(\mathbf{r}_2)\phi_\sigma(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2 \tag{1.8}$$

where $\frac{1}{r_{12}} = \frac{1}{|\mathbf{r}_1 - \mathbf{r_2}|}$, is the Poisson kernel. For large systems it is possible to utilize integral screening techniques, such as the Schwartz inequality,[21] to reduce the complexity of HF algorithm to $\mathcal{O}(N^2 ln(N))$. The HF procedure is a relatively, inexpensive ab initio approach to solving the Schrödinger equation. However, its base assumption, that the wavefunction is a product of N non-interacting functions, is incorrect.[22] Accurately determining exactly how electrons in a chemical system interact with each other is the correlation problem.

1.2.3 Electronic correlation

Accurate modeling of instantaneous, electron-electron interactions is necessary to predict chemical properties and reactions. For example, correlation is necessary to accurately model protein folding and Uracil molecule stacking.[23–25] Uracil is a nucleobase found in RNA. Understanding how Uracil molecules stack is important for biology, genomics, and drug development. We know that the primary force which stabilizes nucleobase stacking is dispersion, a force which exists as a bi-product of electron correlation. Unfortunately, the HF method is flawed because the independent particle picture fails to consider instantaneous effects of Coulomb repulsion; the HF method simply subjects each electron to a mean-field of charge generated by the other N-1 electrons.[22] Thus, methods that don't fully consider

[†]Note, the physicist bra-ket notation uses $\langle \mathbf{r}_1\mathbf{r}_2|\mathbf{r}_1\mathbf{r}_2\rangle$ convention, while the chemist notation uses $(\mathbf{r}_1\mathbf{r}_1|\mathbf{r}_2\mathbf{r}_2)$ and tensors follow the standard $T^{\mathbf{r}_1\mathbf{r}_2}_{\mathbf{r}_1\mathbf{r}_2}$

the effects of electron correlation, like HF, cannot properly model nucleobase stacking and protein folding.[26]

Electron correlation has been defined by Löwdin[27] as:

$$E_{\text{corr}} = \mathcal{E}_0 - E_{\text{HF}} \tag{1.9}$$

where $\mathcal{E}_0$ is the exact non-relativistic energy and E_{HF} is the HF energy at the HF limit, i.e. using an infinite basis set However this definition is inexact because HF does partially correlate electrons of parallel spin. Thus, it is more accurate to define E_{corr} as an observable of a wavefunction, Ψ_{corr}, which is orthogonal to a reference wavefunction (here we only consider a HF reference wavefunction) and encapsulates the physical interactions which the reference wavefunction does not consider,[28] i.e.

$$\Psi_{\text{exact}} = \Psi_{\text{HF}} + \Psi_{\text{corr}}. \tag{1.10}$$

To follow we introduce popular post-HF methods, explain how these methods construct Ψ_{corr} to approximate a system's correlation energy, and discuss the costs associated with each approach.

1.2.4 Perturbation based corrections

Many body perturbation theory (MBPT) was first used in nuclear physics in 1957[29] and was later applied to ab initio electronic structure theory in 1968.[30, 31] Starting from the

exact, eigenvalue representation of the Schrödinger equation

$$\hat{H}_{\text{exact}}|\Psi_{\text{exact}}\rangle = \epsilon_0|\Psi_{\text{exact}}\rangle \tag{1.11}$$

perturbation theory assumes that there exists a zero-order approximation to this equation which is nearly the exact solution. The exact solution, then, can be expressed as the zeroth-order approximation plus a series of small perturbations. For example, the exact Hamiltonian can be expressed as

$$\hat{H}_{\text{exact}} = \hat{H}_0 + \lambda\hat{H}_1 + \ldots \lambda^n \hat{H}_n \tag{1.12}$$

where λ represents a small valued perturbation. To solve for the n-order perturbative correction to the energy, one expands each term in Eq. (1.11), collects equivalent orders of perturbation (λ), and solves for the nth-order of λ.

Here, the presentation of MBPT will follow the popular Moller-Plesset (MPn) definition.[10, 11] In MPn theory, one assumes that HF is the zeroth-order approximation to Eq. (1.11). HF canonical MOs are subject to Brillouin's theorem[19] which ensures that first-order perturbations are identically zero. One of the most popular corrections is the MP2 correction because of its relatively simple formulation and low computational cost:

$$E_{\text{MP2}} = \frac{1}{4}\sum_{ijab} \frac{\|\langle ab||ij\rangle\|^2}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j} \tag{1.13}$$

where $\langle ab||ij\rangle$ is the anti-symmetrized TEI tensor

$$\langle ab||ij\rangle = \langle ab|ij\rangle - \langle ab|ji\rangle \tag{1.14}$$

and ϵ are MO energies, which are the the eigenvalues to the canonical HF MOs. Computation of MP2 has a computational complexity of $\mathcal{O}(N^5)$ and a storage requirement of $\mathcal{O}(N^4)$. The computational scaling comes from the AO to MO orbital transformation via

$$\langle ab|ij\rangle = \sum_{\mu\nu\rho\sigma} C_{\mu a} C_{\nu b} C_{\rho i} C_{\sigma j} \langle \mu\nu|\rho\sigma\rangle \tag{1.15}$$

MP2 is a useful method because it is a relatively inexpensive, single calculation route to recovering many-body correlation. Unfortunately, MP2 has some serious issues accurately predicting correlation energy because it assumes HF mean field orbitals as the reference state.[32]

Another popular perturbative correction is the triple substitution, (T), correction to the coupled cluster with single and double substitution (CCSD) method.[13] The (T) energy consists of two energy contributions

$$E_{(\mathrm{T})} = E_T^{[4]} + E_{ST}^{[5]} \tag{1.16}$$

and can be understood as the MP4 and MP5 order corrections to the HF zeroth-order state. The (T) correction is praised for its balance between computational cost and accuracy in

modeling electron correlation. The energy terms in Eq. (1.16) are defined as

$$E_{\mathrm{T}}^{[4]} = \frac{1}{3}\sum_{ijk}\sum_{abc}\frac{W_{abc}^{ijk}R_{abc}^{ijk}}{\epsilon_a+\epsilon_b+\epsilon_c-\epsilon_i-\epsilon_j-\epsilon_k}, \tag{1.17}$$

$$E_{\mathrm{ST}}^{[5]} = -\frac{1}{3}\sum_{ijk}\sum_{abc}\frac{V_{abc}^{ijk}R_{abc}^{ijk}}{\epsilon_a+\epsilon_b+\epsilon_c-\epsilon_i-\epsilon_j-\epsilon_k}. \tag{1.18}$$

The tensors W, R, and V are constructed from optimized coupled cluster single and double substitution tensors and the TEI tensor; a complete definition be found in Section 3.2. One of the biggest issues with the (T) correction is its restrively large, $\mathcal{O}(N^7)$, computational scaling. However, reducing the computation required to find the (T) energy correction is currently being studied by a number of researchers. In Chapter 3, we make great strides in reducing the cost of the (T) correction by introducing integral tensors approximated using the canonical polyadic (CP) decomposition.

These perturbation based theories are among the least cost intensive tools to incorporate many-body correlation as they require no iterative optimization schemes. However, they work under the assumption that the reference state is nearly the correct answer, which is not the case for HF. The family of CC methods are a successful, though cost intensive, way to systematically introduce correlation into an uncorrelated reference wavefunction.

1.2.5 Coupled cluster methods

Coupled-cluster (CC) methods were first developed by Čížek and Paldus in the 1960's[33–35] and are known for their exceptional accuracy and rapid and systematic inclusion many body correlation.[36–39] The CC family of methods is defined by the exponential wavefunction

ansatz: $|\Psi_{CC}\rangle$,

$$|\Psi_{\mathrm{CC}}\rangle \equiv e^{\hat{T}}|\Phi_0\rangle, \tag{1.19}$$

with $|\Phi_0\rangle$ being a reference state (typically, the ground-state Hartree-Fock determinant) and $\hat{T}$ being a cluster operator composed of 1-, 2-, ... N-particle components:

$$\hat{T} \equiv \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \dots \tag{1.20}$$

In the *traditional* coupled-cluster method $\hat{T}_p$ are pure excitation operators that promote p particles from the occupied MOs in $|\Phi_0\rangle$ to the unoccupied (virtual) orbitals:

$$\hat{T}_p \equiv \frac{1}{(p!)^2} \sum_{i_1 \dots i_p, a_1 \dots a_p} t^{i_1 \dots i_p}_{a_1 \dots a_p} a^\dagger_{a_1} \dots a^\dagger_{a_p} a_{i_p} \dots a_{i_1} \tag{1.21}$$

where $a_t/a_t^\dagger$ are fermionic annihilation/creation operators such that a_t deletes an orbital ϕ_t from Φ_0 and $a_p^\dagger$ creates an orbital ϕ_p in Φ_τ.[36, 40] In traditional CC, t amplitudes in Eq. (1.21) are determined by projecting the electronic Schrödinger equation onto left-hand excited determinants $\langle\Phi_\tau e^{-\hat{T}}|$, where $|\Phi_\tau\rangle$ is a determinant obtained from $|\Phi_0\rangle$ by exciting 1 or more particles. The resulting equations,

$$\langle\Phi_\tau|e^{-\hat{T}}\hat{H}e^{\hat{T}}|\Phi_0\rangle = \langle\Phi_\tau|\bar{H}|\Phi_0\rangle = 0, \tag{1.22}$$

involve the *similarity-transformed* Hamiltonian $\bar{H} \equiv e^{-\hat{T}}\hat{H}e^{\hat{T}}$ that has finite particle rank due to its truncating commutator expansion:[36]

$$\bar{H} = \hat{H} + [\hat{H}, \hat{T}] + \frac{1}{2!}[[\hat{H}, \hat{T}], \hat{T}] + \frac{1}{3!}[[[\hat{H}, \hat{T}], \hat{T}], \hat{T}] + \frac{1}{4!}[[[[\hat{H}, \hat{T}], \hat{T}], \hat{T}], \hat{T}] \tag{1.23}$$

Construction of CC using the exact N-particle cluster operator is infeasible as the computational complexity grows with particle rank k as $\mathcal{O}(N^{2k+2})$. Fortunately, due to the rapid (exponential)[37–39] convergence of the CC ansatz with the particle rank, the cluster operator can be limited to a low particle rank. The least expensive truncation' are the 2-particle substitutions (CCD) and 1- and 2-particle substitutions (CCSD) and such truncations have a computational complexity of $\mathcal{O}(N^6)$ and a storage complexity of $\mathcal{O}(N^4)$. Inclusion of exact 3-particle substitutions (CCSDT)[14] is restrictive for large systems[12, 41–43] as CCSDT has a computational complexity of $\mathcal{O}(N^8)$ and a storage complexity of $\mathcal{O}(N^6)$. Fortunately, as discussed in the previous section, it is possible to approximate the 3-particle substitutions using a perturbative correction on top of the CCSD ansatz, CCSD(T),[13] which has a computational complexity of $\mathcal{O}(N^7)$.

The computational and storage complexity of the CC family of methods, and all the methods discussed up to this point, is an artifact of representing data on computers using dense tensors. Conventional tensors entangle each mode of a tensor with every other mode. For example, scaling a single mode of a k dimensional tensor with d elements requires d^k operations. Determining a sparse representation for a tensor reduces the complexity of storing and using modes of such a tensor. Thus, much work, including that which is presented here, has

been done to find accurate and efficient sparse tensor representations in electronic structure methods.

1.3 Tensor approximations

A tensor is any multidimensional array; commonly, an order-one tensor is known as a vector or array, an order-two tensor is known as a matrix and a tensor of order-three or higher is known as a higher-order tensor (HOT).[44] Because of the topology of tensors and tensor-networks, as a chemical system grows the amount of information and number of computations required to model a system will not depend on the amount of physical information being introduced but will in fact, be directly correlated to the highest-order tensor-network in the method.

The computational scaling introduced by dense tensors algebra severely limits the application of accurate electronic structure methods. Figure 1.1 demonstrates the limitation of standard implementations of of 1- and 2- particle methods, like HF and CCD respectively. [‡] A realistic system of RNA would contain over 200 nucleotides[46] and modeling such a system in a relatively accurate way would take far too long. Fortunately, this is not the end for ab initio quantum chemistry because the computational representation is unphysical and, using mathematical approximations, one can reduce the computational complexity of electronic structure methods. In the following sections will introduce and discuss the motivation and

[‡]In Figure 1.1, The labeled lines represent the time it takes to complete a PhD program in the US (roughly 5 years), the time since the creation of the first computer in 1945 (Electronic Numerical Integrator and Computer, labeled ENIAC), the time since the construction of the first pyramid in 2650 BCE (labeled Pyramids), and the age of the universe (labeled Universe)[45]

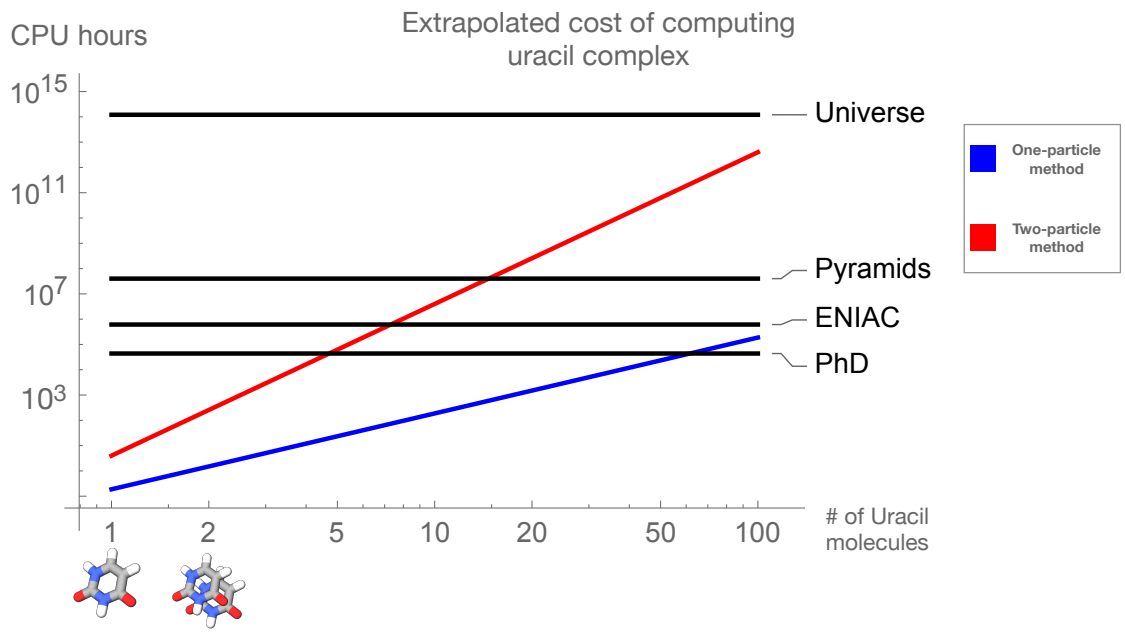


Figure 1.1: Extrapolated cost of computing the energy of a complex which contains $n \in [1, 100]$ uracil molecules with a single-particle approach, like HF, or a two-particle approach, like CCD. Figure courtesy of Dr. Edward F. Valeev

application of tensor decomposition in electronic structure theory methods.

1.3.1 Matrix decompositions of the two-electron integral tensor

One of the most commonly used sparse tensor representation in electronic structure methods is the matrix approximation of the two-electron integral (TEI) tensor, g from Eq. (1.8). The goal is to represent the order-4 tensor g as a product of lower-order tensors

$$g = BB^{\top}. \tag{1.24}$$

This sparse representation can be constructed two ways: using the Cholesky decomposition (CD) or using the density fitting (DF) approximation. Both of theses constructions utilize fact that the TEI tensor, and by extension the Coulomb kernel, $\frac{1}{r_{12}}$, is positive (semi)-definite.

The DF and CD both start by approximating the order-four g tensor as a matrix using a one-electron densities:

$$g_{\mu\sigma,\nu\lambda} = \iint \rho_{\mu\sigma}(\mathbf{r}_1)\hat{M}(\mathbf{r}_1,\mathbf{r}_2)\rho_{\nu\lambda(\mathbf{r}_2)}dr_{\mathbf{1}}d\mathbf{r}_2 \tag{1.25}$$

where

$$\int \rho_{\mu\sigma}(\mathbf{r}_1)d\mathbf{r}_1 = \int \phi^*_\mu(\mathbf{r}_1)\phi_\sigma(\mathbf{r}_1)d\mathbf{r}_1, \tag{1.26}$$

$\hat{M}(\mathbf{r}_1,\mathbf{r}_2)$ is a postive definite two-particle operator, i.e. the Coulomb kernel $\frac{1}{\mathbf{r}_{12}}$, and the comma notation indicates indices which make up the column and row-wise modes of the matricized tensor.

The Cholesky decomposition (CD) is an algebraic matrix decomposition which converts any square, symmetric matrix A into a product of lower triangular matrices

$$A = LL^\top. \tag{1.27}$$

In theory, the CD requires the matrix A to be positive definite; however, in practice, semi-definiteness creates no complication as the algorithm to compute the CD stops when diagonal elements of the matrix A are smaller than a threshold value. The CD was first used to to approximate the TEI tensor in 1977 by Beebe and Linderburg.[4] Using the CD, one can numerically compute the rank-P decomposition of the matrix g which results in the

factorization

$$g_{\mu\sigma,\nu\lambda} = \sum_{p=1}^{P} B^{P}_{\mu\sigma} B^{P}_{\nu\lambda} \tag{1.28}$$

The numerical CD of the TEI typically finds a rank, P, 3-10 times greater than the number of basis functions N.[47] The CD of the TEI is, typically, more accurate and can be more practical than the DF approximation;[48] however, it is more cost intensive to construct than the DF approximation. Much work has put into finding efficient CD algorithms, for more details see the recent developments by Folkestad et al.[49] and Zhang et al.[48]

The density fitting (DF) approximation, also known as the resolution of identity (RI), is a specific formulation of the CD which uses a predetermined auxiliary basis.[49] The DF method approximates one-electron densities using a grid or auxiliary basis:

$$\rho_{\mu\sigma}(r_{\mathbf{1}}) = \sum_{X}^{N_{\text{fit}}} d^{\mu\sigma}_{X} \chi_X(r_{\mathbf{1}}) \tag{1.29}$$

where the auxiliary basis χ_X is abbreviated as X and Y. The DF fitting coefficients d are constructed using the formula

$$d^{\mu\sigma}_{X} = \sum_{Y} (\mu\sigma|Y)[J^{-1}]_{XY}. \tag{1.30}$$

where the three-body integral tensor is defined as

$$(\mu\sigma|Y) = \iint \frac{\phi^*_{\mu}(\mathbf{r}_1)\phi_{\sigma}(\mathbf{r}_1)\chi_Y(\mathbf{r}_2)}{r_{\mathbf{12}}} d\mathbf{r}_1 d\mathbf{r}_2 \tag{1.31}$$

and the two-body metric tensor J is defined as

$$J_{XY} = \iint \chi_X(\mathbf{r}_1)\hat{M}(\mathbf{r}_1\mathbf{r}_2)\chi_Y(\mathbf{r}_2)d\mathbf{r}_1 d\mathbf{r}_2 \tag{1.32}$$

such that $\hat{M}$ is a two-particle kernel. One of the most popular form of the two-index metric tensor is the Coulomb metric[50, 51] where $\hat{M} = \frac{1}{\mathbf{r}_{12}}$. The Coulomb metric is praised for its accuracy in electronic structure calculation, though other metrics have been proposed[52] which can be computed more quickly. Using Eq. (1.29), one can express the TEI matrix, as

$$(\mu\sigma|\nu\lambda) = \sum_{XY}(\mu\sigma|A)[J^{-1/2}]_{XA}[J^{-1/2}]_{BX}(B|\nu\lambda) = \sum_{X} B_{\sigma}^{\mu X} B_{\lambda}^{\nu X} \tag{1.33}$$

provided that the metric tensor J is positive (semi)definite. The *half-transformed* tensor B is defined as

$$B_{\sigma}^{\mu X} = \sum_{A}(\mu\sigma|A)[J^{-1/2}]_{AX} \tag{1.34}$$

. The density fitting basis is typically 2-3 times larger than the number of orbital basis functions. Though the DF approximation is, in general, less accurate than CD, it can be computed relatively quickly, as it does not require the numerical decomposition of an order-four tensor.

Matrix decompositions of the TEI tensor are widely used in electronic structure methods because they can significantly reduce the storage requirements of the TEI tensor. However, their application in the effort to construct reduce-scaling algorithms is relatively limited

because these matrix decompositions do not decouple indices which describe orbitals of the a single particle. Fortunately, there are many applications where these approximations can be used to reduce some high-scaling contractions and, therefore, the computational prefactor of such methods. Understanding the limitation of matrix decompositions applied to higher-order tensor algebra leads us to the next step in tensor approximations: utilization of higher-order tensor decomposition schemes.

1.3.2 Higher-order tensor decomposition and the canonical polyadic decomposition

Because matrix decomposition schemes, like the DF approximation, cannot, in general, reduce the scaling of electronic structure methods, we look to higher-order formulations which can completely disentangle higher-order tensors (HOTs). HOT decompositions may be viewed simply as extrapolations of matrix decomposition schemes but, unfortunately, the well behaved properties of analogous matrix decompositions do not carry. Matrix decomposition methods are well-understood, straightforward approaches which transform general order-two tensors into a standard matrix product representation using a finite algorithm. Matrix factorization can discover latent information, such as the *rank* of the matrix which is the number of unique vectors which span the matrix space, and there are finite algorithms which can discover this information; for example, the rank is the number of linearly independent rows or columns in a matrix. HOT decomposition too transform (order-N) tensors into standard tensor products, however, in general, constructing optimal representations in this form is non-trivial. In the following, we will introduce the canonical polyadic (CP) de-

composition, discuss its strengths and flaws, and show how it may be constructed for tensors used in electronic structure methods.

As discussed before, a HOT is a multi-index array which has three or more modes. In this work we assume tensors are real but, in general, this is not required. Here, we will limit our discussion of HOT decompositions to the CP decomposition, though other decomposition topologies do exist such as the Tucker decomposition[44, 53] and tensor train.[54, 55] The goal of the CP decomposition is to express the tensor $\mathcal{T}$ as a sum of R_{CP} rank-one tensors:

$$\mathcal{T} = \sum_{r}^{R_{\text{CP}}} \lambda_r \mathcal{X}_r \quad \lambda \in \mathbb{R}. \tag{1.35}$$

In Eq. (1.35) λ_r is the weight of the rank-one tensor $\mathcal{X}_r$ and is determined by normalizing the vector components of $\mathcal{X}_r$. A rank-one, order-k tensor is defined as the outer product of k vectors[44]

$$\mathcal{X} = a^{(1)} \otimes a^{(2)} \otimes a^{(3)} \otimes \cdots \otimes a^{(k)} \tag{1.36}$$

such that

$$\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_k} \tag{1.37}$$

$$a^{(1)} \in \mathbb{R}^{I_1}, \quad a^{(2)} \in \mathbb{R}^{I_2}, \quad \ldots \quad a^{(k)} \in \mathbb{R}^{I_k} \tag{1.38}$$

where I_i is the dimension of the ith mode. One can form a *CP factor matrix* using the set

Figure 1.2: a. Graphical representation of an order-3 tensor; b. Graphical representation of a CP decomposition of an order-3 tensor

of R_{CP} vectors which span a single mode of $\mathcal{T}$, i.e.

$$A^{(1)} = \{a_1^{(1)}, a_2^{(1)}, \ldots, a_{R_{\text{CP}}}^{(1)}\} \tag{1.39}$$

Using this definition, one can also represent the CP decomposition as a set of factor matrices coupled by the hyperdimension R_{CP}. Figure 1.2 uses a graphical tensor representation to illustrate how the CP decomposition constructs a sparse representation of an order-3 tensor. In this graphical representation, tensors are nodes, the order of a tensor is the number of edges connected to the node, and the dimension of each mode are labels on the edges.

Representing a HOT using the CP decomposition requires, first, determining the rank of the HOT and, second, computing the optimized CP factor matrices. Unlike matrix decompositions, there is no finite algorithm to determine the rank of a HOT.[56, 57] Typically, one must optimize multiple rank-$R < R_{\text{CP}}$ approximations and choose the R which constructs an acceptable approximation based on the tensor's application. Next, finding the optimal rank-R CP decomposition requires the minimization of a non-linear loss function

$$\min f(\mathbf{A}) = \frac{1}{2}\|\mathcal{T} - \hat{\mathcal{T}}(\mathbf{A})\|^2 \tag{1.40}$$

where $\hat{\mathcal{T}}(\mathbf{A})$ is the CP approximation of the HOT $\mathcal{T}$ computed using the set of k factor matrices and the weighting coefficients λ:

$$\hat{\mathcal{T}}(\mathbf{A}) = \sum_{r}^{R} \lambda_r A_r^{(1)} A_r^{(2)} \dots A_r^{(k)}. \tag{1.41}$$

and $\mathbf{A}$ is a vector of CP factors $\mathbf{A} = \{A^{(1)}, A^{(2)}, \dots, A^{(k)}\}$ Fortunately, there are many algorithms to minimize this objective function.[58–60]

In this work, we numerically optimize the CP decomposition of Coulomb integral tensors using an alternating least squares (ALS) algorithm. The ALS algorithm is a linear approximation of the non-linear objective function, Eq. (1.40). The ALS turns the Eq. (1.40) into a set of k linear objective functions by fixing all factor matrices except for one and solving for that single matrix.[44] The optimization cycles through the k linear objective functions until a stopping condition is met; typically this condition is that the *change* in the nonlinear CP objective function, Eq. (1.40), becomes small.

The ALS algorithm is easy to understand and implement but it does have its shortcomings: ALS optimization is heavily guess-dependent; the ALS algorithm cannot determine minima or stationary points of the objective function, Eq. (1.40); and ALS optimizations can stagnate when factors become nearly degenerate.[44] Robust, gradient and nonlinear least square algorithms exist[59] however, these algorithms can require significantly more computational resources than the ALS algorithm. Because of the complications related to numerical optimization of the CP decomposition, it is useful to point out paths for grid-based CP factor matrix discovery.

1.3.3 Grid-based higher-order tensor decompositions

Much like the CD approximation, formulation of CP approximated HOT via numerical optimization is robust and accurate, but can be time and memory intensive depending on the order of the tensor, the rank of the CP decomposition, and the CP optimization scheme. Therefore, researchers have worked to develop CP approximations of HOT which utilize factor matrices constructed using grids. In Chapter 2, we compare our numerical CP factorization of the DF approximated TEI tensor to two grid-based CP initialization schemes: the pseudospectral (PS)[61–65] and the tensor hypercontraction (THC)[66–73] methods. Both of these schemes seek to reduce the complexity of the TEI tensor using a grid-based order-three CP decompositions. The PS method works by approximating the integration of a single, one-particle density using a finite sum over a quadrature grid, i.e.

$$(\mu\sigma|\nu\lambda) \overset{\text{PS}}{\approx} \sum_g w_g \phi^*_\mu(\mathbf{r}_g)\phi_\sigma(\mathbf{r}_g) \int \hat{M}(\mathbf{r}_g, \mathbf{r}_2)\phi^*_\nu(\mathbf{r}_2)\phi_\lambda(\mathbf{r}_2) d\mathbf{r}_2 \tag{1.42}$$

where w_g is the weight of the quadrature grid point, g. This quadrature factorization allows the TEI tensor to expressed as

$$(\mu\sigma|\nu\lambda) = \sum_g X^*_{\mu,g} Y_{\nu\lambda,g} X_{\sigma,g} \tag{1.43}$$

where

$$X_{\mu,g} \equiv \sqrt{w_g}\phi_\mu(\mathbf{r}_g) \tag{1.44}$$

and

$$Y_{\lambda}^{\nu g} \equiv \int \hat{M}(\mathbf{r}_g, \mathbf{r}_2)\phi_{\nu}^{*}(\mathbf{r}_2)\phi_{\lambda}(\mathbf{r}_2). \tag{1.45}$$

The THC approach is the next step in factorization, as it factorizes both sets of particle densities and can by achieved by substituting each exact one-particle density with a quadrature grid

$$(\mu\sigma|\nu\lambda) = \sum_{g_1}\sum_{g_2} w_{g_1} w_{g_2} \phi_{\mu}^{*}(\mathbf{r}_{g_1})\phi_{\sigma}(\mathbf{r}_{g_1})\hat{M}(r_{g_1}, r_{g_2})\phi_{\nu}^{*}(\mathbf{r}_{g_2})\phi_{\lambda}(\mathbf{r}_{g_2}) \tag{1.46}$$

where w_{g_1} is the weight of the quadrature grid point g_1 and where w_{g_2} is the weight of the quadrature grid point g_2. Equivalently, the THC can be constructed by approximating each order-three DF integral tensor on a quadrature grid

$$(\mu\sigma|X) = \sum_{g_1} w_{g_1} \phi_{\mu}^{*}(\mathbf{r}_{g_1})\phi_{\sigma}(\mathbf{r}_{g_1})\chi_X(\mathbf{r}_{g_1}) \tag{1.47}$$

where, again, w_{g_1} is the weight of the quadrature grid point g_1. In the end THC has the form:

$$(\mu\sigma|\nu\lambda) \overset{\text{THC}}{\approx} \sum_{g_1}\sum_{g_2} X_{a,g_1}^{*} X_{b,g_1} Y_{g_1,g_2} X_{c,g_2}^{*} X_{d,g_2}. \tag{1.48}$$

Where X are the orbital-basis factor matrices and Y is a factor matrix which connects the two grid and contains information about the two-particle kernel $\hat{M}$.

Another grid-based CP factorization is the Laplace tranformation (LT) utilized in pertur-

bation theories like MP2 and (T).[74–78] The LT is used in Chapter 3 to deconstruct the canonical HF orbital energy denominator tensor,

$$D_{abc}^{ijk} = \frac{1}{\epsilon_a + \epsilon_b + \epsilon_c - \epsilon_i - \epsilon_j - \epsilon_k}, \tag{1.49}$$

into a product of 6 tensors using a quadrature grid, i.e.

$$D_{abc}^{ijk} \approx \sum_{\alpha} w_\alpha e^{-\epsilon_a s_\alpha} e^{-\epsilon_b s_\alpha} e^{-\epsilon_c s_\alpha} e^{\epsilon_i s_\alpha} e^{\epsilon_j s_\alpha} e^{\epsilon_k s_\alpha} \tag{1.50}$$

where w_α is the weight of the quadrature grid point α. Introduction of the LT factorized denominator tensor simplifies perturbative energy expressions by removing their dependency on canonical HF orbitals.

Grid-based CP initialization is incredibly useful because it can be constructed much more quickly than numerical optimization based initialization. However, one issue that grid-based schemes face is choosing a grid representation which can maximize accuracy and minimize the number of required grid points. We recognize that grid-based initialization is necessary to accelerate relatively low-cost methods such as HF and MP2 and thus, it is the goal of this work to construct means to minimize the error introduced by grid-based CP factorizations.

Chapter 2

Robust approximation of tensor networks: application to grid-free tensor factorization of the Coulomb interaction

2.1 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[79, 80] and local[81–83]), the pseudospectral[61, 62, 64, 84–90] (PS) approach, Cholesky decomposition (CD),[4, 91–93] the fast multipole method (FMM),[94–96] tensor hypercontraction (THC),[66–68, 70, 71, 73, 97–99] the canonical polyadic (CP) decomposition (also known as CANDECOMP/PARAFAC[100, 101]), [102–108] and many others.[109–118] These approaches can be coarsely classified as (a) abstract (algebraic) approximations of the Hamiltonian tensor (e.g., CD, CP, *global* DF, *algebraic* FMM[119, 120]), 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*[*] approximation of tensor networks, in which the leading-order error due to the approximation of the network constituents is cancelled. Here, we demonstrate the utility of the idea by constructing a robust CP (rCP) approximation for a simple network of two order-3 tensors obtained by the DF-factorization of the 2-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 3 atoms, as demonstrated for the particle-particle ladder (PPL) diagram in the coupled cluster method with single and double excitations (CCSD).

[*]In this work, the term "robust" mirrors its use in the discussion of fitting in quantum chemistry[121] rather than referring to the robust approximation of individual tensors.[122]

The rest of manuscript is organized as follows. In Section 2.2 of this paper we introduce the idea of robust approximation of tensor networks, use it to construct an efficient algebraic approximation to a 2-particle interaction tensor, and discuss how to utilize the proposed factorization to evaluate the particle-particle ladder (PPL) diagram with reduced complexity. Section 2.3 describes the details of the computational experiments. Section 2.4 compares the performances of non-robust and robust approximations applied to the CCSD PPL diagram using standard benchmark sets of noncovalent interaction energies and reaction energies. Section 2.5 summarizes our findings and discusses other possible applications of the idea.

2.2 Formalism

2.2.1 Robust approximation of tensor networks

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

$$\delta_i \equiv \mathcal{T}_i - \hat{\mathcal{T}}_i, \tag{2.1}$$

is "small", i.e., $||\delta_i|| = \mathcal{O}(\epsilon) \ni \epsilon < 1$, the tensor network can be accurately represented in terms of tensor approximants by including terms linear in the error:

$$\{\mathcal{T}_1 \dots \mathcal{T}_k\} = \{\hat{\mathcal{T}}_1 \dots \hat{\mathcal{T}}_k\} + \sum_j \{\hat{\mathcal{T}}_1 \dots \hat{\mathcal{T}}_{j-1} \delta_j \hat{\mathcal{T}}_{j+1} \dots \hat{\mathcal{T}}_k\} + \mathcal{O}(\epsilon^2). \tag{2.2}$$

Note that the *naive* approximation of the network, given by the first term on the right-hand side, is only accurate to $\mathcal{O}(\epsilon)$. A *robust* approximation, accurate to $\mathcal{O}(\epsilon^2)$, is obtained by plugging Eq. (2.1) into Eq. (2.2):

$$\{\mathcal{T}_1 \dots \mathcal{T}_k\} = (1-k)\{\hat{\mathcal{T}}_1 \dots \hat{\mathcal{T}}_k\} + \sum_j \{\hat{\mathcal{T}}_1 \dots \hat{\mathcal{T}}_{j-1} \mathcal{T}_j \hat{\mathcal{T}}_{j+1} \dots \hat{\mathcal{T}}_k\} + \mathcal{O}(\epsilon^2). \tag{2.3}$$

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

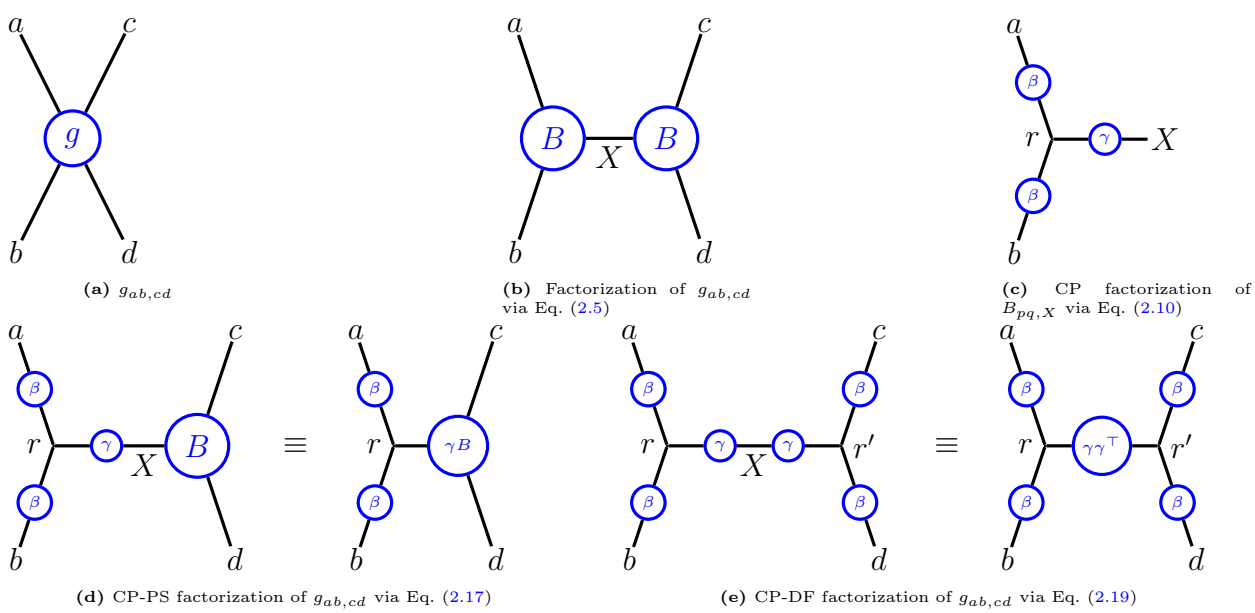
In the context of numerical tensor approximations, the robust approximation has enjoyed a long use by the electronic structure community.[113, 121, 123, 124] Despite its simplicity and/or apparent lack of novelty, in the context of tensor computation the idea has potentially significant unexplored utility. Its utility came as a real surprise to us when we stumbled on its novel application, described below.

2.2.2 Robust approximation of factorized 2-particle interaction tensor

Consider a tensor representation of a 2-particle interaction[†] in a generic basis of size n:

$$g_{ab,cd} \equiv \iint \phi_a^*(\mathbf{r}_1)\phi_b(\mathbf{r}_1)g(\mathbf{r}_1,\mathbf{r}_2)\phi_c^*(\mathbf{r}_2)\phi_d(\mathbf{r}_2)\,d\mathbf{r}_1\,d\mathbf{r}_2. \tag{2.4}$$

The comma separator between indices defines the default *matricization*; namely, matrix $\mathbf{O}$ will refer to the matricized form of tensor O, with element $O_{p_1p_2\ldots,q_1q_2\ldots}$ located in row $p_1p_2\ldots$ and column $q_1q_2\ldots$ of the matrix. It is also useful to convey tensor expressions diagrammatically; in Penrose notation tensor g is represented as a single node (Figure 2.1a).



(a) $g_{ab,cd}$

(b) Factorization of $g_{ab,cd}$ via Eq. (2.5)

(c) CP factorization of $B_{pq,X}$ via Eq. (2.10)

(d) CP-PS factorization of $g_{ab,cd}$ via Eq. (2.17)

(e) CP-DF factorization of $g_{ab,cd}$ via Eq. (2.19)

Figure 2.1: Graphical representation of the 2-particle interaction tensor (Eq. (2.4)) and factorizations thereof considered in this work.

To efficiently approximate g, it is important to retain the analytic properties, such as symme-

[†]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 non-multiplicative kernels is straightforward.

tries 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 2-particle and 1-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,

$$g_{ab,cd} \approx \sum_X B_{ab,X} B_{cd,X}, \tag{2.5}$$

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

$$\mathbf{g} \approx \mathbf{B}\mathbf{B}^\top. \tag{2.6}$$

Such "generalized square root" factorization is not unique. One way to compute the factorization efficiently is by a (rank-revealing) Cholesky decomposition (CD);[93] for any finite precision the CD rank (i.e., the number of columns of $\mathbf{B}$) is $\mathcal{O}(n)$. Another way to compute this symmetric factorization is via DF, where

$$B_{ab,X} = C_{ab,Y} \left(\mathbf{G}^{1/2}\right)_{Y,X}, \tag{2.7}$$

the fitting coefficients $C_{pq,Y}$ are determined by weighted least-squares fitting,[79–83] typically, using the Coulomb "metric":

$$(\mathbf{G})_{X,Y} \equiv \iint \phi_X(1) g(1,2) \phi_Y(2)\, d1\, d2, \tag{2.8}$$

and the square root of $\mathbf{G}$ is defined by Eq. (2.6), 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, in 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 $\mathbf{B}$:

$$B_{ab,X} \approx \sum_{r}^{R} \beta_{a,r}\kappa_{b,r}\gamma_{X,r} \tag{2.9}$$

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

$$B_{ab,X} \approx \hat{B}_{ab,X} \equiv \sum_{r}^{R} \beta_{a,r}\beta_{b,r}\gamma_{X,r} \tag{2.10}$$

It is well known[56, 57] 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.[58–60]

Tensor factorization of Coulomb interaction Eq. (2.4) 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.[61] His work led to the pseudospectral family

of methods[61, 62, 64, 84, 86–90] which approximate Coulomb integrals using a numerical quadrature over one electron. This quadrature approximation is also employed in the COSX method[112–115, 117, 118, 125] and in the approximation of many-electron integrals in explicitly correlated F12 methods.[126]

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:

$$g_{ab,cd} \overset{\text{PS}}{\approx} \sum_g w_g \phi_a^*(\mathbf{r}_g)\phi_b(\mathbf{r}_g) \int g(\mathbf{r}_g, \mathbf{r}_2)\phi_c^*(\mathbf{r}_2)\phi_d(\mathbf{r}_2)\, d\mathbf{r}_2 \tag{2.11}$$

where w_g is the weight of the g quadrature point; Introducing

$$X_{a,g} \equiv \sqrt{w_g}\phi_a(\mathbf{r}_g), \tag{2.12}$$

$$Y_{g,cd} \equiv \int g(\mathbf{r}_g, \mathbf{r}_2)\phi_c^*(\mathbf{r}_2)\phi_d(\mathbf{r}_2)\, d\mathbf{r}_2, \tag{2.13}$$

leads to the *algebraic* form of the PS approximation,

$$g_{ab,cd} \overset{\text{PS}}{\approx} \sum_g X_{a,g}^* X_{b,g} Y_{g,cd}, \tag{2.14}$$

which makes the connection to CP factorization obvious; note that the summation over grid points g corresponds to the 3-way *hyperedge* in the diagrammatic representation of Eq. (2.14) in Figure 2.1d. In practice, an accurate implementation of the PS approximation is sensitive to choice of grid and requires various measures to reduce the error.[65, 68, 113, 115, 118] 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 (2.11), but by arbitrary fitness conditions.

Inserting a quadrature once for *every* particle leads to, what Martinez and co-workers termed, the tensor hypercontraction[‡] (THC) approximation[66–71, 73, 97] of $g_{ab,cd}$,

$$g_{ab,cd} \overset{\text{THC}}{\approx} \sum_{g_1,g_2} w_{g_1} w_{g_2} \phi_a^*(\mathbf{r}_{g_1}) \phi_b(\mathbf{r}_{g_1}) g(\mathbf{r}_{g_1}, \mathbf{r}_{g_2}) \phi_c^*(\mathbf{r}_{g_2}) \phi_d(\mathbf{r}_{g_2}), \tag{2.15}$$

and its algebraic form:

$$g_{ab,cd} \overset{\text{THC}}{\approx} \sum_{g_1} \sum_{g_2} X^*_{a,g_1} X_{b,g_1} Y_{g_1,g_2} X^*_{c,g_2} X_{d,g_2}. \tag{2.16}$$

The diagrammatic representation of Eq. (2.16), shown in Figure 2.1e, includes *two* 3-way hyperedges. Clearly, the same idea can be applied to a matrix element of any (local) n-body operator.[69] THC approximation was originally exploited in the algebraic form, using algebraic CP decomposition of 3-center overlap integrals in the context of (non-robust) overlap-metric DF to define factors X and Y in Eq. (2.16) ("PF-THC").[66] It was subsequently formulated using real-space quadrature to define factors X and least-squares fitting to determine factor Y in Eq. (2.16) ("LS-THC")[68, 71, 99]. What these approaches have in common with each other and with other related factorizations[105] is use of the tensor network topology of Eq. (2.16); how the factors are determined can differ widely between the methods.

Although our focus in this manuscript is on the 3-way CP factorization (CP3) we should also

[‡]The term "hypercontraction" presumably refers to the appearance of *hyperedges* in the diagrammatic representation of CP-like tensor networks, e.g., Figure 2.1d.

note that the direct 4-way algebraic CP factorization of Coulomb integrals (CP4) has been employed by Benedikt and co-workers.[102–104] Related 4-way factorizations of Coulomb integrals has been considered by Peng and Kowalski, who proposed to compress the Cholesky factors of the Coulomb tensor by the SVD; the use of factorized integrals has been explored in the CC method.[127] More recently, Motta and co-workers employed a similar multi-step factorization to reduce the cost of auxiliary-field Quantum Monte Carlo methods.[128] The similarity of these factorizations to the 4-way CP decomposition is due to the appearance of the 4-way hyperedge, whereas all of the factorizations considered in this work are limited to 3-way hyperedges only.

To introduce the main result of our work consider how to best introduce the CP3 approximation (Eq. (2.10)) for the symmetric (CD/DF-like) factorization in Eq. (2.5). Using CP3 *once* produces a PS-like factorization, to which we will refer as CP-PS:

$$g_{ab,cd} \overset{\text{CP-PS}}{\approx} \sum_{X} \sum_{r}^{R} \beta_{a,r} \beta_{b,r} \gamma_{X,r} B_{cd,X} = \sum_{r}^{R} \beta_{a,r} \beta_{b,r} (\gamma B)_{cd,r}, \tag{2.17}$$

where we introduced

$$(\gamma B)_{cd,r} \equiv \sum_{X} \gamma_{X,r} B_{cd,X}; \tag{2.18}$$

compare Eq. (2.17) to Eq. (2.14) 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:

$$g_{ab,cd} \overset{\text{CP-DF}}{\approx} \sum_{X} \sum_{r}^{R} \beta_{a,r} \beta_{b,r} \gamma_{X,r} \sum_{r'}^{R} \beta_{c,r'} \beta_{d,r'} \gamma_{X,r'} = \sum_{r}^{R} \beta_{a,r} \beta_{b,r} \sum_{r'}^{R} \beta_{c,r'} \beta_{d,r'} (\gamma \gamma^{\top})_{r,r'}, \tag{2.19}$$

where we introduced $(\gamma\gamma^\top)_{r,r'} \equiv \sum_X \gamma_{X,r}\gamma_{X,r'}$; compare Eq. (2.19) to Eq. (2.16) 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. (2.10)). As discussed in Section 2.2.1, it is possible to eliminate the linear error using the *robust* form of CP-DF, to which we will refer as rCP-DF:

$$g_{ab,cd} \overset{\text{rCP-DF}}{\approx} 2g^{\text{CP-PS}}_{ab,cd} - g^{\text{CP-DF}}_{ab,cd} = \sum_r^R \beta_{a,r}\beta_{b,r}\left(2(\gamma B)_{cd,r} - \sum_{r'}^R \beta_{c,r'}\beta_{d,r'}(\gamma\gamma^\top)_{r,r'}\right). \tag{2.20}$$

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 ($\mathcal{O}(N^5)$) 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.

2.2.3 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[86, 90] and THC factorizations[71, 105] can reduce the computational complexity of the PPL term in the canonical MO basis from $\mathcal{O}(N^6)$ to $\mathcal{O}(N^5)$, hence the same should be possible for the PPL term in the rCP-DF approximation. Indeed, plugging in Eq. (2.17) into the spin-free

PPL expression (permutational symmetry is ignored for simplicity) yields:

$$\sum_{bd} g_{ab,cd} t_{bdij} \overset{\text{CP-PS}}{\approx} \text{PPL}^{\text{CP-PS}} \equiv \sum_{r}^{R} \beta_{a,r} \left(\sum_{d} (\gamma B)_{cd,r} \left(\sum_{b} \beta_{b,r} t_{bdij} \right) \right). \tag{2.21}$$

The order of evaluation which 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, $\sum_b \beta_{b,r} t_{bdij} \rightarrow (I_1)_{rdij}$, 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 2 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. (2.18)) 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 $R < u^2/3$.[§]

The PPL term can be similarly reformulated with the $\mathcal{O}(N^5)$ cost using the CP-DF approximation. One approach, utilized by Parrish et al.[71] and Hummel et al.[105], uses the CP-PS route (Eq. (2.21)) by recomputing the appropriate intermediates:

$$\sum_{bd} g_{ab,cd} t_{bdij} \overset{\text{CP-DF}}{\approx} \text{PPL}^{\text{CP-DF}} \equiv \sum_{r}^{R} \beta_{a,r} \left(\sum_{d} (\gamma \hat{B})_{cd,r} \left(\sum_{b} \beta_{b,r} t_{bdij} \right) \right), \tag{2.22}$$

[§]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.

where $(\gamma\hat{B})_{cd,r}$ is the CP-factorized intermediate $(\gamma B)_{cd,r}$, obtained by inserting Eq. (2.10) into Eq. (2.18)[¶]:

$$(\gamma\hat{B})_{cd,r} \equiv \sum_{X} \gamma_{X,r} \left(\sum_{r'}^{R} \beta_{a,r'} \beta_{b,r'} \gamma_{X,r'} \right). \tag{2.23}$$

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.[105] and Mardirossian et al.[129] introduces order-4 tensors with 2 CP indices:

$$\sum_{bd} g_{ab,cd} t_{bdij} \overset{\text{CP-DF}}{\approx} \text{PPL}^{\text{CP-DF}} \equiv \sum_{r}^{R} \beta_{a,r} \sum_{r'}^{R} \left(\beta_{c,r'} \left((\gamma\gamma^{\top})_{r,r'} \left(\sum_{b} \beta_{b,r} \left(\sum_{d} \beta_{d,r'} t_{bdij} \right) \right) \right) \right). \tag{2.24}$$

Compared to 3 tensor products in the CP-PS approach, the CP-DF route has 5 products, with all but the third product of $(\gamma\gamma^{\top})$ 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. (2.24) lowers the computational efficiency of this approach. For these reasons, throughout our work

[¶]N.B. if $5X > 3R$, Eq. (2.23) can be reordered to compute $(\gamma\hat{B})$ more efficiently

we used the CP-PS-based approach, Eq. (2.22), to implement CP-DF PPL.

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

$$\sum_{bd} g_{ab,cd} t_{bdij} \overset{\text{rCP-DF}}{\approx} 2 \times \text{PPL}^{\text{CP-PS}} - \text{PPL}^{\text{CP-DF}}. \tag{2.25}$$

Plugging Eq. (2.21) and Eq. (2.22) into Eq. (2.25) and refactoring leads to the following evaluation scheme with optimal operation count:

$$\sum_{bd} g_{ab,cd} t_{bdij} \overset{\text{rCP-DF}}{\approx} \text{PPL}^{\text{rCP-DF}} \equiv \sum_{r}^{R} \beta_{a,r} \left(\sum_{d} (\gamma \tilde{B})_{cd,r} \left(\sum_{b} \beta_{b,r} t_{bdij} \right) \right), \tag{2.26}$$

in which we introduced

$$(\gamma \tilde{B})_{cd,r} \equiv 2(\gamma B)_{cd,r} - (\gamma \hat{B})_{cd,r} \tag{2.27}$$

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 3-way CP approach in the context of the PPL evaluation.

2.3 Computational details

CP approximations for order-3 tensors were computed using the standard alternating least squares (ALS) method.[130, 131] Although ALS can be slow to converge and the quality of

the solution can strongly depend on the initial guess,[132] we found that our solver converged robustly with an initial guess of vectors generated using quasi-random numbers taken from the uniform distribution on [-1,1]. No consistent benefit was found from an initial guess scheme which generated factor matrices using the higher-order SVD (HOSVD)[44] 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.[133] The use of non-linear and gradient-based solvers[58, 59] 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[134] as well as a 12-system representative set of 12 complexes (S66/12)[‖]; some computations utilized a smaller 7-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).[135] Additional assessments utilized the HJO12 set of isogyric reaction energies,[136, 137] the 8 low-lying conformers of $(H_2O)_6$[138] and a conformer of $(H_2O)_{20}$.[139, 140] All of the above computations utilized the cc-pVDZ-F12 (abbreviated as DZ-F12) orbital basis set (OBS)[141]. The 2-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).[142] Assessment of the basis set variation in the performance of rCP-DF used the following additional OBS/DFBS pairs: the aug-cc-pVDZ[143, 144] (aVDZ) OBS paired with the aVDZ-RI DFBS, the aug-cc-pVTZ (aVTZ) OBS[143, 144] paired with the aug-cc-pVTZ-RI[142] (aVTZ-RI) DFBS, and the cc-

[‖] 1 water ... water, 2 water ... MeOH, 3 water ... $MeNH_2$, 4 $MeNH_2$ ... MeOH, 5 benzene ... benzene (π-π), 6 pyridine ... pyridine (π-π), 7 uracil ... uracil (π-π), 8 pentane ... pentane, 9 benzene ... benzene (TS), 10 benzene ... ethyne (CH-π), 11 ethyne ... water (CH-O), 12 $MeNH_2$ ... pyridine

pVTZ-F12[141] (TZ-F12) OBS paired with the aVTZ-RI DFBS. The CP approximations of Coulomb integral tensors was 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 2 Intel Xeon E5-2683 v4 CPUs, and high-memory nodes, each with 4 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

$$\text{speedup} = \frac{t_{\text{DF-CCSD}}}{t_{\text{CP-PPL-DF-CCSD}} + t_{\text{CP-ALS}}} \tag{2.28}$$

where $t_{\text{DF-CCSD}}$ and $t_{\text{CP-PPL-DF-CCSD}}$ 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 $t_{\text{CP-ALS}}$ 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 Subroutines (BTAS) library.[145] The CP-DF, CP-PS and rCP-DF approximations are implemented in a developmental version of the Massively Parallel Quantum Chemistry (MPQC) package.[146]

2.4 Results

The discussion of computational experiments is organized as follows. In Section 2.4.1 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 2 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 2.4.2 and 2.4.3 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.

2.4.1 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 2.2 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 notation introduced in Eq. (2.6), it is clear that the leading-order

error of the CP-DF factorization should be roughly twice the error of CP-PS:

$$\mathbf{g}^{\text{DF}} - \mathbf{g}^{\text{CP-PS}} = \mathbf{B}\mathbf{B}^\top - \frac{1}{2}\left(\hat{\mathbf{B}}\mathbf{B}^\top + \mathbf{B}\hat{\mathbf{B}}^\top\right) = \frac{1}{2}\left(\boldsymbol{\delta}\mathbf{B}^\top + \mathbf{B}\boldsymbol{\delta}^\top\right), \tag{2.29}$$

$$\mathbf{g}^{\text{DF}} - \mathbf{g}^{\text{CP-DF}} = \mathbf{B}\mathbf{B}^\top - \hat{\mathbf{B}}\hat{\mathbf{B}}^\top = \boldsymbol{\delta}\hat{\mathbf{B}}^\top + \hat{\mathbf{B}}\boldsymbol{\delta}^\top + \boldsymbol{\delta}\boldsymbol{\delta}^\top = 2\left(\mathbf{g}^{\text{DF}} - \mathbf{g}^{\text{CP-PS}}\right) + \boldsymbol{\delta}\boldsymbol{\delta}^\top, \tag{2.30}$$

where $\hat{\mathbf{B}}$ is the matricized form of the CP approximant in Eq. (2.10), and

$$\boldsymbol{\delta} \equiv \mathbf{B} - \hat{\mathbf{B}} \tag{2.31}$$

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 one 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.,[68, 71] and full relaxation of the CP-DF network cost was implemented by Schutski et al.[73] (e.g., see the discussion of their THC-ALS-RI solver). To investigate whether the subopti-

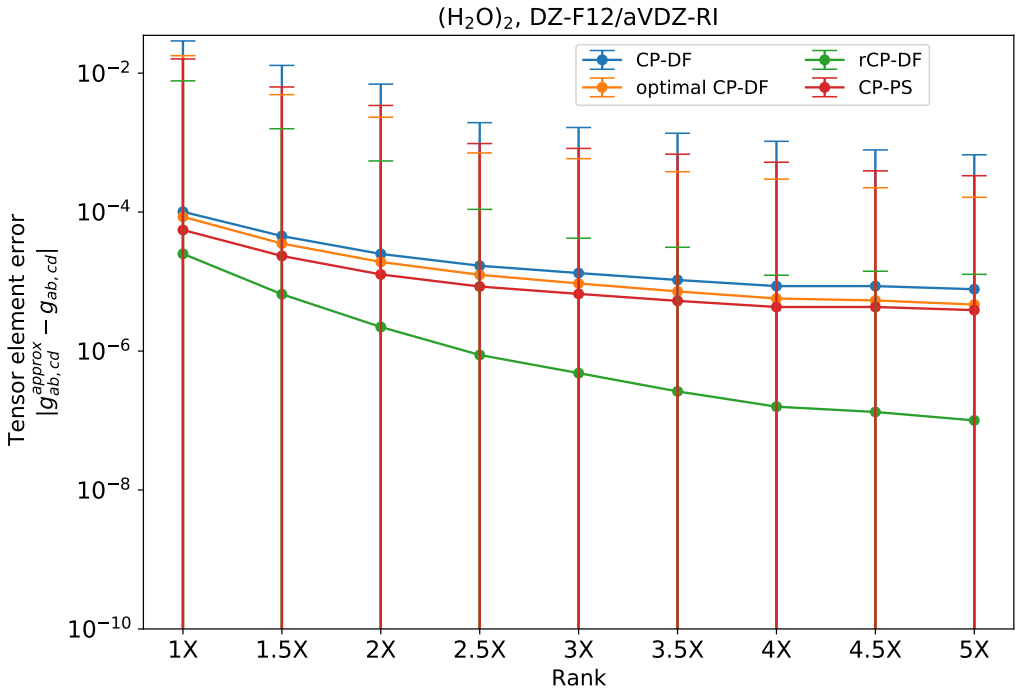


Figure 2.2: Absolute errors in matrix elements of $g_{ab,cd}$ for a water dimer with S66 configuration approximated by the CP-PS, CP-DF, and rCP-DF factorizations obtained with ALS precision of $\epsilon = 10^{-3}$. The error bars denote the max/min unsigned errors.

mality of the CP-DF network using the **B**-optimized factors is significant we implemented an ALS solver that minimizes the CP-DF error in $\mathbf{g}^{\mathrm{DF}}$;** the operation complexity of such solver is identical to the $\mathcal{O}(N^4)$ complexity of the ALS solver for the CP decomposition of **B**, albeit the prefactor is somewhat larger. Only few iterations are needed to relax the CP-DF network fully with respect to **g** if we use, as the initial guess, the factors obtained by CP3 decomposing **B**.

As the data in Sections 2.4.2 and 2.4.3 indicates, the tensor element errors obtained with the **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

**See the Supporting Information for the detailed algorithm description.

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 $\mathbf{g}^{\mathrm{DF}}$) of the factors in the CP-DF network and due to the deficient CP rank.

2.4.2 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. (2.10) as

$$\Delta \equiv 1.0 - \frac{\|B_{ab,X} - \sum_r^R \beta_{a,r}\beta_{b,r}\gamma_{X,r}\|}{\|B_{ab,X}\|} = 1.0 - \frac{\|\boldsymbol{\delta}\|}{\|B_{ab,X}\|} \tag{2.32}$$

where $\boldsymbol{\delta}$ is the CP error tensor as defined in Eq. (2.31). 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.

Variation of the CP error with the ALS solver precision

Figure 2.3 report 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.

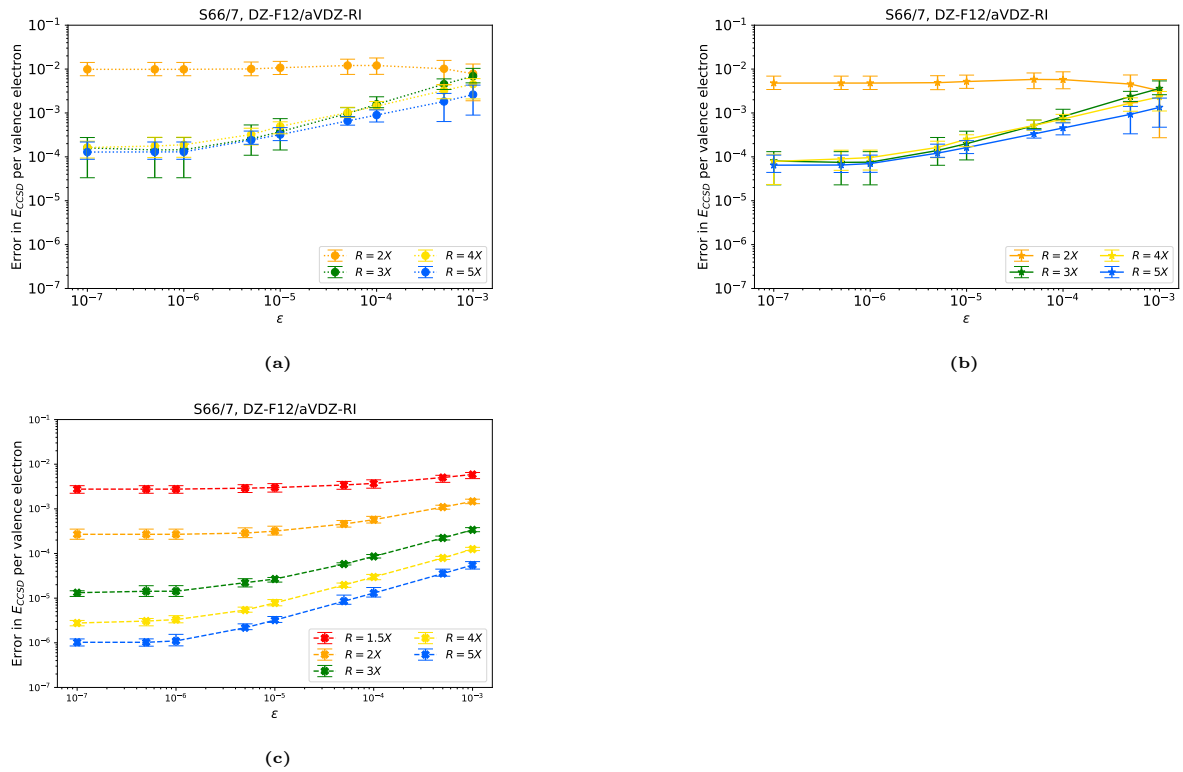
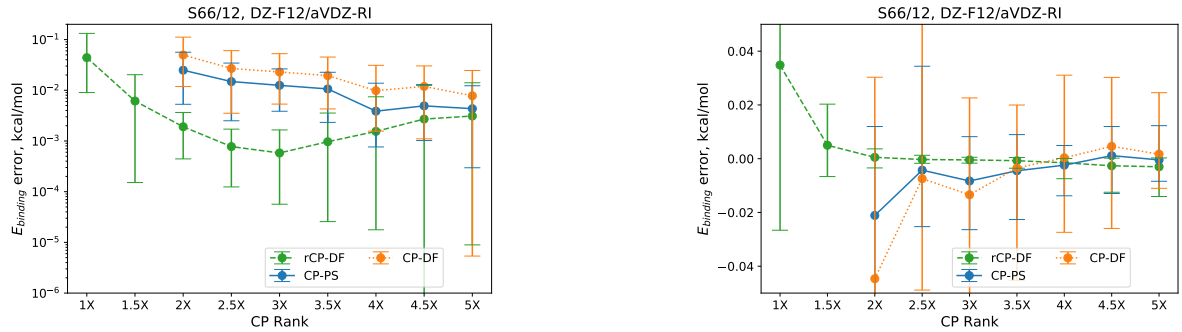


Figure 2.3: Mean unsigned errors in the per-electron CCSD correlation energies (kcal/mol) of molecules in the S66/12 dataset, relative to canonical CCSD, induced by the (a) CP-DF, (b) CP-PS or (c) rCP-DF approximations to PPL vs the ALS precision (ϵ). The error bars denote the max/min unsigned errors.



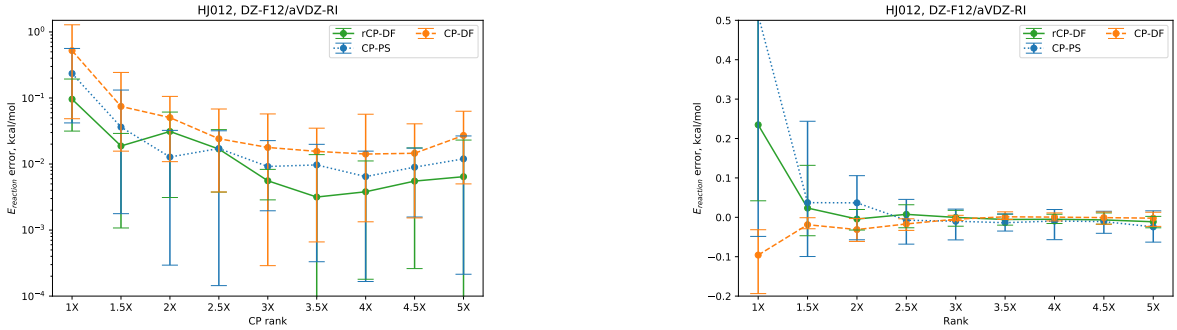
(a) (b)

Figure 2.4: Mean unsigned (a) and signed (b) errors, respectively, in the CCSD binding energies (kcal/mol) of the S66/12 dataset, 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.

Variation of the CP error with the CP rank

Figure 2.3 indicated that increasing the CP rank R reduced the error in the CCSD energy monotonically. These figures also gave the first evidence of performance advantage of rCP-DF over CP-DF and CP-PS. At $R = 1.5X$ (the red line in Figure 2.3c), rCP-DF is more accurate than both CP-DF and CP-PS with $R = 2X$ (the orange line in Figures 2.3a and 2.3b). Furthermore, the error in the CCSD energy is reduced at a fast rate, with respect to CP rank, for rCP-DF, which corroborates our discussion in Section 2.4.1. For each R and at converged ϵ, the rCP-DF approximation introduces error which is at least an order of magnitude smaller than the error introduced by either CP-DF or CP-PS.

Next 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 2.4 and 2.5, respectively. Because, compared to R, ϵ has a relatively small influence on E_{CCSD}, we have limited this assessment to using relatively loose



(a) **(b)**

Figure 2.5: Mean unsigned (a) and signed (b) errors, respectively, in the CCSD reaction energies (kcal/mol) of the HJO12 dataset, 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.

ALS tolerances of $\epsilon = 10^{-3}$.[††] 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 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 2.6.[‡‡] 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.

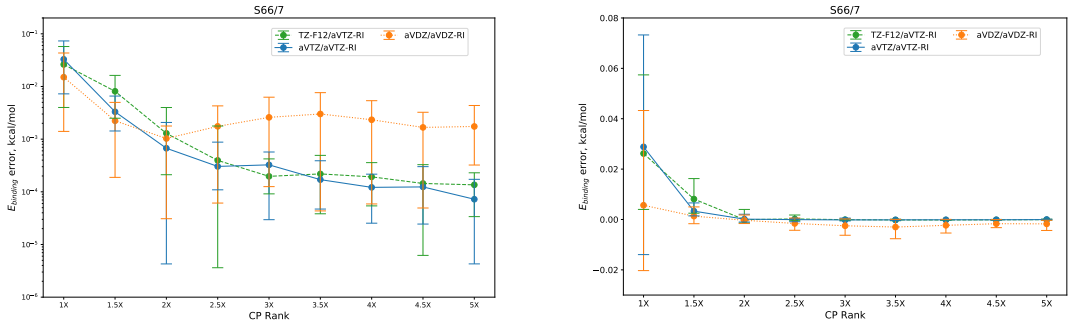
[††]The corresponding results for a tighter ALS tolerance, $\epsilon = 10^{-4}$, are reported in the Supporting Information.

[‡‡]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

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.[71] Table 2.1 juxtaposes the maximum absolute and relative CCSD energy errors due to the rCP-DF and the THC PPL approximations for the 8 low-lying $(H_2O)_6$ conformers. The same OBS/DFBS basis set pair, TZ/TZ-RI, was utilized for all computations. The

	Maximum Absolute Error	Maximum Relative Error
rCP-DF	0.45	0.036
LS-THC(DF)[71]	2.13	0.18
W-LS-THC(DF)[71]	0.29	0.03

Table 2.1: Maximum absolute and relative errors in valence TZ/TZ-RI DF-CCSD correlation energies (mE_h) of 8 low-lying $(H_2O)_6$ clusters.[138] For the rCP-DF approximation CP rank and ALS precision were fixed at $R = 1.3X$ and $\epsilon = 10^{-3}$, respectively.



(a) **(b)**

Figure 2.6: Mean unsigned (a) and signed (b) errors, respectively, in the CCSD binding energies (kcal/mol) for the S66/7 dataset, relative to canonical CCSD, induced by the rCP-DF approximation to PPL vs CP rank R (in units of the fitting basis, X) using 3 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.

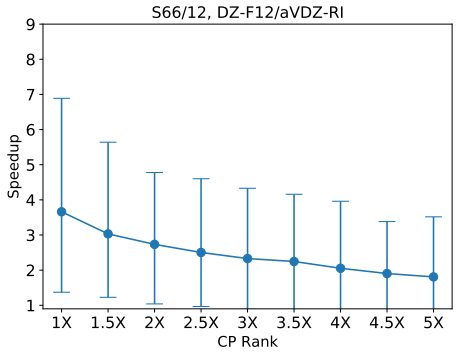
rCP-DF approach used $R = 1.3X$, whereas the corresponding LS-THC grid size corresponds to $R \approx 4X$, i.e., roughly 3 times larger than used by our method. Although the absolute energies are most accurate with the W-LS-THC(DF) method of Parrish et al., the relative energies of the clusters are nearly as accurate with our method, despite its much smaller CP rank. Most importantly, the rCP-DF approach greatly outperforms its true THC counter-

part, 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.[71]

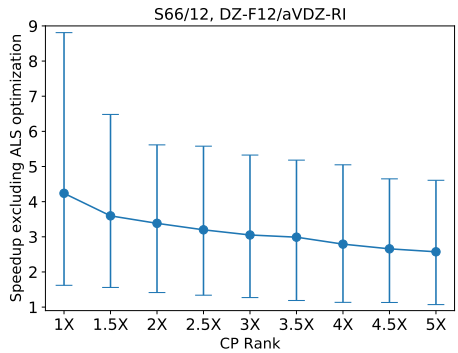
2.4.3 Cost reduction vs DF-CCSD

Next 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.

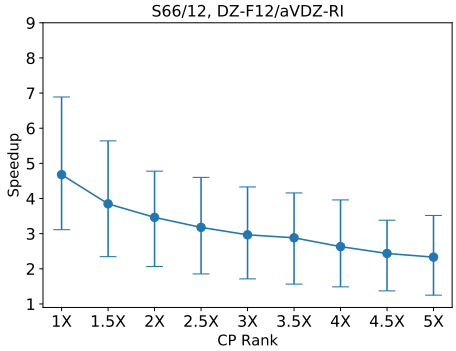
The observed speedups in the DF-CCSD computations due to the CP-based PPL refor-



(a) Average speedup (Eq. (2.28)) 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. (2.28), **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. (2.28)) 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 2.7: Speedup of CP-PPL CCSD

mulations are illustrated for the clusters in the S66/12 test set in Figure 2.7a. Just as in Section 2.4.2, only $\epsilon = 10^{-3}$ are reported in the manuscript, 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 2.7b demonstrates the speedups obtained by excluding the cost of ALS. We see that ALS has the most dramatic effect on cost when ϵ is tighter and R is larger.

Unsurprisingly, ALS optimization had the greatest impact on the smallest molecules. Figure 2.7c demonstrates that the speedup for the 7 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 dataset with 3 additional basis set pairs (Figure 2.8). 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.

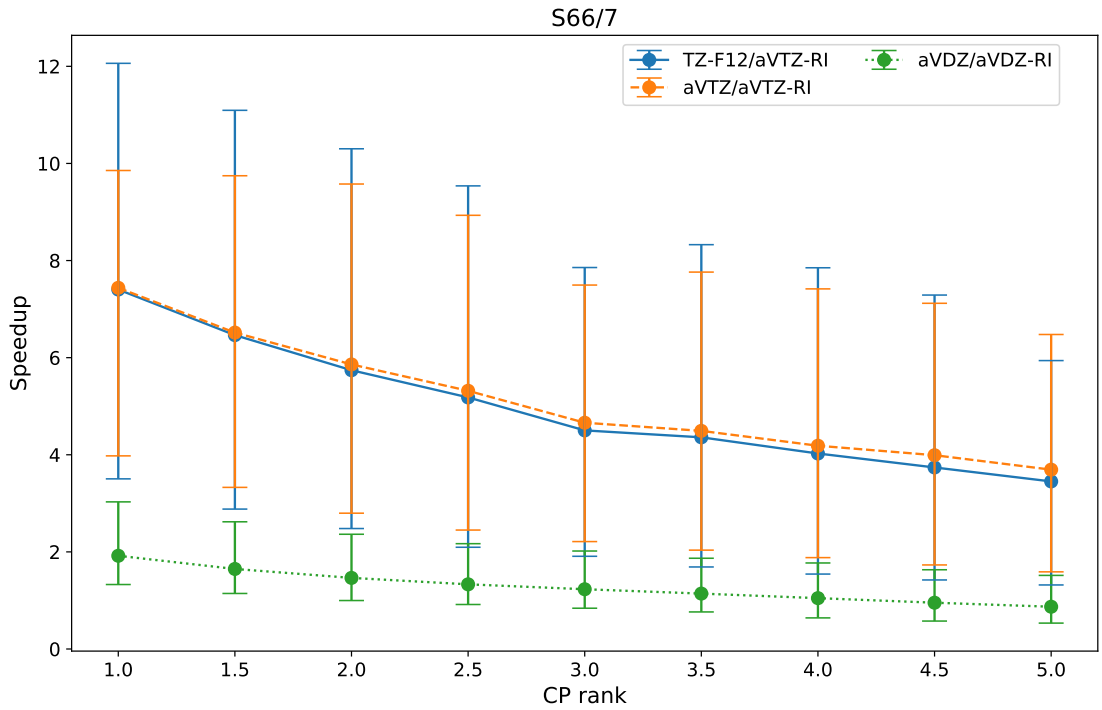


Figure 2.8: Average speedup (Eq. (2.28)) of CCSD with rCP-DF-approximated PPL vs CP rank R (in units of the fitting basis, X) for the S66/7 dataset. ALS precision fixed at $\epsilon = 10^{-3}$. The error bars denote the max/min speedup

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 2.9 and 2.10. For all systems, the errors introduced by

	E_{CCSD}	D_e	t_{CCSD}	t_{PPL}	$t_{\text{CP-ALS}}$
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	

Table 2.2: 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 $(\text{H}_2\text{O})_{20}$ cluster. The total time of the CP ALS optimization is also reported ($t_{\text{CP-ALS}}$, s). CP rank and ALS precision are fixed at $R = 1.3X$ and $\epsilon = 10^{-3}$, respectively.

rCP-DF are significantly less than 0.1 kcal/mol, and the computational savings are realized for all systems, with 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

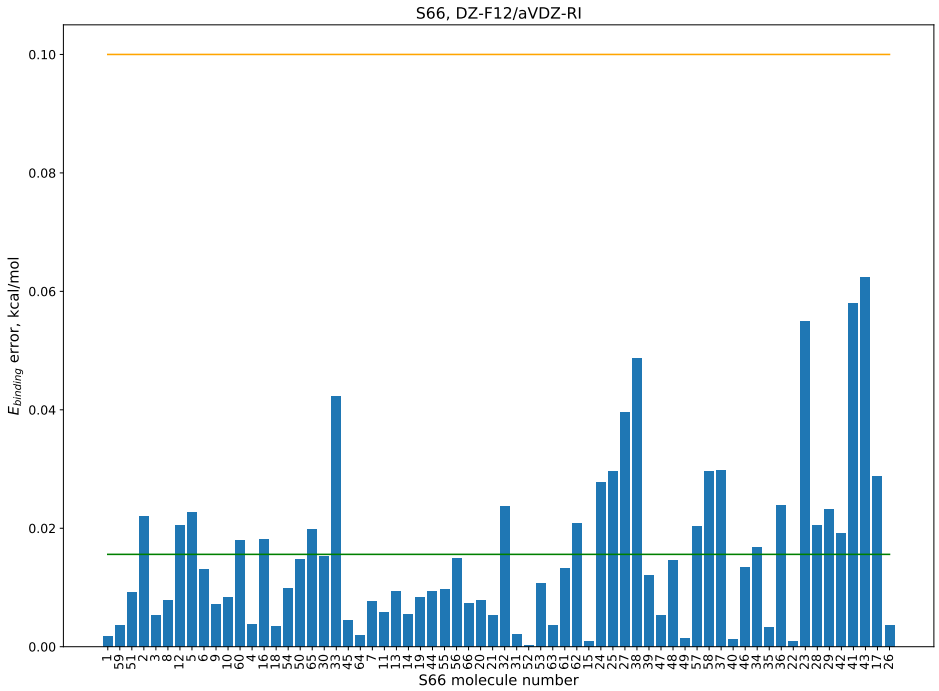


Figure 2.9: Unsigned errors in the S66 CCSD binding energies (kcal/mol), relative to canonical CCSD, induced by the rCP-DF approximation to PPL. CP rank and ALS precision are fixed at $R = 1.3X$ and $\epsilon = 10^{-3}$, respectively. Molecules ordered from smallest to largest number of occupied orbitals. The orange line is the target maximum error, 0.1 kcal/mol, and the green line is the average error of the set.

optimization: as we increase the systems 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 2.11. 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 Section 2.4.3. 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.

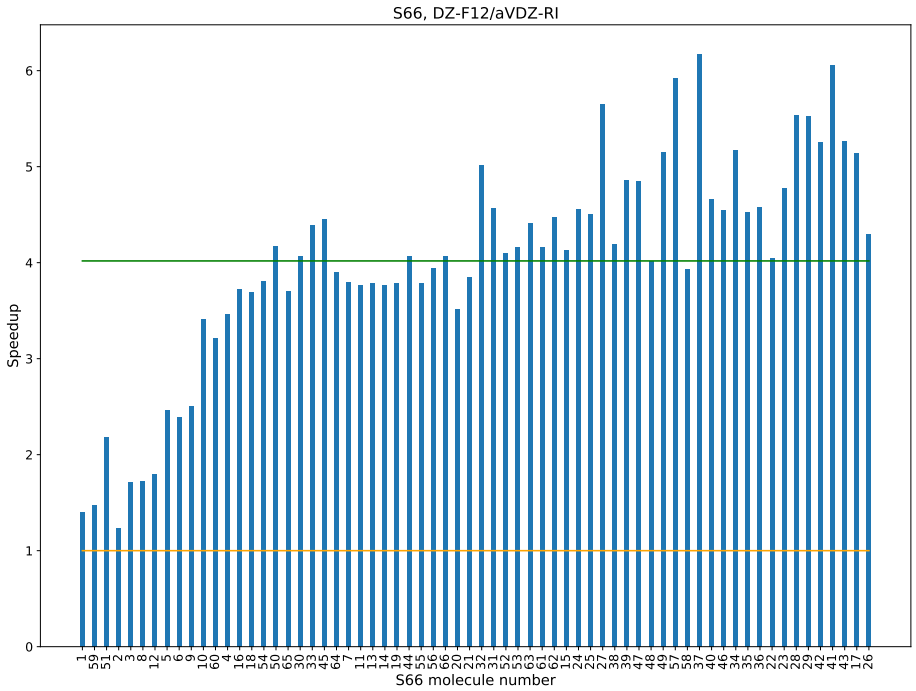


Figure 2.10: Speedup (Eq. (2.28)) of CCSD with rCP-DF-approximated PPL for the entire S66 dataset. CP rank and ALS precision are fixed at $R = 1.3X$ and $\epsilon = 10^{-3}$, respectively. Molecules are ordered according to the number of occupied orbitals, from smallest to largest. The orange line represents no speedup over CCSD and the green line is average speedup of the set.

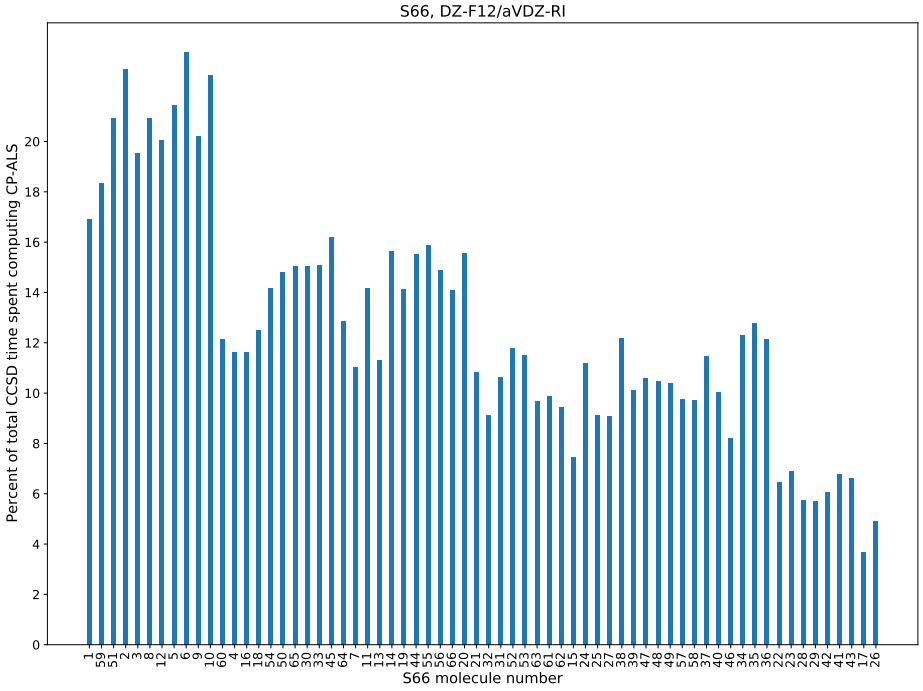


Figure 2.11: Percent of the total CCSD time spent in ALS for each cluster molecule in S66 dataset using rCP-DF with CP rank $R = 1.3X$ and ALS precision of $\epsilon = 10^{-3}$. Molecules are ordered according to the number of occupied orbitals, from smallest to largest.

2.5 Summary and perspective

In this work, we considered how *robust* (in the Dunlap sense[121]) approximation of tensor networks, in which the leading-order error due to the approximation of the network constituents is explicitly cancelled, can be used profitably to construct efficient factorizations of the 2-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. Single use of the CP decomposition leads to a tensor network resembling the factorization in the well-known pseu-

dospectral (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, resulting in a 1 to 2 order reduction of the error over either naive substitution scheme. Deeper analysis of the errors in the Coulomb interaction tensor revealed 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 $\mathcal{O}(N^6)$ to $\mathcal{O}(N^5)$. 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 3 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 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,[147] CIM,[148] DEC,[149, 150] and others[151], 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 2.4.2, albeit limited, suggests that grid size reductions of a factor of 3 or more are realistic). Robust factorization should also simplify formulation of analytic gradients.

- A combination with other ideas such as the use of orbital-biasing explored in LS-THC-based coupled-cluster[71] 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.

2.6 Acknowledgement

This work was supported by the U.S. National Science Foundation (awards 1550456 and 1800348). We also acknowledge Advanced Research Computing at Virginia Tech (www.arc.vt.edu) for providing computational resources and technical support that have contributed to the results reported within this paper.

Supporting Information

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

Chapter 3

Effective use of 4-way canonical polyadic decomposition for accelerating the coupled-cluster perturbative triples.

3.1 Introduction

The canonical polyadic (CP) decomposition[100, 101] is a powerful tool used to construct mathematically sparse representations of higher-order tensor and to eliminate the curse of dimensionality from computational modeling. In electronic structure methods, introduction of the algebraic CP decomposition[5, 102–107] is fairly limited for a number of reasons: there is no finite algorithm to discover the rank of a CP decomposition[56, 57], optimization of CP factor matrices requires solving a set of nonlinear equations, and CP based algorithms require the implementation of non-standard tensor-algebra kernels. To solve some of the

issues related to construction of CP represented tensors, researchers have developed grid-based initialization schemes such as the pseudospectral[61, 62, 64, 84–90] (PS) approach, and tensor-hypercontraction[66–68, 70, 71, 73, 97–99] (THC). Both the PS and THC approaches construct sparse representations for the order-four tensors using order-3 CP (CP3) decompositions. However, electronic structure methods routinely require tensors of order-four and higher. Additionally, depending on the approximated network topology, the CP3 approximation may fail to reduce the complexity of high-scaling contractions. In the cases where the CP3 based approximations are ineffective, it can be beneficial to construct accurate CP approximations to order-four tensors. In this work, we develop a reduced-scaling approach to construct high-accuracy, order-four CP decompositions of the Coulomb integral tensor by substituting the density fitting (DF) approximation into the CP loss function. We demonstrate the efficiency and accuracy of the approach by introducing order-3 and order-4 CP approximated integral tensors into the Laplace-transformed (LT) triples correction to the coupled cluster method with single and double substitutions [CCSD(T)].

CCSD(T)[13] plays a key role as the most economical route to approaching the chemical accuracy for chemical energies and other properties, and thereby surpasses the accuracy of mainstream density-functional approximations by significant margin. However, CCSD(T), in its naive formulation, is restricted to relatively small chemical systems due to its $\mathcal{O}(N^7)$ operation complexity. Therefore, much recent effort has been focused on reducing the computational operations and time-to-solution of CCSD(T). The former can be achieved by exploiting massively parallel HPC resources[152–167], including various types of accelerated hardware (such as general-purpose graphical processing units)[153, 168–171]. It is possible to

reduce the complexity of CCSD(T) by exploiting numerical sparsity of the CCSD amplitudes and Hamiltonian tensors. This sparsity can be reveled using a localized representation such as localized occupied and unoccupied orbitals or even atomic orbitals[166, 172–190]. However, because the zeroth-order Møller-Plesset Hamiltonian used in the CCSD(T) method is not diagonal in localized representations, these approaches usually involve a low-rank representation of the zeroth-order Hamiltonian inverse tensor via a Laplace transformation[78] (LT) or a Cholesky decomposition[191]. Furthermore, fragment-based approaches avoid some technical complexity by using pre-defined partitioning of an entire molecule into polyatomic fragments.[148, 192–197]

Here, we introduce CP approximated Hamiltonian tensors into the LT CCSD(T) formulation introduced by Constans and co-workers[78]. To demonstrate the utility of the order-3 and order-4 CP approximations we reduce the complexity of 9, rate-limiting contractions from $\mathcal{O}(N^6)$ to $\mathcal{O}(N^5)$. The error of CP approximated tensor networks is made small by explicitly removing leading order error using the robust tensor network approximation, which we developed earlier this year[5]. The CP decomposition is not a physics-based or orbital-specific approximation therefore it is possible to combine the CP approximation with coarse-grain fragmented CCSD(T) approaches, such as the LNO CCSD(T) method presented by Kállay, Nagy, et al[166, 177, 182, 183]. Because such extensions are non-trivial it will not be demonstrated here. However, we will discuss, in detail, how the robust CP decomposition can be deployed in such context. Large-scale parallelization of the CP approximated LT CCSD(T) approach, developed here, is facilitated by its implementation using the TiledArray massively parallel-distributed tensor framework[1, 109, 198].

The rest of this manuscript is organized as follows. In Section 3.2 we introduce the DF and CP decompositions and demonstrate how one can construct the CP4 approximations in $\mathcal{O}(N^4)$ time. Furthermore, we introduce LT CCSD(T) and show how we reduce the complexity of 9 rate-limiting $\mathcal{O}(N^6)$ contractions using CP3 and CP4. In Section 3.3 we discuss the details of our computational experiments. Section 3.4 reports on the accuracy of the rCP-DF-LT CCSD(T) approach and compares the computational performance of rCP-DF-LT CCSD(T) to the DF-LT CCSD(T) and the DF CCSD(T) methods. Section 3.5 discusses how the polynomial complexity of the presented approach can be lowered to linear by coarse-graining the LT CCSD(T) energy expression. Finally, Section 3.6 summarizes our findings.

3.2 Formalism

3.2.1 Approximating the Coulomb integral tensor

The most popular factorization of the Coulomb integral tensor is the density fitting[50, 79, 80, 199] (DF) approximation. The DF approximation seeks to factorize the order-4 2-particle interaction tensor, here expressed in a generic basis

$$g_{rs}^{pq} = \iint \phi_p^*(\mathbf{r}_1)\phi_q^*(\mathbf{r}_2)g(\mathbf{r}_1,\mathbf{r}_2)\phi_r(\mathbf{r}_1)\phi_s(\mathbf{r}_2)d\mathbf{r}_1 d\mathbf{r}_2, \tag{3.1}$$

as a product of lower order tensors. In general, this factorization can be constructed as

$$g_{ab,cd} \approx \sum_{XY} C_c^{aY} V_Y^X C_d^{bX}, \tag{3.2}$$

where C's are DF fitting coefficients[79–83] and V is the two-center integral

$$V_{Y,X} = \int d\mathbf{r}_1 d\mathbf{r}_2 \phi_Y^*(\mathbf{r}_1) g(\mathbf{r}_1, \mathbf{r}_2) \phi_X(\mathbf{r}_2). \tag{3.3}$$

In this work we utilize the Poisson kernel: $g(\mathbf{r}_1, \mathbf{r}_2) \equiv |\mathbf{r}_1 - \mathbf{r}_2|^{-1}$, which is a positive-definite kernel. For any positive-definite kernel, Eq. (3.1) can be factorized into the symmetric form

$$g_{ab,cd} \approx \sum_{X} B_c^{aX} B_d^{bX}, \tag{3.4}$$

where $B_{ac,X} = \sum_Y C_c^{aY} [V^{-1/2}]_Y^X$. Additionally, the size of the fitting basis $\{\phi_X\}$, denoted as $X, Y, \ldots$, is proportional to V. It is also possible to construct canonical polyadic (CP) factorizations of the order-4 coulomb integral tensor and the order-3 DF-approximated Coulomb integral tensors, defined in Eq. (3.4).

The CP approximation[100, 101] is a higher-order decomposition that factorizes an order-N tensor into a polyadic sum of rank-one tensors. For example, the CP decomposition of the tensor $\mathcal{T} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is

$$\mathcal{T} = \sum_{r}^{R_{\text{CP}}} \lambda_r a_r^{(1)} \otimes a_r^{(2)} \otimes \cdots \otimes a^{(N)} \quad | \; a_r^{(i)} \in \mathbb{R}^{I_i} \tag{3.5}$$

where the component vectors $a_r^{(i)}$ are unit-normalized and λ_r contains the scaling information

for the rank-one component tensor, $\chi = a_r^{(1)} \otimes a_r^{(2)} \otimes \cdots \otimes a^{(N)}$. The number of rank-one tensors, R_{CP}, necessary to satisfy the equality in Eq. (3.5) is known as the CP rank. Unlike matrix decomposition, determining the CP rank of a non-trivial tensor is a difficult problem[56, 57]. In practice, determining a suitable rank requires optimizing multiple rank-$R < R_{\mathrm{CP}}$ CP approximation and choosing the value of R which satisfies the desired accuracy requirements. Taking the set of R vectors which span $\mathbb{R}^{I_i}$, one can form the ith factor matrix $A^{(i)} = \{a_1^{(i)}, a_2^{(i)}, \ldots, a_R^{(i)}\}$. This factor matrix representation of the CP dceomposition allows one to express the CP factorization of an order-N tensor as a set of N order-2 tensors which are coupled by the hyperdimension R.

We construct the CP decomposition of the order-4 two-electron integral tensor, Eq. (3.1), using the set of factor matrices $\mathbf{x} = \{\gamma, \rho, \beta, \kappa\}$:

$$g_{bc}^{ia} \approx \hat{g}_{bc}^{ia} = \sum_r^{R_{\mathrm{CP4}}} \gamma_r^i \rho_r^a \beta_r^b \kappa_r^c. \tag{3.6}$$

We denote this factorization the CP4 approximation because our target, g, is an order-4 tensor. Later, we show that the size of the CP4 rank is proportional to V. Additionally, we construct CP decompositions of the order-3 DF Coulomb integral tensors

$$B_b^{iX} \approx \hat{B}_b^{iX} = \sum_r^{R_{\mathrm{CP3}}} \gamma_r^i \beta_r^b \chi_r^X \tag{3.7}$$

and

$$B_c^{aX} \approx \hat{B}_c^{aX} = \sum_r^{R_{\mathrm{CP3}}} \rho_r^a \kappa_r^c \chi_r^X, \tag{3.8}$$

which are denoted as CP3 approximations. Because it is known that the quality and efficiency of the the CP optimization processes can strongly depend on the initial guess, we use optimized CP3 factor matrices as inputs to the CP4 optimization. We allow the CP3 and CP4 ranks to differ, so when the $R_{\mathrm{CP4}} > R_{\mathrm{CP3}}$ we pad the CP4 guess with vectors of quasi-random numbers taken from the uniform distribution on [-1,1]. We have found that this padded CP4 initial guess rapidly converges to a high accuracy approximation.

Recently we have shown[5] that the leading-order error associated with approximating tensors in a arbitrary tensor network, $\{T_1 T_2 \ldots T_k\}$ where $k > 1$, can be removed using a robust tensor network formulation. This robust approximation is defined as

$$\{T_1, \ldots, T_k\} \approx (1-k)\{\hat{T}_1, \ldots, \hat{T}_k\} + \sum_j \{\hat{T}_1, \ldots, \hat{T}_{j-1}, T_j, \hat{T}_{j+1}, \ldots, \hat{T}_k\} \tag{3.9}$$

where $\hat{T}_j$ is the approximation of the tensor T_j. We use the robust approximation in this work to minimize both the propagation/amplification of error introduced by approximating tensor networks and the error associated with rank-deficient CP3 and CP4 approximations. Next, we introduce a reduced-scaling scheme to optimize the CP4 factor matrices.

3.2.2 Reduced-scaling CP decompositions using tensor network approximations

In general, constructing an order-N CP decomposition of a tensor, $T \in \mathbb{R}^{I \times I \times \cdots \times I}$, using an alternating least squares (ALS) optimization scheme has a computational scaling of $\mathcal{O}(I^N R)$ per iteration, where R is the rank of the decomposition. This means, given $R \propto N$, a naive

optimization of the CP4 decomposition has an iterative cost of $\mathcal{O}(N^5)$. However, we realize that it is possible to reduce the complexity of the ALS scheme by replacing a target tensor in any CP loss function with a *representative* tensor-network. For example, CP4 seeks to optimize the order-four Coulomb integral tensor g using the loss function

$$f(\mathbf{x}) = \frac{1}{2}\|g^{ia}_{bc} - \hat{g}(\mathbf{x})^{ia}_{bc}\|^2. \tag{3.10}$$

However, we can accurately approximate g tensor as a product of order-three tensors using the conventional DF approximation. Substituting the DF tensor network into the CP4 loss function creates the approximate loss function

$$f(\mathbf{x}) \approx \frac{1}{2}\|B^{iX}_{b} B^{aX}_{c} - \hat{g}(\mathbf{x})^{ia}_{bc}\|^2. \tag{3.11}$$

Using the DF-network approximated CP loss function, we construct an ALS optimization with a reduced-scaling cost of $\mathcal{O}(N^4)$ per iteration. Although others have introduced tensor networks into the CP loss function[73], this is the first implementation which we know of that computes factor matrices for *only* the external modes of a tensor network. The CP decomposition, in this way, provides one a route to approximate the result of a tensor-network without the apriori computation of such tensor network.

Now, both the CP3 and CP4 ALS optimizations have a computational cost of $\mathcal{O}(N^4)$. In Section 3.4, to demonstrate the effectiveness of the DF approximated CP4 ALS optimization algorithm, we approximate the order-four coulomb integral tensor, g, and use such approximation to effectively reduce the cost of LT CCSD(T) while introducing negligibly small errors

in relevant energy differences. Next, we discuss the Laplace transformation formulation of the (T) energy correction and subsequently, we demonstrate how we use CP approximations to reduce 9 rate-limiting contractions in LT CCSD(T).

3.2.3 The traditional and Laplace-transformed (T) energy correction

In the canonical Hartree-Fock basis the spin-restricted closed-shell (T) energy consists of 2 contributions:[13]

$$E_{(\mathrm{T})} = E_{\mathrm{T}}^{[4]} + E_{\mathrm{ST}}^{[5]}, \tag{3.12}$$

with

$$E_{\mathrm{T}}^{[4]} = \frac{1}{3} \sum_{ijk} \sum_{abc} \frac{W_{abc}^{ijk} R_{abc}^{ijk}}{D_{abc}^{ijk}}, \tag{3.13}$$

$$E_{\mathrm{ST}}^{[5]} = -\frac{1}{3} \sum_{ijk} \sum_{abc} \frac{V_{abc}^{ijk} R_{abc}^{ijk}}{D_{abc}^{ijk}}. \tag{3.14}$$

Note that the superscripts in Eq. (3.12) denote the Møller-Plesset (MP) order relative to an HF zeroth-order state. * In Eqs. (3.13) and (3.14), the order-6 tensors, W_{abc}^{ijk}, V_{abc}^{ijk}, R_{abc}^{ijk} and

*Additionally, for non-HF references (or for spin-restricted open-shell HF references; these cases will not be considered here) both contributions of these energies are 4-th order in MP, and, furthermore, there is another 4th-order contribution to the energy that must be included.[200]

D^{ijk}_{abc}, are defined as:

$$W^{ijk}_{abc} = \hat{P}^{ijk}_{abc}\left(\sum_f g^{if}_{ab} t^{kj}_{cf} - \sum_m g^{ij}_{am} t^{mk}_{bc}\right), \tag{3.15}$$

$$V^{ijk}_{abc} = g^{ij}_{ab} t^k_c + g^{jk}_{bc} t^i_a + g^{ki}_{ca} t^j_b, \tag{3.16}$$

$$R^{ijk}_{abc} = 4W^{ijk}_{abc} + W^{kij}_{abc} + W^{jki}_{abc} - 2W^{kji}_{abc} - 2W^{ikj}_{abc} - 2W^{jik}_{abc}, \tag{3.17}$$

$$D^{ijk}_{abc} = f^a_a + f^b_b + f^c_c - f^i_i - f^j_j - f^k_k, \tag{3.18}$$

where t^{ij}_{ab} and t^i_a are the converged CCSD cluster amplitude tensors, g^{pq}_{rs} is the Coulomb repulsion two-electron integral tensor, Eq. (3.1), f^p_p is the diagonal matrix elements of the Fock operator, and $\hat{P}^{ijk}_{abc}$ operator defines index permutations for a given six-index tensor:

$$\hat{P}^{ijk}_{abc}(O^{ijk}_{abc}) = (O^{ijk}_{abc}) + (O^{jki}_{bca}) + (O^{kij}_{cab}) + (O^{ikj}_{acb}) + (O^{kji}_{cba}) + (O^{jik}_{bac}). \tag{3.19}$$

Notice, the indices $i, j, k, l, \ldots$ and $a, b, c, d, \ldots$ represent the occupied and unoccupied subspaces of the molecular orbital (MO) basis,respectively, while the indices $p, q, r, s \ldots$ represent general indices.

The asymptotic complexity of the (T) energy is controlled by the construction of the W^{ijk}_{abc} intermediate whose cost is $O^3V^4 + O^4V^3 = \mathcal{O}(N^7)$, where O and V correspond to the size of the occupied and unoccupied orbital subspaces, respectively. Typically, the canonical HF basis is delocalized over the entire chemical system and, unfortunately, orbital localization techniques, such as the pair natural orbital[185, 201–203] (PNO) technique, cannot be introduced via simple unitary rotations. The complexity of orbital localization comes from the

fact that the orbital energy denominator, Eq. (3.18), can only be written in this simple form for the diagonalized HF basis. However, it is possible to factorize the inverse orbital energy denominator tensor using a Laplace transformation (LT)[74–77]:

$$(D^{ijk}_{abc})^{-1} = \int_0^\infty e^{-D^{ijk}_{abc}s} ds \approx \sum_\alpha^{n_q} w_\alpha e^{-D^{ijk}_{abc}s_\alpha}. \tag{3.20}$$

The LT is approximated on a quadrature grid using n_q quadrature points and a weight of w_α for every quadrature point, α. Rewriting the (T) energy contributions, Eqs. (3.13) and (3.14), with this factorization restores the freedom to change representation via unitary orbital rotations[78]. The LT (T) energy terms can be written as

$$\begin{aligned} E^{[4]}_{\mathrm{T}} =& \frac{1}{3}\sum_\alpha \sum_{ijkabc} w_\alpha W^{ijk}_{abc} R^{ijk}_{abc} e^{-D^{ijk}_{abc}s_\alpha} \\ =& \frac{1}{3}\sum_\alpha w_\alpha W^{ijk}_{abc} \left(\mathbf{f}^{(\alpha)}_o\right)_{i,i'} \left(\mathbf{f}^{(\alpha)}_o\right)_{j,j'} \left(\mathbf{f}^{(\alpha)}_o\right)_{k,k'} \left(\mathbf{f}^{(\alpha)}_u\right)_{a,a'} \left(\mathbf{f}^{(\alpha)}_u\right)_{b,b'} \left(\mathbf{f}^{(\alpha)}_u\right)_{c,c'} R^{i'j'k'}_{a'b'c'}, \end{aligned} \tag{3.21}$$

$$\begin{aligned} E^{[5]}_{\mathrm{ST}} =& -\frac{1}{3}\sum_{ijkabc}\sum_\alpha w_\alpha V^{ijk}_{abc} R^{ijk}_{abc} e^{-D^{ijk}_{abc}s_\alpha} \\ =& -\frac{1}{3}\sum_\alpha w_\alpha V^{ijk}_{abc} \left(\mathbf{f}^{(\alpha)}_o\right)_{i,i'} \left(\mathbf{f}^{(\alpha)}_o\right)_{j,j'} \left(\mathbf{f}^{(\alpha)}_o\right)_{k,k'} \left(\mathbf{f}^{(\alpha)}_u\right)_{a,a'} \left(\mathbf{f}^{(\alpha)}_u\right)_{b,b'} \left(\mathbf{f}^{(\alpha)}_u\right)_{c,c'} R^{i'j'k'}_{a'b'c'}. \end{aligned} \tag{3.22}$$

In Eqs. (3.21) and (3.22) we employ Einstein summation convention for all *covariant* dummy indices (those indices that appear twice on distinct tensors) and introduce the following exponentiated occupied/unoccupied Fock matrices

$$\mathbf{f}^{(\alpha)}_o \equiv e^{s_\alpha \mathbf{f}_o}, \quad \text{where} \quad (\mathbf{f}_o)_{i,i'} \equiv f_{i,i'}, \tag{3.23}$$

$$\mathbf{f}^{(\alpha)}_u \equiv e^{-s_\alpha \mathbf{f}_u}, \quad \text{where} \quad (\mathbf{f}_u)_{a,a'} \equiv f_{a,a'}. \tag{3.24}$$

After eliminating the orbital energy denominator in Eqs. (3.13) and (3.14) it is no longer necessary to construct the order-6 tensor intermediates, W and R, before evaluating the energies. It is, instead, possible to expand and refactorize the terms in Eqs. (3.21) and (3.22) in such a way that only order-4 tensor intermediates appear and, in doing such, one reduces the complexity of the method to $\mathcal{O}(N^6)$. Starting from Constans[78] factorization of LT CCSD(T), we introduce the DF approximation and form 75 unique energy contributions[†]. In Table 3.1 we have reported number of $\mathcal{O}(N^6)$ contractions found in this DF-LT (T) factorization and we report the time it takes to compute the most cost-intensive contractions per quadrature point for a $(H_2O)_{12}$ cluster in TIP4P geometry[139, 140] using the aug-cc-pVDZ orbital basis with an aug-cc-pVDZ-RI DF basis[143, 144]. This calculation was run using the Virginia Tech Advanced Research Computing's (ARC)Cascades cluster which utilizes standard nodes that contain 2 Intel Xeon E5-2683 v4 CPUs. As one can see, the most

	Number of terms	Cost (s)
OV^5	1	2,264†
O^2V^4	8	3,139$^\top$
O^3V^3	22	637
O^4V^2	12	61
O^5V	5	–

Table 3.1: The number of terms in DF-LT CCSD(T) which have an $\mathcal{O}(N^6)$ computational scaling, and the average cost of computation per quadrature point for a $(H_2O)_{12}$ cluster using TIP4P[139, 140] optimized geometry using and aug-cc-pVDZ orbital basis[143, 144] set and a cc-pVDZ-RI[142, 143, 204] DF basis set.
†: The timing of this computation actually contains 1 OV^5 contraction and 1 O^2V^4 contraction.
⊤: The timing of this computation contains 7 of 8 O^2V^4 contractions.

time-intensive portions of this DF-LT CCSD(T) method are the eight O^2V^4 contractions and the one OV^5 contraction. In the sections to follow we, first, demonstrate the limitations of CP3 approximated tensor networks and finally, show how we use CP approximations to reduced the complexity of these 9 rate-limiting contractions.

[†]The complete set of DF-LT CCSD(T) equations can can be found in Section 4.1.1

3.2.4 The limitations of the CP3 approximation

In this work, we seek to reduce the complexity of the 9 $\mathcal{O}(N^6)$ contractions in LT CCSD(T). However, in our efforts to reduce the complexity of these contractions we realized the limitations of the CP3 approximation of the DF Coulomb integral tensor network. One tensor network in LT CCSD(T) where CP3 fails is

$$\sum_{abef}\sum_{ic} g_{ci}^{fa} g_{ci}^{eb} \sum_{ij} t_{ij}^{ea} t_{ij}^{fb}, \tag{3.25}$$

and here we walk through the key points which lead to the failure of CP3.

First consider the sub-network

$$\sum_{ic} g_{ci}^{fa} g_{ci}^{eb}, \tag{3.26}$$

which has a cost of OV^5 Notice that the labels i and c occupy different particle indices for both Coulomb integral tensors. Because of this fact, it is not possible to use DF to reduce the complexity of this sub-network. It is, however, possible to reduce the complexity of the sub-network using using the following CP3 formulation:

$$\sum_{ic} g_{ci}^{fa} g_{ci}^{eb} \approx \sum_{r}^{R_{\mathrm{CP3}}} \rho_r^a (\sum_i (\sum_c g_{ci}^{fa} \kappa_r^c)(\sum_X B_i^{bX} \chi_r^X)). \tag{3.27}$$

However, as one can see that reducing the complexity of this sub-network leaves the full network with a computational scaling of O^2V^4. Upon further investigation, one realizes that

there is no way in which CP3 can be used to reduce the complexity of the full network. The CP3 approximation has a limited degree of freedom because each DF tensor requires its own CP3 factorization, and thereby each DF tensor introduces a sum over a *different* rank hyperdimension, i.e.

$$g_{ci}^{ab} \approx \sum_X B_c^{aX} B_i^{bX} \approx \sum_X (\sum_r^{R_{\text{CP3}}} \rho_r^a \kappa_r^c \chi_r^X)(\sum_{r'}^{R_{\text{CP3}}} \gamma_{r'}^i \beta_{r'}^b \chi_{r'}^X). \tag{3.28}$$

This limitation, coupled with the fact that each CCSD amplitude tensor, t, contains a single index from each Coulomb integral tensor, g, leads to the failure of CP3.

Fortunately, this inadequacy can be overcome by increasing the flexibility of our model using the CP4 approximation. We rewrite Eq. (3.25) using the CP4 approximation as

$$\sum_{rfa}^{R_{\text{CP4}}} (\sum_{ci} g_{ci}^{fa} \gamma_r^{ci})(\sum_{ij} (\sum_e \rho_r^e t_{ij}^{ea})(\sum_b \beta_r^b t_{ij}^{fb})) \tag{3.29}$$

where γ is constructed using the Khatri Rao product

$$\gamma_r^{ci} = \gamma_r^i \odot \kappa_r^c. \tag{3.30}$$

As predicted, the CP4 approximation of Eq. (3.25) as Eq. (3.29) reduces the complexity of both restrictively large $\mathcal{O}(N^6)$ contraction. Next we walk through how we reduce the complexity of 9 $\mathcal{O}(N^6)$ terms in DF-LT CCSD(T).

3.2.5 Reduced-scaling algorithms: rCP-DF-LT CCSD(T)

Finally, we demonstrate how how to reduce the scaling of nine rate-limiting contractions using CP approximated integral tensors. Additionally, we analyze the cost of the steps involved in these reduced-scaling computations and compare the reduced-cost to the term's first conventional counterpart.

Because of their similar network topology, we factorize the following three terms in nearly the same manor as,

$$(g^{ie}_{bc}g^{if}_{ac})(t^{jk}_{ea}t^{jk}_{bf} - 2t^{jk}_{ea}t^{jk}_{fb}) = \sum_{r}^{R_{cp4}}(\gamma^{ic}_{r}g^{if}_{ac})\left[(\rho^{e}_{r}t^{jk}_{ea})(\kappa^{b}_{r}(t^{jk}_{bf} - 2t^{jk}_{fb}))\right] \tag{3.31}$$

$$(g^{ie}_{cb}g^{if}_{ca})(4t^{jk}_{ea}t^{jk}_{bf} - 2t^{jk}_{ea}t^{jk}_{fb}) = \sum_{r}^{R_{cp3}}(\chi^{X}_{r}B^{iX}_{c})g^{if}_{ca}\left[(\rho^{e}_{r}t^{jk}_{ea})(\kappa^{b}_{r}(4t^{jk}_{bf} - 2t^{jk}_{fb}))\right] \tag{3.32}$$

and

$$(g^{ie}_{bc}g^{if}_{ca})(4t^{jk}_{ea}t^{jk}_{bf} - 2t^{jk}_{ea}t^{jk}_{fb}) = \sum_{r}^{R_{cp3}}(\chi^{X}_{r}B^{iX}_{c})g^{ie}_{bc}\left[(\kappa^{a}_{r}t^{jk}_{ea})(\rho^{f}_{r}(4t^{jk}_{bf} - 2t^{jk}_{fb})\right]. \tag{3.33}$$

The cost of computing these contractions in these three terms are nearly identical, therefore, during the analysis of the cost I will use a general R to mean either R_{CP3} or R_{CP4}.[‡] First, computation of the contraction $(\gamma^{ic}_{r}g^{if}_{ac})$ [which is similar to $(\chi^{X}_{r}B^{iX}_{c})g^{if}_{ca}$] costs RV^3O. Second, the computation of$(\rho^{e}_{r}t^{jk}_{ea})$ [which is similar to $(\rho^{f}_{r}(4t^{jk}_{bf}-2t^{jk}_{fb})$] costs RO^2V^2. Next, each of the

[‡]Please note, to introduce the robust network approximation we utilize symmetry in the CP factorizations for Eqs. (3.31) and (3.32). However, for Eq. (3.33) we must use both CP3 and CP4 factorizations.

three terms has the non-covariant contraction $t_{ra}^{jk}t_{rf}^{jk}$ which costs RO^2V^2 to compute. Finally, each term has the dot product $g_f^{ra}t_f^{ra}$ which costs ROV; thus, the final cost of these terms is $RV^3O + 3RO^2V^2 + ROV$. The conventional cost of these contractions is $OV^5 + V^4O^2 + V^4$ and, taking the leading order costs, we recognize that we will find a cost reduction when $R < V^2$. Later, we will show that R values for CP3 and CP4 are on the order of V.

Next we show the factorizations for the remaining O^2V^4 contractions. Starting with Eq. (4.6), from the appendix, we factorizes $G2$ and $G3$ using CP3 as

$$G2_{fi}^{ea} = g_{bj}^{ea}t_{ji}^{fb} = \sum_{r}^{R_{cp3}} \beta_r^a(\chi_r^X B_b^{eX})(\gamma_r^j t_{ji}^{fb}) \tag{3.34}$$

and

$$G3_{fi}^{ea} = g_{bj}^{ea}t_{ij}^{fb} = \sum_{r}^{R_{cp3}} \beta_r^a(\chi_r^X B_b^{eX})(\gamma_r^j t_{ij}^{fb}). \tag{3.35}$$

These factorizations reduce the cost of these contractions from O^2V^4 to $O^2V^2R_{\text{CP3}}$ $+OV^3R_{\text{CP3}}$ and, therefore, we will find speedup in these terms when $R < VO$. Please note, the factorization of $G2$ and $G3$ are the only ones in this work where we do not use the robust network formulation.

Next, we reformulate the g times g contractions found in Eqs. (4.17) and (4.18) using symmetric robust CP4 approximations:

$$g_{bi}^{ec}g_{cj}^{fb} = \sum_{r}^{R_{cp4}} \gamma_r^{ei}(\gamma_r^{cb}(2.0g_{cj}^{fb} - \hat{g}_{cj}^{fb})) \tag{3.36}$$

and

$$g_{bi}^{ec} g_{bj}^{fc} = \sum_{r}^{R_{cp4}} \gamma_r^{ei} (\gamma_r^{cb} (2g_{bj}^{fc} - \hat{g}_{bj}^{fc})) \tag{3.37}$$

where $\hat{g}$ is CP4 approximation of g. Subsequently, we compute the dot product of the result of Eqs. (3.36) and (3.37) with $T1$ and $T2$ which are defined in the Appendix, Section 4.1.1. These CP4 factorizations reduce the scaling of these contractions from O^2V^4 to V^3OR.

Finally, we factorize Eq. (4.35) using the CP3 approximation

$$-t_{kl}^{ab}(g_{ak}^{ec}(t_{ij}^{eb}(2g_{il}^{cj} - g_{jl}^{ci}))) = -\sum_{r}^{R_{cp3}} (\chi_r^X B_a^{eX})(\beta_r^k t_{kl}^{ab})(t_{ij}^{eb}(\gamma_r^c(2\tilde{g}_{il}^{cj} - \tilde{g}_{jl}^{ci}))) \tag{3.38}$$

$$t_{kl}^{ba}(g_{ak}^{ec}(t_{ij}^{eb}(g_{il}^{cj} - 2g_{jl}^{ci}))) = \sum_{r}^{R_{cp3}} (\chi_r^X B_a^{eX})(\beta_r^k t_{kl}^{ba})(t_{ij}^{eb}(\gamma_r^c(\tilde{g}_{il}^{cj} - 2\tilde{g}_{jl}^{ci}))) \tag{3.39}$$

where, to utilize the robust tensor network approximation, we use

$$\tilde{g}_{il}^{cj} = 2B_i^{cX} B_l^{jX} - \hat{B}_i^{cX} B_l^{jX} \tag{3.40}$$

in place of the canonical g_{il}^{cj} tensor. These factorizations reduce the complexity of the $t_{kl}^{ab} g_{ak}^{ec}$ contraction from O^2V^4 to OV^3R. Notice, however, that this reformulation does not reduce the O^3V^3, $t_{ij}^{eb} g_{il}^{cj}$ contraction.

In the sections to follow, we will show that the CP3 and CP4 ranks required to preserve the accuracy of relevant energetic differences is proportional to the size of DF fitting basis and we

demonstrate realized cost reduction over DF-LT (T) and canonical DF (T) for medium-sized molecules.

3.3 Computational details

rCP-DF-LT CCSD(T) and DF-LT CCSD(T) calculations have been computed using a developmental version of the massively parallel quantum chemistry (MPQC) software package[1] while DF CCSD(T) calculations were computed using the Psi4 software package.[205] All of these computations were run on the Virginia Tech Advanced Research Computing's (ARC) Cascades cluster which utilizes standard nodes that contain 2 Intel Xeon E5-2683 v4 CPUs. The CP approximations of order-3 and order-4 tensors were computed using the standard alternating least squares (ALS) method.[130, 131] In this work, as was done in our previous work,[5] initial guess of vectors for the CP3 decomposition were generated using quasi-random numbers taken from the uniform distribution on [-1,1]. As discussed earlier in this work, we utilized optimized CP3 factors and quasi-random numbers as an initial guess strategy for the CP4 decomposition. Using these initialization approaches, we see rapid convergence to high accuracy approximations.

Quadrature accuracy was studied using a $(H_2O)_{18}$ cluster with TIP4P optimized geometry[139, 140] and the two largest molecules from the S66 benchmark set of weakly bound complexes[134] (uracil ... uracil and pentacene ... pentacene). These S66 geometries were taken from the Benchmark Energy and Geometry Database (BEGDB).[135] For the $(H_2O)_{18}$ cluster we utilized the cc-pVDZ (DZ) basis set accompanied by the corresponding cc-pVDZ-

RI (DZ-RI) DF basis set,[142, 143, 204] for the uracil ... uracil computation we utilized the cc-pVDZ-F12 (DZ-F12) basis set with the aug-cc-pVDZ-RI (aVDZ-RI) DF basis set,[141, 142] and the pentacene ... pentacene computation utilized the 6-31G*(0.25) basis set with the aVDZ-RI DF basis set[204, 206].

CCSD(T) Energy and timing data was drawn from water clusters with 5 to 12 molecules in the TIP4P optimized geometry[139, 140], using this data we determine the optimal CP rank for rCP-DF-LT CCSD(T) and show the speedup of rCP-DF-LT CCSD(T) over DF-LT CCSD(T). Here we define speedup as

$$\text{speedup} = \frac{t_{\text{rCP-DF-LT CCSD(T)}} + t_{\text{ALS}}}{t_{\text{DF-LT CCSD(T)}}} \tag{3.41}$$

where t_{ALS} is the time it takes to optimize the CP3 and CP4 approximations using the ALS optimization scheme. Using this optimal CP rank, we show that the error in reaction energies for the Neese, Wennmohs, Hansen (NWH) dataset[207], a set of 23 reactions with molecules up to 36 atoms, is less than our desired error threshold (0.01 kcal/mol). [§] The water cluster and reaction energy calculations in this work utilized the aug-cc-pVDZ (aVDZ) basis set[143, 144] with the aVDZ-RI DF basis set.

[§]The NWH dataset conatins the following molecules, sorted from least number of vitaul orbitals to greatest: 1. H_2, 2. C_2H_4, 3. H_2CN_2, 4. propyne, 5. allene, 6. cyclopropene, 7. N-methylenemethanamin, 8. aziridin, 9. propene, 10. cyclopropane, 11. acetic acid, 12. methyl formate, 13. ethanediol, 14. dimethylperoxide, 15. propan-2-one, 16. oxetane, 17. buta-1,3- diene, 18. cyclobutene, 19. 1-pyrazoline, 20. cyclopentene, 21. vinylcyclopropane, 22. dihydrofuran-2(3H)-one, 23. vinyl acetate, 24. heptahexane, 25. heptatriyne, 26. 2-hydroxypyridine, 27. 2-hydropyridone, 28. neo-pentane, 29. n-pentane, 30. tetrahydropyran-2-one, 31. pentan-2,4-dione, 32. toluene, 33. norbornadiene, 34. 2,3-dimethylbut-2-ene, 35. styrene, 36. cyclooctatetraene, 37. C9O3 (carbo-[3]oxocarbon), 38. C9O3(D3hiso2), 39. p-xylene, 40. hexanoic acid, 41. methyl pivalate, 42. n-octane, 43. 2,2,3,3 -tetramethylbutane, 44. $C_{12}H_{12}$, 45. cyclopentadienyletropenyle, 46. octamethylcyclobutane, 47. [2,2]paracyclophane.

3.4 Results

3.4.1 Quadrature accuracy

Though the work presented here does not explicitly depend on the number of quadrature points, we would like to use enough points to properly represent realistic calculations and demonstrate the accuracy of the rCP formulation. Therefore, in this section, we study how the number of quadrature points affects the accuracy of absolute and relative energy values and determine how many points we should use for the remainder of this study. We limit this investigation to a few large molecules, $(H_2O)_{18}$, uracil ... uracil and pentacene ... pentacene. Our results are summarized in Table 3.2. In this table, ΔE_{abs} represents the DF-LT CCSD(T) absolute energy error per electron with respect to DF CCSD(T) while $\Delta E_{b.e.}$ represents the binding energy error with respect to DF CCSD(T).

	$(H_2O)_{18}$		Uracil dimer		Pentacene dimer	
n_q	ΔE_{abs}	$\Delta E_{\text{b.e.}}$	ΔE_{abs}	$\Delta E_{\text{b.e.}}$	ΔE_{abs}	$\Delta E_{\text{b.e.}}$
2	0.00244	0.180	0.114	0.561	0.313	44.9
3	-2.22×10^{-5}	-0.00406	0.0185	0.113	0.0496	9.10
4	-1.48×10^{-5}	-0.00215	0.00369	0.0272	0.00499	0.979
5	-8.71×10^{-6}	-0.00119	7.22×10^{-4}	0.00645	4.18×10^{-4}	0.0841
6	-5.66×10^{-6}	-0.00071	1.24×10^{-4}	0.00119	3.56×10^{-5}	0.00740

Table 3.2: Error in absolute per electron energy, Δ_{abs} (kcal/mol), and binding energy, $\Delta_{b.e.}$ (kcal/mol), for LT-DF CCSD(T) compared to DF CCSD(T) with respect to the number of quadrature points n_q for selected chemical systems.

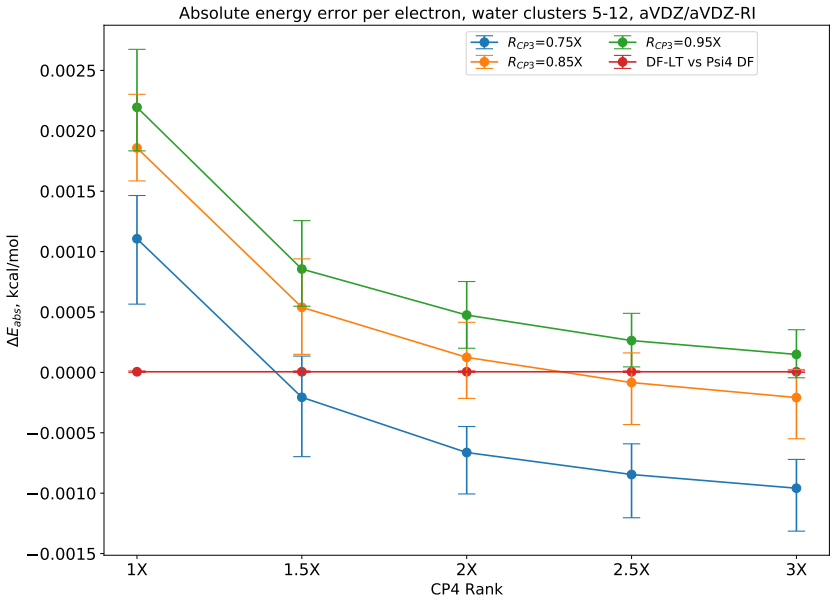
Both the absolute and relative energy errors fall monotonically with number of quadrature points which is in agreement with the findings presented in Constans' LT CCSD(T)[78] work. To determine an optimal number of quadrature points we utilize the standard chemical accuracy requirement of less than 1 kcal/mol error in relative energy. We find this chemical

accuracy requirement is satisfied when $n_q \geq 4$, and, in the following, choose to use $n_q = 4$.

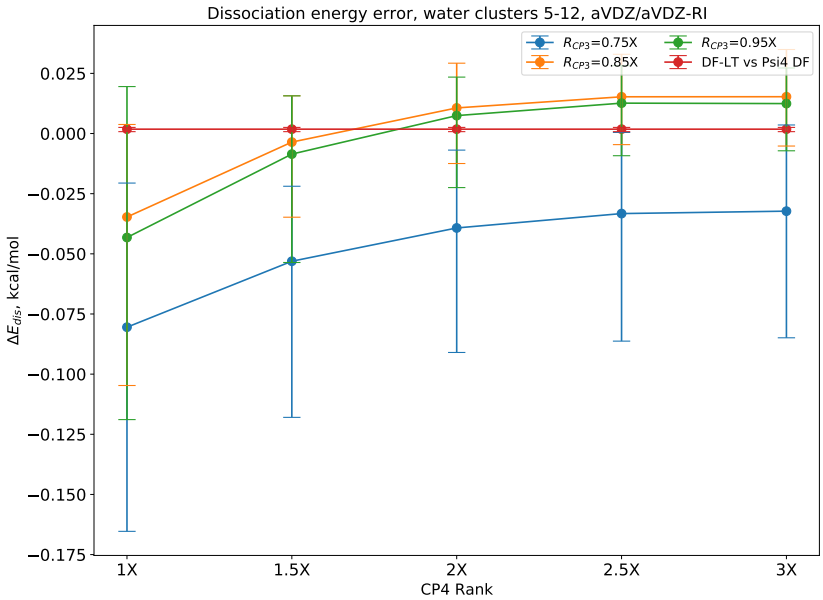
3.4.2 Computational performance of rCP-DF-LT CCSD(T)

In this results section it is our goal to determine minimum CP3 and CP4 ranks for rCP-DF-LT CCSD(T) and to show that the method can reduce the cost of DF-LT CCSD(T) for medium-sized molecules. We desire these molecules to be relatively equal in size to the molecule fragments generated by Nagy's LNO CCSD(T) method. In their recent work, Nagy has shown that LNO CCSD(T) requires localized fragments which contain between 108 and 563 unoccupied orbitals[183, 208] and between 29 and 92 local occupied orbitals[183, 208] per fragment . To represent fragments of this size we use water clusters with between 5 to 12 water molecules, which have between 180 and 435 unoccupied orbitals. After determining a minimum CP3 and CP4 rank for these water clusters we, further, compute the error in reaction energy for the Neese, Wennmohs, Hansen (NWH) dataset and demonstrate that the chosen ranks introduce little error into DF-LT (T) computations.

First, we seek to understand the error introduced by the CP approximations. We use this error analysis to help determine optimal CP3 and CP4 ranks. As discussed in Section 3.4.1, we fixed the number of quadrature points to 4 in these computations. In Figure 3.1a we compare both rCP-DF-LT and DF-LT CCSD(T) to Psi4's DF CCSD(T) and compute the error in absolute (T) energy, ΔE_{abs}, per electron. Notice, the CP3 and CP4 ranks are defined as a multiple times the size of the DF basis, X. Figure 3.1a helps us understand the nature of the error introduced by the CP3 and CP4 approximations. We notice that, low rank CP3 and CP4 approximations, in this scope, benefit from fortuitous cancellation of errors. However,



(a)

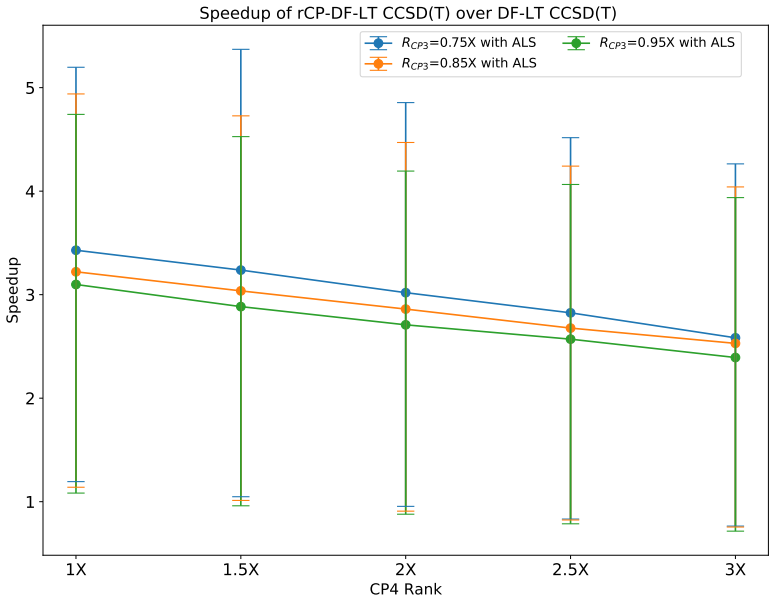


(b)

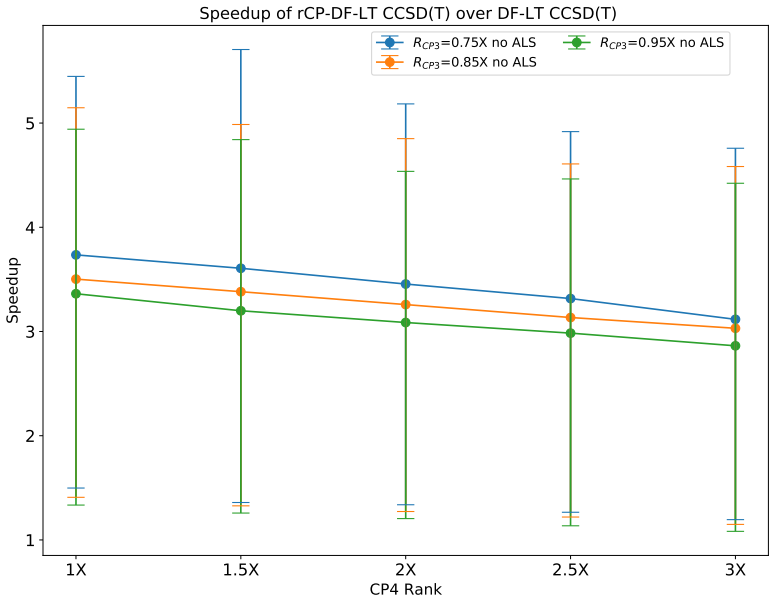
Figure 3.1: (a) Absolute energy error per electron (ΔE_{abs}, kcal/mol) for approximated LT CCSD(T) methods versus Psi4's DF CCSD(T). (b) Unsigned dissociation energy error (ΔE_{dis}, kcal/mol) for approximated LT CCSD(T) methods versus Psi4's DF CCSD(T). Error bars represent max and min error

when error associated with CP4 rank is reduced (at large values of R_{CP4}), the error due to CP3 is revealed. Therefore, we approximate CP3 error using data points at $R_{\mathrm{CP4}} = 3X$. CP3 error rapidly converges, from $R_{\mathrm{CP3}} = 0.75X$ to $R_{\mathrm{CP3}} = 0.95X$ the error per electron drops by a factor of about 9, from 9.5×10^{-4} kcal/mol to 1.4×10^{-4} kcal/mol. CP4 error converges less rapidly, for example from $R_{\mathrm{CP4}} = X$ to $R_{\mathrm{CP4}} = 1.5X$ we observe an error reduction of about 2. Additionally, we note that error associated with the CP4 approximation is significantly higher than the error associated with CP3. These observations elude to our later finding, that R_{CP4} must be larger than R_{CP3}. It is not reasonable to derive minimum CP3 or CP4 ranks using absolute energy errors so, next, we compute and compare the cluster dissociation energies.

In Figure 3.1b we compute the dissociation energy error, ΔE_{dis}, for water clusters $(H_2O)_5$ through $(H_2O)_{12}$ and use the results to determine the minimum-rank values for CP3 and CP4 rank. In this work, we require minimum-rank combinations to introduce a maximum error of less than 0.1 kcal/mol in relative energy, a threshold 10 times smaller than the conventional chemical accuracy measure. From Figure 3.1b, we see, again, that the CP4 error is slow to converge relative to CP3 error. Maximum CP3 error at $R_{\mathrm{CP3}} = 0.75X$ is already less than 0.1 kcal/mol and by $R_{\mathrm{CP3}} = 0.95X$ it is just slightly larger than 0.025 kcal/mol. While CP4 error at rank $R_{\mathrm{CP4}} = X$ is roughly 0.5 kcal/mol. One anomaly worth noting is the $R_{textCP3} = 0.85X$ curve is lower in error than the $R_{textCP3} = 0.95X$, when $R_{textCP4} \leq 1.5X$. This anomaly can be attributed to fortuitous cancellation of error since CP3 and CP4 errors are opposite in sign. This cancellation of error is, in general, not guaranteed and potentially a relic of these water systems tested. Therefore, for high accuracy calculation we suggest



(a)



(b)

Figure 3.2: (a) Speedup of rCP-DF-LT CCSD(T) over DF-LT CCSD(T) including (a) and excluding (b) the ALS optimization time for TIP4P water clusters (H_2O_5 through H_2O_{12}). Error bars provide max and min speedup.

using larger values of $R_{textCP4}$ when $R_{textCP3} < 0.95X$. In Figure 3.1b, we can identify the minimum-rank [CP3, CP4] pairs which satisfy our error threshold as [$0.75X$, $2X$], [$0.85X$,

$1.5X$] and [$0.95X$, $1.5X$]. However, to ensure the accuracy of $R_{textCP3} = 0.85X$ for systems other than water clusters, we will use the pair [$0.85X$, $2.0X$]. To chose the best minimum-rank combination, we need to understand the cost associated with CP3 and CP4 and how these costs change with rank.

Next, we seek to determine which minimum-rank CP3/CP4 pair, determined previously, has the capacity to reduce the cost of computing DF-LT CCSD(T) and which pair finds the greatest cost reduction. Figure 3.2 shows the speedup of the rCP-DF-LT CCSD(T) method over the DF-LT CCSD(T) method for the 8 water clusters, H_2O_5 through H_2O_{12}. Note that the maximum bars in Figure 3.2 are correlated with large system computations and the minimum bars are correlated with small system computations. The correlations found in Figure 3.2 develops for two reasons.

First, as system size grows the cost of computing optimized CP3 and CP4 decomposition becomes small relative to the cost of computing the CP-DF-LT (T) energy. However, when we assess the impact of the CP optimization by comparing the speedups in Figure 3.2a, which includes the ALS optimization, to the speedups in Figure 3.2b, which doesn't include the ALS optimization, we find that ALS optimization reduces the speedup of the max, min and average points by about 0.3. The consistency of cost between all of these points tells us that the ALS optimization is relatively fast for all 8 of these water clusters. Furthermore, in Figure 3.3 we show the percent of time spent computing the ALS optimization ($\frac{t_{\text{ALS}}}{t_{\text{rCP-DF-LT (T)}}} *$ 100) in rCP-DF-LT (T) calculations is less than 40 percent for all clusters and CP3 and CP4 ranks and, for water cluster of with 7 or more molecules, these three ALS optimizations takes less than a quarter of the total CP-DF-LT (T) time.

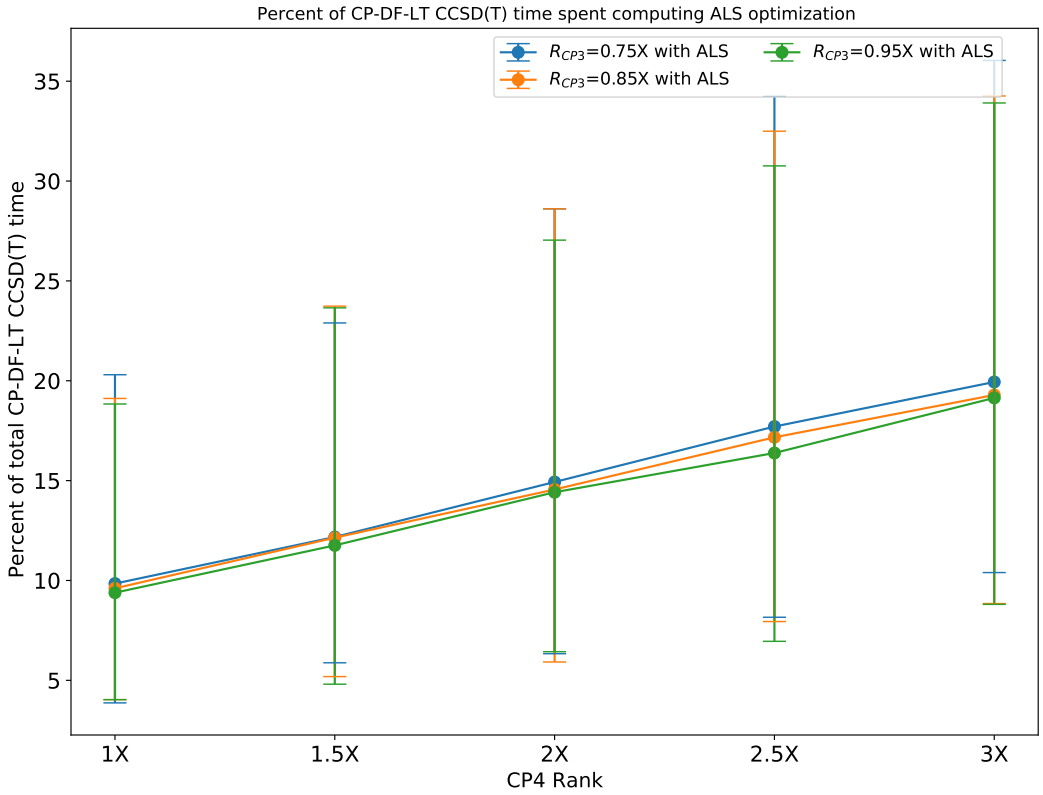


Figure 3.3: Percent of rCP-DF-LT CCSD(T) time spent computing the CP3 and CP4 ALS optimizations

Second, the general-type contractions (necessary in these CP algorithms) are tensor-algebra kernels which modern computer architectures are not optimized to compute. The low efficiency of the general-type contraction kernel (compared to the modern gemm kernel efficiency) becomes apparent when the rank is closer to the CP crossover, $R = VO$, i.e. in relatively small molecular systems. The general-type contractions are primarily used in algorithms in CP3 factorizations and, for this reason, we seek to minimize the CP3 rank. Moreover, Figure 3.2 shows that increasing CP4 rank has a relatively small influence on the cost of the rCP-DF-LT (T) method, compared to the cost of increasing the CP3 rank. Therefore, based on the conditions laid out in this section, we determine the best, minimum

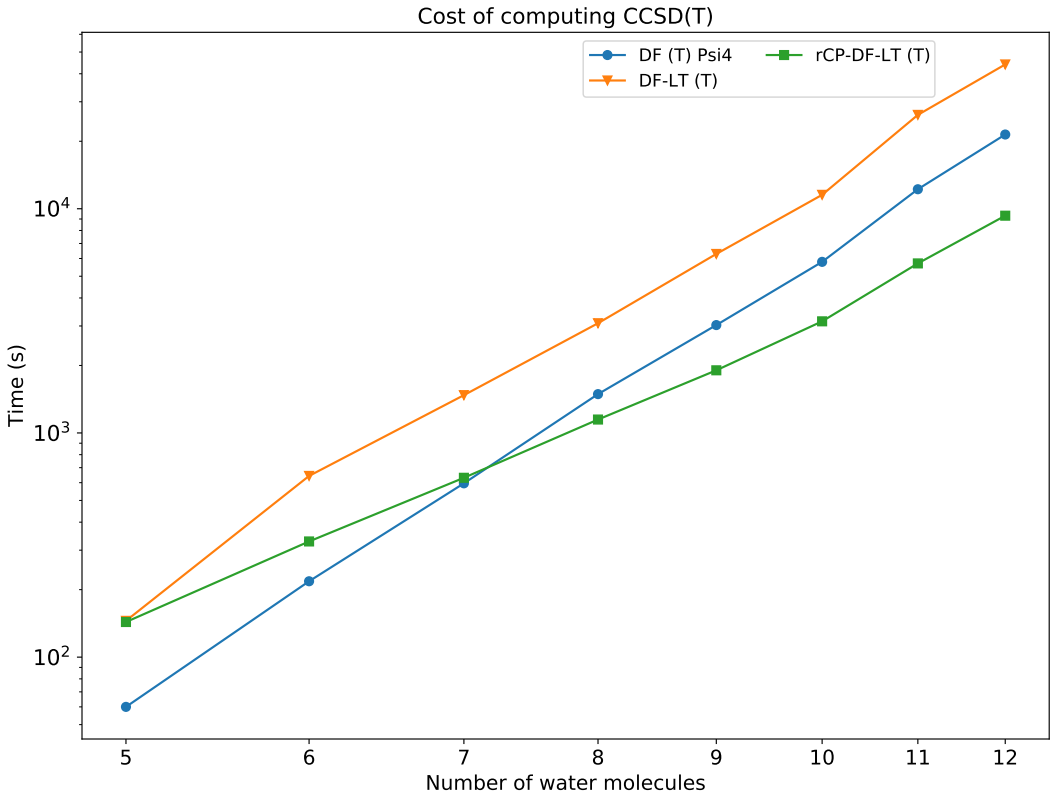


Figure 3.4: Cost of computing (T) energy for water clusters with between 5 and 12 molecules in aVDZ/aVDZ-RI using the conventional DF (T) method from Psi4[205], DF-LT (T) and rCP-DF-LT (T) with $R_{\mathrm{CP3}} = 0.75$ and $R_{\mathrm{CP4}} = 2.2$

CP3/CP4 pair to be [0.75X,2X]. Using this pair, we show, in Figure 3.4, the full cost of computing the (T) energy using the three different methods, DF (T), DF-LT (T) and rCP-DF-LT (T). rCP-DF-LT (T) quickly becomes less expensive than canonical DF (T), while the crossover of DF-LT (T) is not in the range of this plot. Next, we compute the reaction energies for the NWH dataset using CP3/CP4 ranks of [0.75X,2.2X] to further demonstrate the accuracy and performance of the rCP-DF-LT (T) approach.

In Figure 3.5 we present the reaction energy errors for the NWH dataset using rCP-DF-LT CCSD(T) with respect to DF-LT CCSD(T) using 4 quadrature points and $R_{\mathrm{CP3}} = 0.75X$

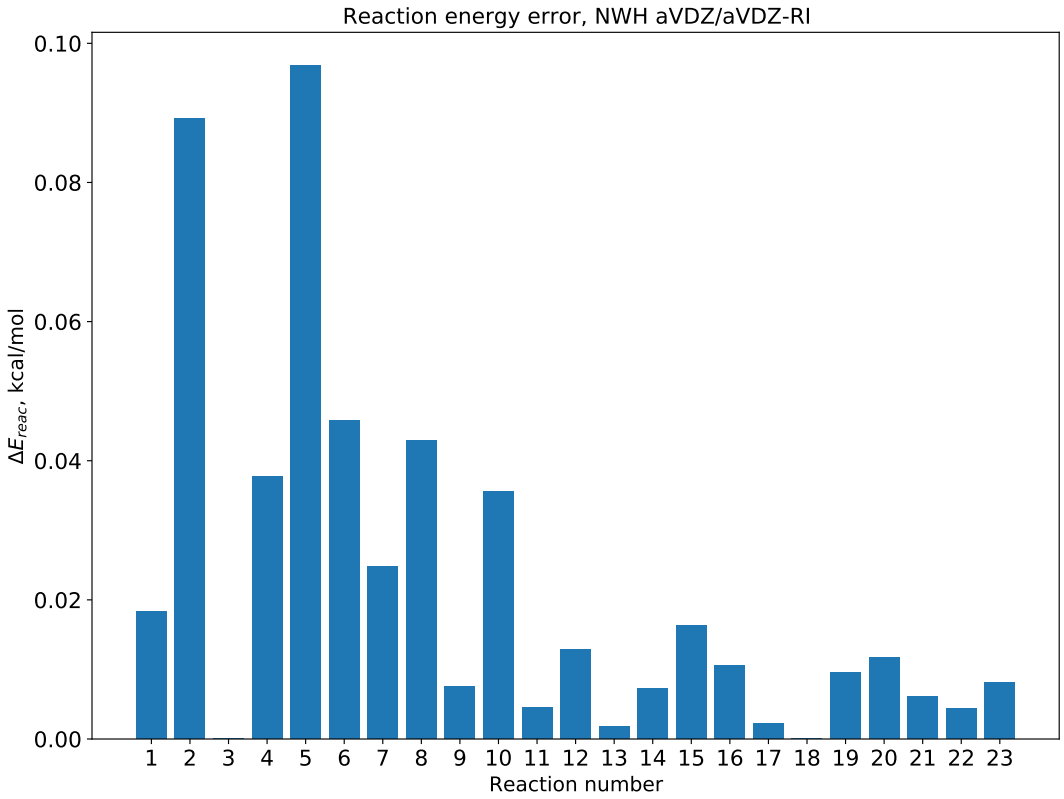


Figure 3.5: Reaction energy error (ΔE_{reac}, kcal/mol) for the NWH dataset in the aVDZ/aVDZ-RI basis. Reaction numbers correspond to those found in table IV in the original work[207]

and $R_{\mathrm{CP4}} = 2.2X$. All reaction energy computations here introduce less than 0.1 kcal/mol. We find that the largest reaction energy errors comes from reactions 5. (2 p-xylene → [2,2] paracyclophane + 2 H_2) and 2. (2 2,3-dimethylbut-2-ene → octamethylcyclobutane). These reactions contain our largest molecules, however the size of the molecule doesn't seem to strongly effect the accuracy of the (T) energy. Figure 3.6 shows the absolute energy error per electron for all the systems in the NWH dataset sorted from smallest to largest number of virtual orbitals. From this figure, one can see that the size of the systems has little effect on the magnitude of the absolute per electron energy errors associated with CP. However, it is obvious that larger molecules tend to have a negative deviation from the exact value,

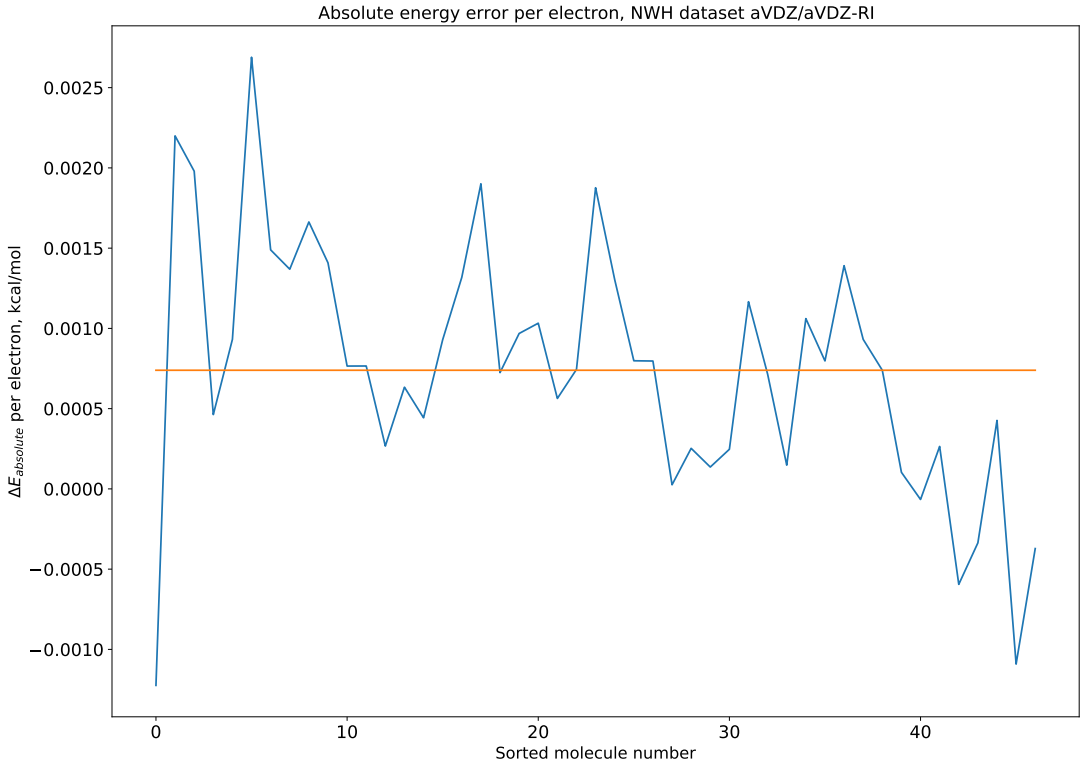


Figure 3.6: Absolute energy error per electron ($\Delta E_{absolute}$, kcal/mol) for the NWH dataset in the aVDZ/aVDZ-RI basis. Line represents average per electron error.

while smaller molecules have a positive deviation. The relatively high reaction energy error associated with these large molecule reactions could, thus, be a result of the difference in signed error of molecules involved in the reactions. For higher accuracy computations, we suggest increasing the CP3 rank to minimize this signed error problem.

In Figure 3.7 we show the speedup of the rCP-DF-LT CCSD(T) method for the largest 17 molecules in the NWH dataset. These molecules have 18 and more occupied orbitals and between 206 and 456 unoccupied orbitals which falls in the range of fragment sizes used in LNO CCSD(T) computations. We see definitive speedup over conventional DF-LT (T) when the number of unoccupied orbitals exceeds 208 and increased benefits as the number

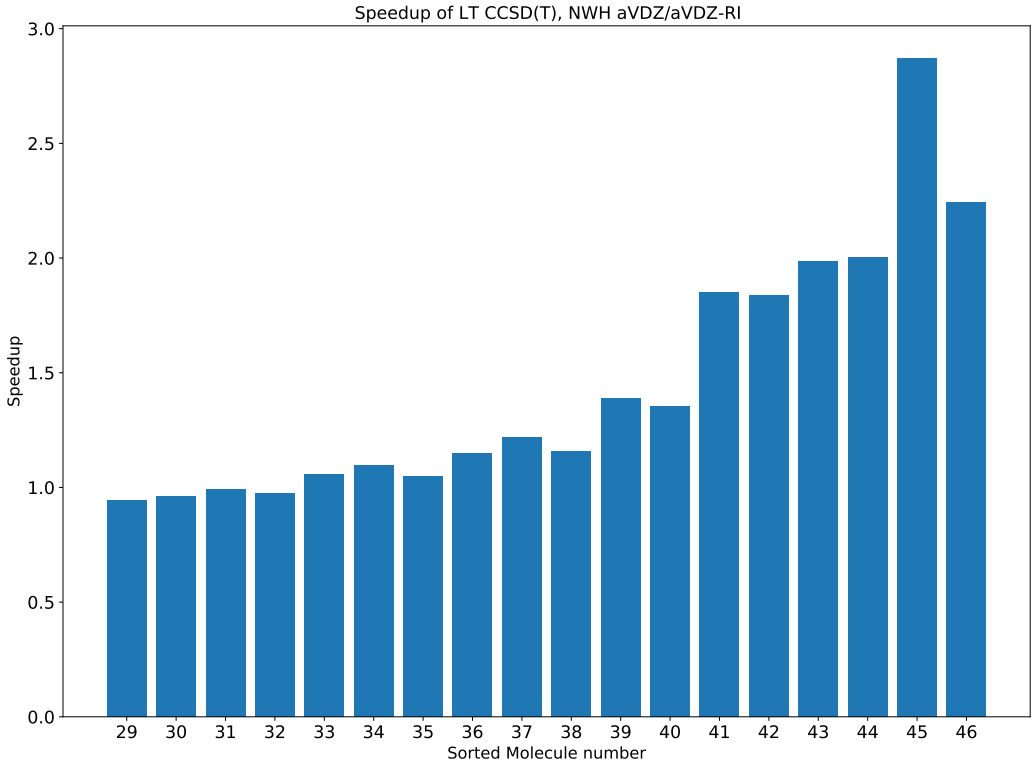


Figure 3.7: Speedup of largest 29 molecules in the NWH dataset with a CP rank of X

of unoccupied orbitals grows.

In conclusion, the CP4 approximation, introduced in this work, can be accurately computed using the DF network in the CP loss function. We have shown that the CP3 and CP4 approximations can be introduced accurately into the LT CCSD(T) method and can be used to reduce the complexity of the most restrictive terms in this method. We have demonstrated that both the CP3 and CP4 approximation require a relatively low CP rank ($R \approx X$) to achieve sub 0.1 kcal/mol accuracy in relative energies. Furthermore, we have shown that the cost of (T) can be significantly reduced for medium sized molecules using this rCP-DF-LT CCSD(T) method.

3.5 Further Complexity Reduction by Coarse-Graining

In traditional CCSD(T), for the sake of memory, contractions are course-grained over virtual indices. However,Constan's LT CCSD(T) implementation[78] avoids construction of the approximate order-6 three-particle amplitude tensor T_{abc}^{ijk} by completely expanding the tensor W and V in Eq. (3.13) and Eq. (3.14). Expansion of Eq. (3.13) and Eq. (3.14) significantly reduces the memory requirements of CCSD(T) and, therefore, gives us freedom to course-grain our computations over either occupied or unoccupied indices. Introduction of local molecular orbitals (LMOs), as done by Nagy et. al[166, 177, 182, 183], leads naturally to occupied-index coarse-grain algorithms. Unfortunately, an LMO implementation would also require the computation of CP decomposed Hamiltonian integral tensors for every localized occupied orbital, which may make the cost of constructing the CP4 decomposition prohibitively large. However, it may be reasonable to decompose fragmented two-particle amplitude tensors. We are currently working on an optimized implementation of this strategy.

3.6 Conclusions

In this work, we present means to reduce the scaling of the canonical polyadic (CP) factorization of higher-order tensors. We do this via the introduction of a representative tensor network into the CP optimization problem. To demonstrate the abilities of this reduced-scaling CP factorization, we compute the decomposition of the order-4 Coulomb integral

tensor using a density fitting (DF) tensor network factorization, denoted as the CP4 approximation. Additionally, we compute the CP factorization of the order-3 DF Coulomb integral tensors, which we denote the CP3 approximation. Using the CP3 and CP4 approximations, we reduce the complexity of 9 rate-limiting contractions found in the DF and Laplace-transformed approximated perturbative triples correction to the coupled cluster with single and double substitution [DF-LT CCSD(T)] method. Furthermore, we minimize the error associated with rank-deficient CP approximations using using a robust tensor network formulation, denoted as rCP. The robust tensor network formulation explicitly removes leading order error associated with approximating a tensor-network.

The rCP-DF-LT CCSD(T) implementation introduced here is able to reduce the cost of computing DF-LT CCSD(T) for molecules with 208 and 500 virtual orbitals, on average, by a factor of about 3.5. Furthermore, we show that CP3 and CP4 ranks necessary to introduce negligible error in relevant relative energies is roughly the size of the DF fitting basis, $R_{\mathrm{CP3}} = 0.75X$ and $R_{\mathrm{CP4}} = 2X$. In our future work, we plan to integrate these CP approximations into an orbital localized, fragmented version of DF-LT CCSD(T) to further reduce the cost of computing accurate energies of large molecular systems.

3.7 Acknowledgement

This work was supported by the U.S. National Science Foundation (awards 1550456 and 1800348). We also acknowledge Advanced Research Computing at Virginia Tech (www.arc.vt.edu) for providing computational resources and technical support that have contributed to the

results reported within this paper.

Chapter 4

Conclusion and outlook

The goal of this research has been to address some of the major issues associated with introducing the canonical polyadic (CP) decomposition to reduce the computational scaling of electronic structure methods. In previous studies, relatively large rank CP decompositions have been necessary to ensure reasonably small error introduction. These high rank decompositions made it difficult to see a cost reduction for reasonable sized chemical systems. We have shown that introduction of the robust tensor network approximation significantly increases the accuracy of rank-deficient CP approximated tensor networks. This accuracy, in turn, allows one to use relatively low CP rank decompositions and to reduce the cost of electronic structure methods. In Chapter 2, our reduced cost coupled cluster method shows a significant speedup for systems as small as a single water molecule.

It is well understood that optimizing factors of a rank-R CP is a difficult task. In this work, we found the numerical optimization of order-three and order-four integral tensors to be a relatively time consuming process. In relatively low scaling methods, like the coupled

cluster method with single and double substitutions (CCSD), computing optimal CP factor matrices can significantly reduce the effective speedup. To maximize the benefit of using the CP approximation we look towards combining the robust network approximation with grid-based CP initialization. Additionally, we found that, provided a good initial guess, the CP ALS solver can quickly find accurate CP approximations but it can take many iterations to converge CP algorithms tightly. Combining the CP ALS solver algorithm with robust, non-linear optimization methods will allow us to maximize accuracy in the numerical CP approximations.

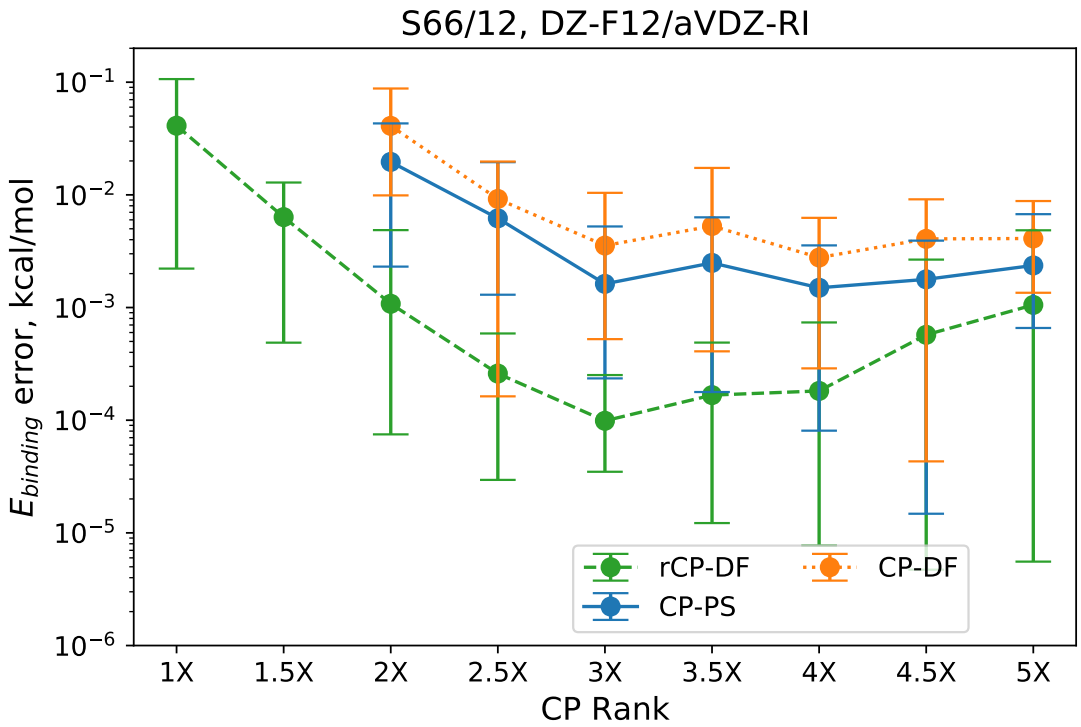
Furthermore, we have shown that CP loss functions can be approximated or modified to reduce the complexity of computing higher-order CP decompositions. We show that one can reduce the complexity of computing the CP decomposition of order-four, Coulomb integral tensors from $\mathcal{O}(N^5)$ to $\mathcal{O}(N^4)$ by introducing the DF approximation into the CP loss function. In Chapter 3, we utilize CP Factor matrices of the Coulomb integral tensor found using this reduced scaling DF approximated CP optimization and in the Laplace transform (LT) perturbative triple substitution (T) correction to CCSD and introduce negligible errors into chemically relevant energy differences. We found that the application of order-three DF approximated integral tensors can be limited, one such instance is pointed out in Section 3.2.4, and these limitations can, in some cases, be circumvented using the order-four CP decomposition of the Coulomb integral tensor. In future work, we hope to use both the order-three and the order-four CP approximations of the Coulomb integral tensor to reduce the computational complexity of other methods, such as the coupled cluster method with single, double and triple substitutions (CCSDT).

Finally, we recognize the cost reduction in this work is relatively modest. However, we believe that reduced-cost CP algorithms will be found beneficial in divide-and-conquer reduced-scaling approaches. The CP decomposition does not depend on electronic sparsity and can be computed for integral tensors of each fragment without worry of adding unphysical artifacts. Particularly, we look look forward to introducing CP approximated integral tensors into the local fragmentation (T) schemes developed by Nagy et al.[166, 177, 182, 183] It is our hope that this work leads the many avenues of electronic structure methods towards reduced cost and scaling algorithms via introduction of the CP decomposition.

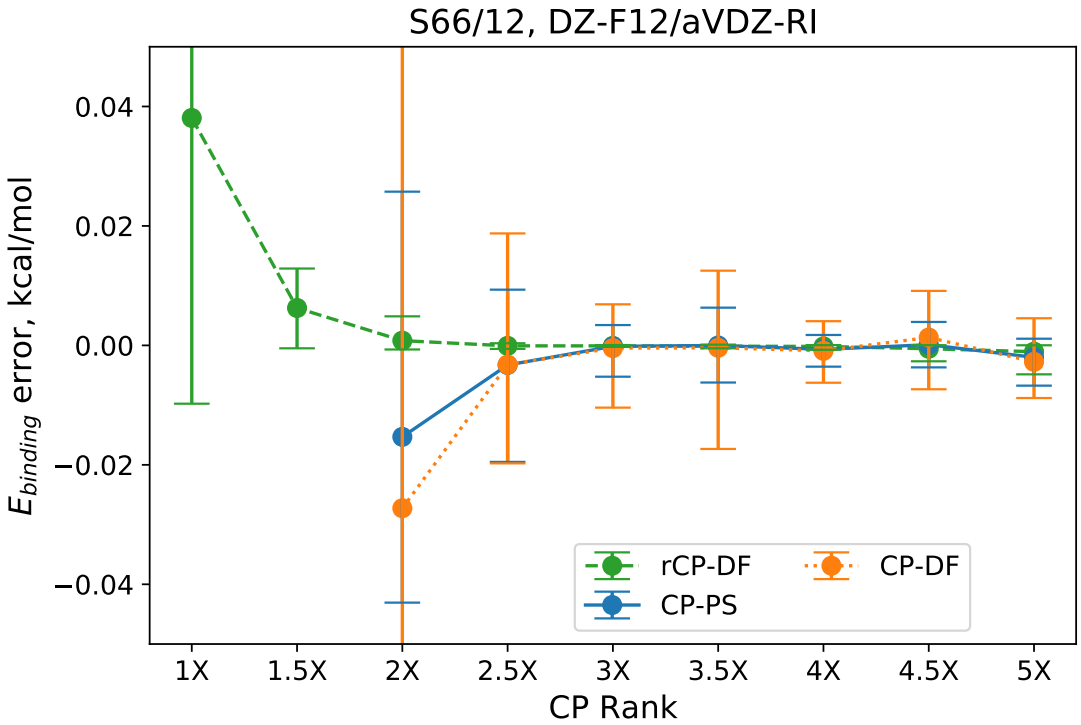
Appendix A

Supporting information for "Robust approximation of tensor networks: application to grid-free tensor factorization of the Coulomb interaction"

Results with $\epsilon = 10^{-4}$

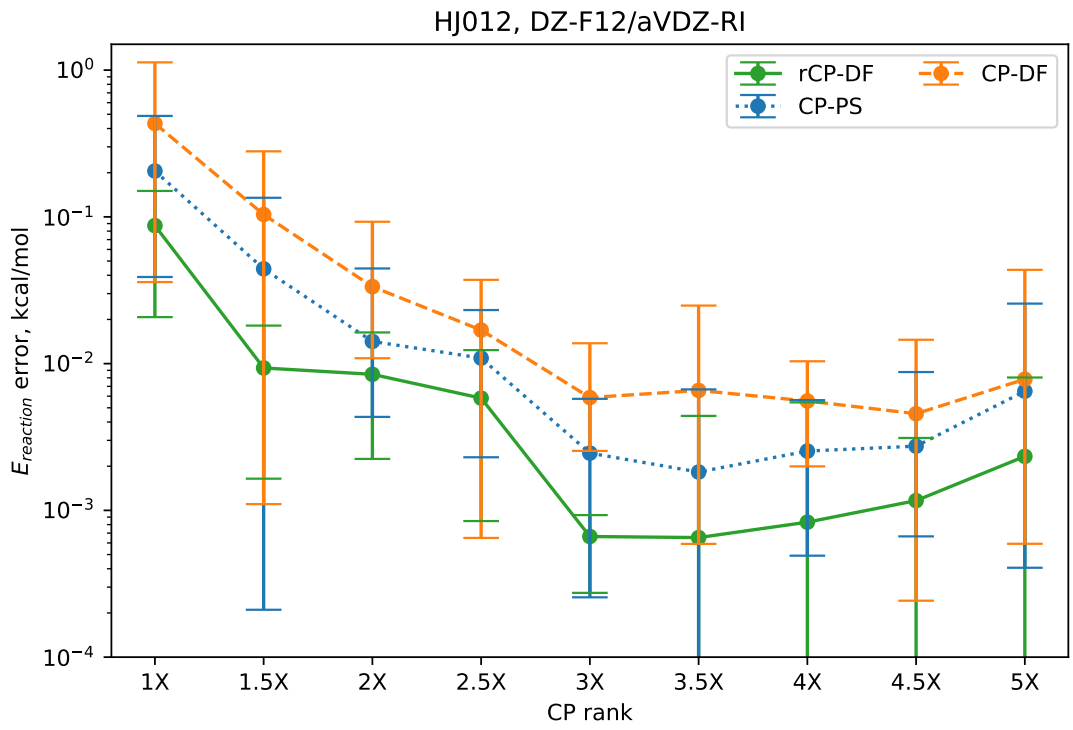


(a)

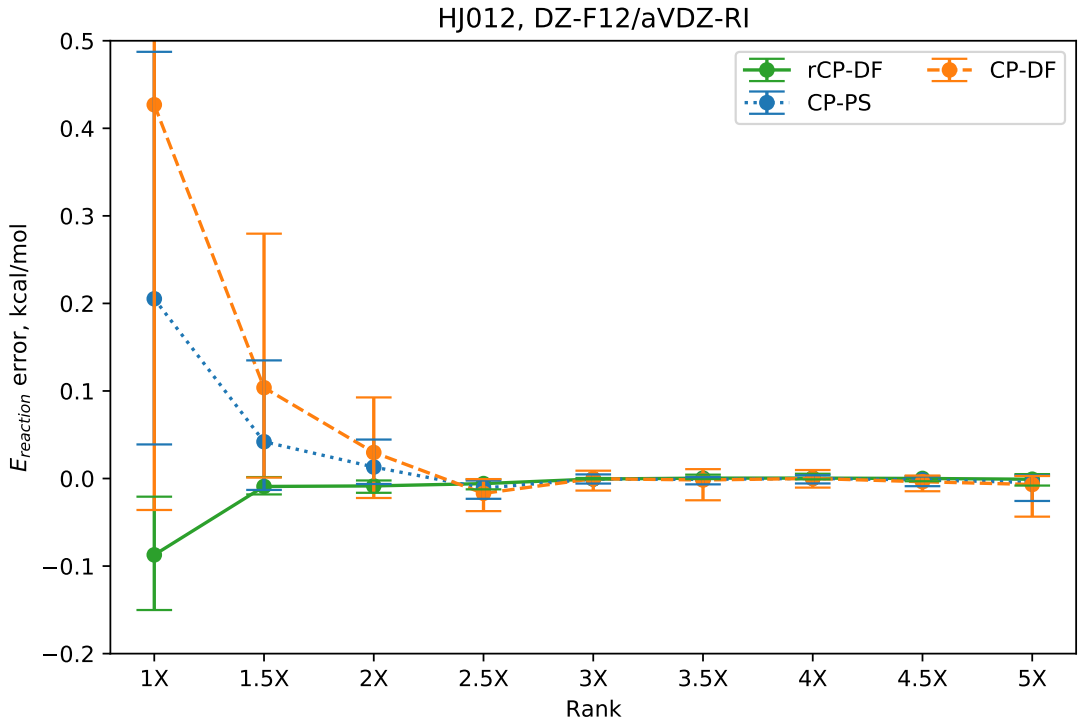


(b)

Figure 4.1: Mean unsigned (a) and signed (b) errors, respectively, in the CCSD binding energies (kcal/mol) of the S66/12 dataset, 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 $\epsilon = 10^{-4}$. The error bars denote the max/min errors.

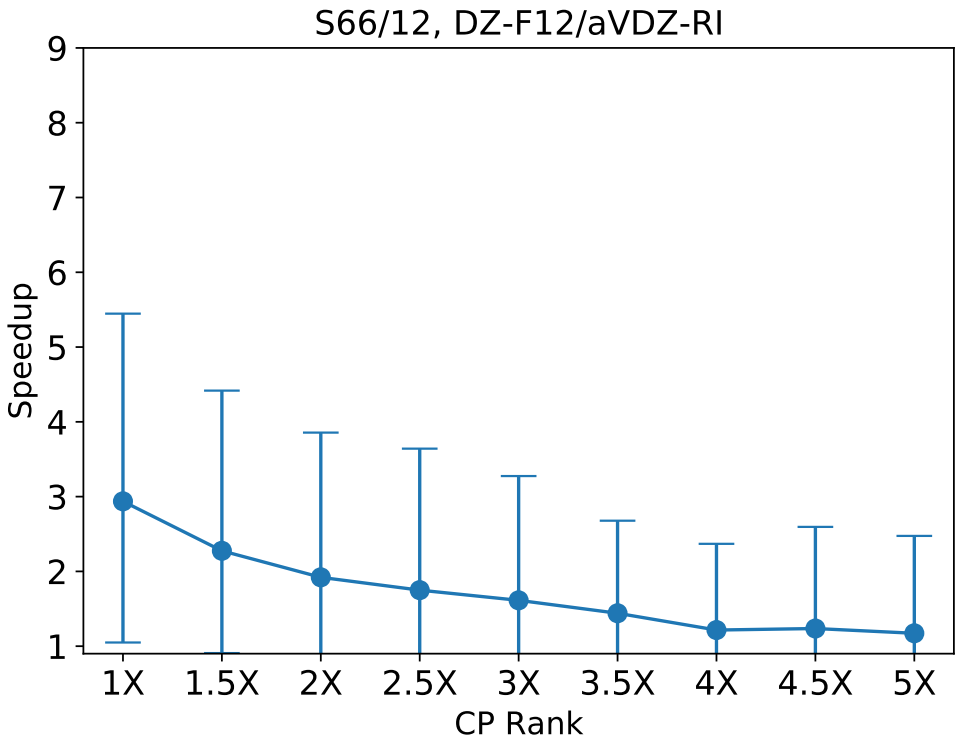


(a)

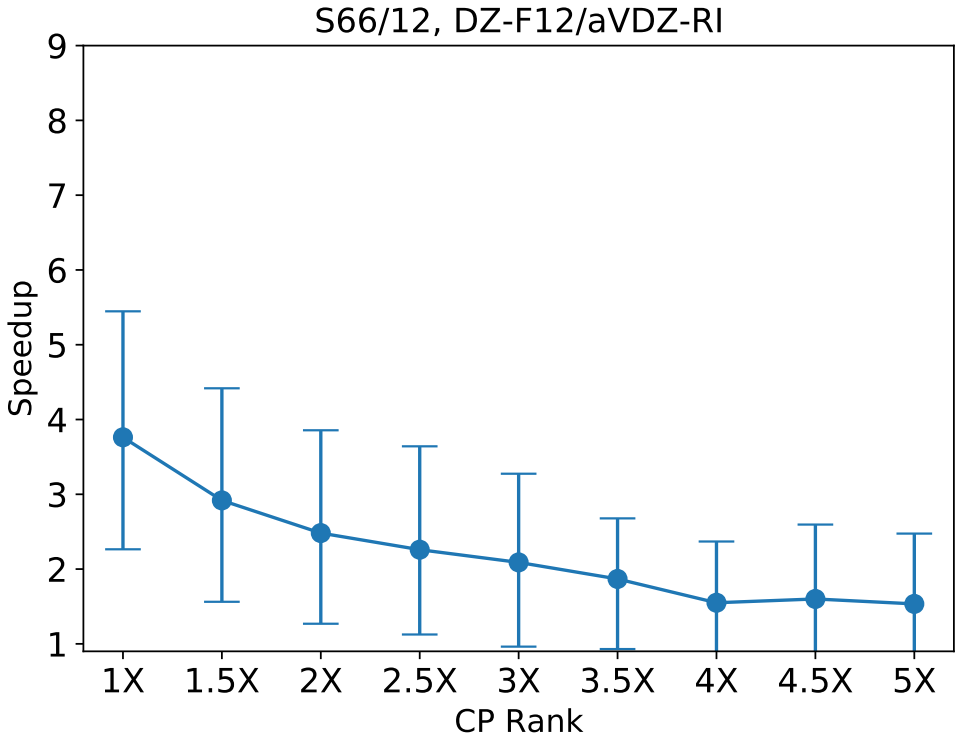


(b)

Figure 4.2: Mean unsigned (a) and signed (b) errors, respectively, in the CCSD reaction energies (kcal/mol) of the HJO12 dataset, 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 $\epsilon = 10^{-4}$. The error bars denote the max/min errors.



(a) Average speedup (Eq. (2.28)) 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 $\epsilon = 10^{-4}$. The error bars denote the max/min speedup.



(b) Average speedup (Eq. (2.28)) 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. ALS precision fixed at $\epsilon = 10^{-4}$. The error bars denote the max/min speedup.

The algorithm to compute g optimal CP-DF.

Algorithm 0: CP-DF optimal $g_{ab,cd}$ factorization

Solve: $\text{argmin}\|B_{ab,X}B_{cd,X} - \hat{B}_{ab,Y}\hat{B}_{cd,Y}\|$

Such that $\hat{B}_{ab,Y} = \sum_r^R \beta_{a,r}\beta_{b,r}\gamma_{Y,r}$

and $\hat{B}_{cd,Y} = \sum_{r'}^{R'} \beta_{c,r'}\beta_{d,r'}\gamma_{Y,r'}$

Result: $\beta_{a,r}$ $\beta_{b,r}$ $\gamma_{Y,r}$ $\beta_{c,r'}$ $\beta_{d,r'}$ $\gamma_{Y,r'}$

initialization:

Provided: R, ϵ_{CP} ;

Set:

$\beta_{a,r}\beta_{b,r}\gamma_{Y,r} \leftarrow argmin(\|B_{ab,Y} - \hat{B}_{ab,Y}\|)$,

$\hat{B}_{cd,Y} \leftarrow \hat{B}_{ab,Y}$,

$\xi^{n+1} \leftarrow \|B_{ab,X}B_{cd,X} - \hat{B}_{ab,Y}\hat{B}_{cd,Y}\|$,

$\xi^n = 0$,

$n = 0$;

while $(\xi^{n+1} - \xi^n) > \epsilon$ **do**

// Compute outer index factor update ;

for $C^n \in \{\beta_{a,r}, \beta_{b,r}\}$ **do**

$M_{ab,Y} \leftarrow B_{ab,X}B_{cd,X}\hat{B}_{cd,Y}$;

if $C^n == \beta_{a,r}$ **then**

$V_{bY,r} \leftarrow \beta_{b,r} \odot \gamma_{Y,r}$; // $\odot$ indicates Khatri-Rao product

else

$V_{aY,r} \leftarrow \beta_{a,r} \odot \gamma_{Y,r}$; // $\odot$ indicates Khatri-Rao product

end

$W \leftarrow V^\top V$;

$C^{n+1} \leftarrow MVW^\dagger$; // $\dagger$ indicates inverse

end

// Compute fitting basis factor update

$K_{ab,r} \leftarrow \beta_{a,r} \odot \beta_{b,r}$;

$V_{Y,Y} \leftarrow \hat{B}_{cd,Y}\hat{B}_{cd,Y}^\top$;

$M_{ab,Y} \leftarrow B_{ab,X}B_{cd,X}\hat{B}_{cd,Y}$;

$W \leftarrow K^\top K$;

$\gamma^{n+1} \leftarrow W^\dagger(K^\top M)V^\dagger$;

// Compute new CP approximation

$\hat{B}_{ab,Y}^{n+1} \leftarrow \sum_r^R \beta_{a,r}^{n+1}\beta_{b,r}^{n+1}\gamma_{Y,r}^{n+1}$;

// Repeat for the other DF factor

$\vdots$

// Check the fit of the approximation

$\xi^n \leftarrow \xi^{n+1}$;

$\xi^{n+1} \leftarrow \|B_{ab,X}B_{cd,X} - \hat{B}_{ab,Y}\hat{B}_{cd,Y}\|$;

$n \leftarrow n + 1$;

end

Appendix B

Supporting information for "Accelerating coupled-cluster perturbative triples via canonical polyadic decomposition of the Hamiltonian"

4.1 Appendix

Accuracy of the quadrature scheme

There exist a number of ways to determine quadrature weights and roots (w_α and s_α) for CCSD(T): using least squares,[74–76] the Euler-McLaurin formula (also known as the radial quadrature scheme in density functional theory [DFT]),[81, 209] the minimax algorithm[210], and the Gauss-Laguerre and Gauss-Legendre quadrature schemes.[78] Because of its simplicity, in this work we utilize the weights and roots determined via Gauss-Legendre integration.

Using Gauss-Legendre integration one introduces a convenient variable substitution, $s = -\frac{1}{\beta}\ln x$, into Eq. (3.18),

$$\int_0^\infty e^{-D_{abc}^{ijk}s}ds = -\frac{1}{\beta}\int_0^1 x^{(D_{abc}^{ijk}\beta^{-1})-1}dx \approx -\frac{1}{\beta}\sum_\alpha^{n_\alpha} w_\alpha x_\alpha^{(D_{abc}^{ijk}\beta^{-1})-1} \tag{4.1}$$

where the parameter β is defined as the triples gap. From here, the quadrature roots and weights, s_α and w_α, are determined using the Gaussian quadrature integration approach. Using this technique, the accuracy of CCSD(T) is dependent on the number of quadrature points.

In Constan's original LT-CCSD(T) publication, researchers studied the energetic accuracy of a set of small molecules versus the number of quadrature points and found that, to achieve an accuracy of several μE_h, at least 4 quadrature points were required. For slightly larger molecules, $(H_2O)_3$ and $(H_2O)_6$, researchers found only 3 quadrature points were required

to achieve μE_h accuracy. Unfortunately, Constan's found that computing LT-CCSD(T) for $(H_2O)_3$ with more than 1 quadrature point and $(H_2O)_6$ with more than 3 quadrature points took more time than a conventional CCSD(T) computation.

4.1.1 LT CCSD(T) equations

In this section, the indices a, b, c and i, j, k are, implicitly, transformed via the Laplace quadrature scheme. Additionally, we split the up the contractions by multiple energy contributions and LT (T) energy is computed per quadrature point, α as

$$\text{energy}_{(\text{T})} += 3.0 * w_\alpha * (2.0 * \text{energy}_{vv} + 2.0 * \text{energy}_{oo} - 4.0 * \text{energy}_{vo} + \text{energy}_{v-s} - \text{energy}_{o-s}); \tag{4.2}$$

energy$_{vv}$

This term has an OV^5 and an O^2V^4 contraction

$$\text{energy}_{vv} += \sum_{ebfa} (\sum_{ck} g_{ck}^{eb} g_{ck}^{fa}) \sum_{ij} (t_{ij}^{ea} * (t_{ji}^{fb} - 2t_{ij}^{fb})) \tag{4.3}$$

With DF, the next two terms have an O^2V^4 cost

$$\text{energy}_{vv} += \sum_{eafb}\sum_{XY}(B_b^{eY}(\sum_{ck} B_k^{cY} B_k^{cX})\sum_{fa} B_a^{fX})\sum_{ij} t_{ij}^{ea} * (4.0t_{ji}^{fb} - 2.0t_{ij}^{fb}) \tag{4.4}$$

$$\text{energy}_{vv} -= \sum_{eafb}\sum_{X}\sum_{ck}(g_{ck}^{eb} B_k^{cX}) B_a^{fX}\sum_{ij} t_{ij}^{ea} * (4.0t_{ji}^{fb} - 2.0t_{ij}^{fb}) \tag{4.5}$$

The next portion has 4 intermediates 2 (G2 and G3) which cost O2V4 and 2 which cost $\mathcal{O}(N^5)$ with DF

$$G1_{ai}^{ef} = \sum_{X} B_a^{eX} \sum_{bj} B_j^{bX} t_{ij}^{fb} \tag{4.6}$$

$$G2_{fi}^{ea} = \sum_{bj} g_{bj}^{ea} t_{ji}^{fb} \tag{4.7}$$

$$G3_{fi}^{ea} = \sum_{bj} g_{bj}^{ea} t_{ij}^{fb} \tag{4.8}$$

$$G4_{ai}^{ef} = \sum_{X} B_a^{eX} \sum_{bj} B_j^{bX} t_{ji}^{fb} \tag{4.9}$$

$$\tag{4.10}$$

These 4 intermediates contribute to the energy as

$$\text{energy}_{vv} += \sum_{eaif} G2^{ea}_{fi}(G2^{fa}_{ei} - 4G3^{fa}_{ei} - 4G4^{fa}_{ei} + 8G1^{fa}_{ei}) \tag{4.11}$$

$$\text{energy}_{vv} += \sum_{eaif} G4^{ef}_{ai} G4^{fe}_{ai} \tag{4.12}$$

$$\text{energy}_{vv} += \sum_{eaif} G3^{ea}_{fi}(2G4^{fe}_{ai} + G3^{fa}_{ei} - 4G1^{fa}_{ei}) \tag{4.13}$$

$$\text{energy}_{vv} += \sum_{eafi} G1^{ef}_{ai}(4G1^{fe}_{ai} - 4G4^{fe}_{ai}) \tag{4.14}$$

and cost $\mathcal{O}(N^4)$.

The next section has 2 intermediate T tensors which each cost O^3V^3 to construct:

$$T1^{ef}_{ji} = \sum_{ak} 4t^{ea}_{jk}t^{fa}_{ik} + \sum_{ak} t^{ea}_{kj}t^{fa}_{ki} - 4\sum_{ak} t^{ea}_{jk}t^{fa}_{ki} \tag{4.15}$$

$$T2^{ef}_{ji} = \sum_{ak} -2t^{ea}_{jk}t^{fa}_{ik} - 2\sum_{ak} t^{ea}_{kj}t^{fa}_{ki} + 2\sum_{ak} t^{ea}_{kj}t^{fa}_{ik}. \tag{4.16}$$

These intermediates are, then, contracted with coulomb integral tensors. Computing these contractions has an O^2V^4 cost:

$$\text{energy}_{vv} += \sum_{efji} T1^{ef}_{ji} \sum_{bc} g^{ec}_{bi} g^{fb}_{cj} \tag{4.17}$$

$$\text{energy}_{vv} += \sum_{efji} T2^{ef}_{ji} \sum_{bc} g^{ec}_{bi} g^{fc}_{bj} \tag{4.18}$$

This is the final portion of energy_{vv} only costs $\mathcal{O}(N^5)$, containing 2 OV^4 and $2O^2V^3$ con-

tractions.

$$\text{energy}_{vv} += \sum_{ef}(\sum_{cbk} g^{ec}_{bk} g^{fc}_{bk})(4\sum_{ija} t^{ea}_{ij} t^{fa}_{ij} - 2\sum_{ija} t^{ea}_{ij} t^{fa}_{ji}) \tag{4.19}$$

$$\text{energy}_{vv} += \sum_{ef}(\sum_{cbk} g^{ec}_{bk} g^{fb}_{ck})(-2\sum_{ija} t^{ea}_{ij} t^{fa}_{ij} + \sum_{ija} t^{ea}_{ij} t^{fa}_{ji}) \tag{4.20}$$

energy$_{oo}$

The first contributions to energy$_{oo}$ have a cost of V^2O^3

$$\text{energy}_{oo} += \sum_{cjk} g^{cj}_{kl} g^{cj}_{km}(4\sum_{abi} t^{ab}_{il} t^{ab}_{im} - 2\sum_{abi} t^{ab}_{il} t^{ba}_{im}) \tag{4.21}$$

$$\text{energy}_{oo} += \sum_{cjk} g^{cj}_{kl} g^{ck}_{jm}(-2\sum_{abi} t^{ab}_{il} t^{ab}_{im} + \sum_{abi} t^{ab}_{il} t^{ba}_{im}) \tag{4.22}$$

The next contributions have a cost of O^4V^2 and O^5V:

$$\text{energy}_{oo} += \sum_{mijn}(\sum_{bc} t^{bc}_{im} t^{cb}_{jn})(4\sum_{ak} g^{aj}_{km} g^{ai}_{kn} + \sum_{ak} g^{ak}_{jm} g^{ak}_{in} - 4.0\sum_{ak} g^{aj}_{km} g^{ak}_{in}) \tag{4.23}$$

$$\text{energy}_{oo} += sum_{mijn}(\sum_{bc} t^{bc}_{im} t^{bc}_{jn})(2\sum_{ak} g^{ak}_{jm} g^{ai}_{kn} - 2.0\sum_{ak} g^{ak}_{jm} g^{ak}_{in} - 2\sum_{ak} g^{aj}_{km} g^{ai}_{kn}) \tag{4.24}$$

In the next section 4 intermediates are constructed 2, $T1$ and $T2$ are constructed in $\mathcal{O}(N^5)$

time using DF while the other 2 cost O^4V^2.

$$T1_{in}^{ma} = \sum_X B_n^{iX} (\sum_{bj} t_{jm}^{ba} B_j^{bX}) \tag{4.25}$$

$$T2_{in}^{ma} = \sum_X B_n^{iX} (\sum_{bj} t_{jm}^{ab} B_j^{bX}) \tag{4.26}$$

$$T3_{in}^{ma} = \sum_{bj} t_{jm}^{ba} g_{in}^{bj} \tag{4.27}$$

$$T4_{in}^{ma} = \sum_{bj} t_{jm}^{ab} g_{in}^{bj}. \tag{4.28}$$

These four intermediates contribute to energy$_{oo}$ using the following equations

$$\text{energy}_{oo} += \sum_{main} T4_{im}^{na} (8T1_{in}^{ma} + T4_{in}^{ma} - 4T3_{on}^{ma}) \tag{4.29}$$

$$\text{energy}_{oo} += \sum_{main} T3_{im}^{na} (2T2_{in}^{ma} + T3_{in}^{ma} - 4T1_{on}^{ma}) \tag{4.30}$$

$$\text{energy}_{oo} += \sum_{main} T1_{im}^{na} (4T1_{in}^{ma}) \tag{4.31}$$

$$\text{energy}_{oo} += \sum_{main} T2_{im}^{na} (2T2_{in}^{ma} - 4T4_{in}^{ma} - 4T1_{on}^{ma}) \tag{4.32}$$

The final contribution to energy$_{oo}$ has a scaling of O^4V^2

$$\text{energy}_{oo} += \sum_{mabn} (\sum_{jk} g_{km}^{bj} g_{jn}^{ak}) (\sum_{ci} 4t_{im}^{ca} t_{in}^{cb} - 4.0 t_{im}^{ca} t_{in}^{bc} + t_{im}^{ac} t_{in}^{bc}) \tag{4.33}$$

$$\text{energy}_{oo} += \sum_{mabn} (\sum_{jk} g_{km}^{bj} g_{kn}^{aj}) (\sum_{ci} -2t_{im}^{ca} t_{in}^{cb} + 2t_{im}^{ac} t_{in}^{cb} - 2t_{im}^{ac} t_{in}^{bc}) \tag{4.34}$$

energy$_{vo}$

First, energy$_{vo}$ has two terms which have a cost of O^2V^4

$$\text{energy}_{vo} += -\sum_{abkl} t_{kl}^{ab} \sum_{ec} g_{ak}^{ec} \sum_{ij} t_{ij}^{eb}(2.0g_{il}^{cj} - g_{jl}^{ci}) \tag{4.35}$$

$$\text{energy}_{vo} += \sum_{abkl} t_{kl}^{ba} \sum_{ec} g_{ak}^{ec} \sum_{ij} t_{ij}^{eb}(g_{il}^{cj} - 2g_{jl}^{ci}) \tag{4.36}$$

and a third term whose cost is reduced from O^2V^4 to O^3V^3 using DF

$$\text{energy}_{vo} += \sum_{Xec} B_c^{eX} \sum_{ck} B_k^{cX}(2t_{kl}^{cb} - t_{kl}^{bc})(\sum_{ij} t_{ij}^{eb}(2.0g_{il}^{cj} - g_{jl}^{ci}) \tag{4.37}$$

The next portion requires the construction of four intermediates which cost O^3V^3 to compute

$$T1_{an}^{ei} = \sum_{bj} t_{ji}^{eb} t_{jn}^{ab} \tag{4.38}$$

$$T2_{an}^{ei} = \sum_{bj} t_{ij}^{eb} t_{jn}^{ba} \tag{4.39}$$

$$T3_{an}^{ei} = \sum_{bj} t_{ji}^{eb} t_{jn}^{ba} \tag{4.40}$$

$$T4_{an}^{ei} = \sum_{bj} t_{ij}^{eb} t_{jn}^{ab} \tag{4.41}$$

These intermediates contribute to energy$_{vo}$ in the following way

$$\text{energy}_{vo} += \sum_{eian} 3(\sum_{X} B_a^{eX}(\sum_{bj} B_j^{bX} g_{jn}^{bi})T1_{an}^{ei} \tag{4.42}$$

$$\text{energy}_{vo} += \sum_{eian}(\sum_{bj} g_{bj}^{ea} g_{in}^{bj})(T1_{an}^{ei} + 4T2_{an}^{ei} - 2T3_{an}^{ei} - 2T4_{an}^{ei}) \tag{4.43}$$

$$\text{energy}_{vo} += \sum_{eian}\left(\sum_{X} B_n^{iX}(\sum_{bj} g_{bj}^{ea} B_j^{bX})) + (\sum_{X} B_a^{eX}(\sum_{bj} B_j^{bX} g_{in}^{bj})\right)(T3_{an}^{ei} + T4_{an}^{ei} - 2T1_{an}^{ei} - 2T2_{an}^{ei}) \tag{4.44}$$

Two of these energy contributions cost N^5 using DF and one costs O^3V^3

For the next portion, again four intermediates are constructed, two of these intermediates cost N^5 using the DF approximation and the other two cost O^3V^3

$$GT1_{jn}^{ei} = \sum_{X} B_n^{iX}(\sum_{ak} t_{jk}^{ea} B_k^{aX}) \tag{4.45}$$

$$GT2_{jn}^{ei} = \sum_{X} B_n^{iX}(\sum_{ak} t_{kj}^{ea} B_k^{aX}) \tag{4.46}$$

$$GT3_{jn}^{ei} = (\sum_{ak} t_{jk}^{ea} g_{in}^{ak}) \tag{4.47}$$

$$GT4_{jn}^{ei} = (\sum_{ak} t_{kj}^{ea} g_{in}^{ak}). \tag{4.48}$$

These intermediates contribute to energy$_{vo}$ using the following two equations and cost O^3V^3:

$$\text{energy}_{vo} += \sum_{eijn}(4GT1_{jn}^{ei} + GT4_{jn}^{ei} - 2GT3_{jn}^{ei} - 2GT2_{jn}^{ei}) \sum_{bc} g_{ci}^{eb} t_{jn}^{cb} \tag{4.49}$$

$$\text{energy}_{vo} += \sum_{eijn}(GT2_{jn}^{ei} + GT3_{jn}^{ei} - 2GT4_{jn}^{ei} - 2GT1_{jn}^{ei}) \sum_{bc} g_{ci}^{eb} t_{jn}^{bc}. \tag{4.50}$$

The final contribution to energy$_{vo}$ has a cost of N^5

$$\text{energy}_{vo} += \sum_{en}(\sum_{bck} g^{eb}_{ck} t^{bc}_{kn})(\sum_{aij} 4t^{ea}_{ij} g^{ai}_{jn} - 2t^{ea}_{ij} g^{aj}_{in}) \tag{4.51}$$

$$\text{energy}_{vo} += \sum_{en}(\sum_{bck} g^{eb}_{ck} t^{cb}_{kn})(\sum_{aij} -2t^{ea}_{ij} g^{ai}_{jn} + t^{ea}_{ij} g^{aj}_{in}) \tag{4.52}$$

energy$_{v-s}$

Here we construct 3 intermediates, 2 which cost O^3V^3 and one that is reduced to $\mathcal{O}(N^5)$ using DF

$$GT1^{fc}_{jk} = \sum_{ai} g^{ac}_{ji} t^{fa}_{ik} \tag{4.53}$$

$$GT2^{cf}_{jk} = \sum_{X} B^{cX}_{j} (\sum_{ai} B^{aX}_{i} t^{fa}_{ki}) \tag{4.54}$$

$$GT3^{fc}_{jk} = \sum_{ai} g^{ac}_{ji} t^{fa}_{ki} \tag{4.55}$$

contributions to energy$_{v-s}$ are computed using the following equations and cost $\mathcal{O}(N^5)$

$$\text{energy}_{v-s} += \sum_{bk} t^{b}_{k} (2 \sum_{fcj} GT1^{fc}_{jk} g^{fc}_{bj}) \tag{4.56}$$

$$\text{energy}_{v-s} += \sum_{bk} t^{b}_{k} (-4 \sum_{fcj} GT1^{fc}_{jk} g^{fb}_{cj}) \tag{4.57}$$

$$\text{energy}_{v-s} += \sum_{kb} t^{b}_{k} (-4 \sum_{cfj} GT2^{cf}_{jk} g^{fb}_{cj}) \tag{4.58}$$

$$\text{energy}_{v-s} += \sum_{kb} t^{b}_{k} (\sum_{fcj} GT3^{fc}_{jk} g^{fc}_{bj}). \tag{4.59}$$

Next, there is a contribution to energy_{v-s} which costs O^3V^3

$$\text{energy}_{v-s} += \sum_{ijkf}(\sum_{bc} g_{ij}^{bc} g_{ck}^{fb})(\sum_{ak} t_k^a(2t_{ij}^{fa} - 4t_{ji}^{fa}) \tag{4.60}$$

The next portion contains two terms both which have an O^3V^3 contraction, a contraction reduced to $\mathcal{O}(N^5)$ using DF and an additional $\mathcal{O}(N^5)$ contraction

$$\text{energy}_{v-s} += \sum_{aijf}(\sum_{X} B_i^{aX}(\sum_{ck} B_k^{cX}(8t_{jk}^{fc} - 4t_{kj}^{fc}))(\sum_{b} g_{bi}^{fa} t_j^b) \tag{4.61}$$

$$\text{energy}_{v-s} += \sum_{aijf}(\sum_{ck} 2g_{ki}^{ac} t_{jk}^{fc} + 2\sum_{X} B_i^{aX}(\sum_{ck} B_k^{cX} t_{kj}^{fc}))(\sum_{b} g_{ai}^{fb} t_j^b) \tag{4.62}$$

The final portion of energy_{v-s} are three terms which cost $\mathcal{O}(N^5)$

$$\text{energy}_{v-s} += \sum_{fi}(\sum_{bck} g_{ck}^{fb}(8g_{ki}^{bc} - 4g_{ik}^{bc})(\sum_{aj} t_{ij}^{fa} t_j^a) \tag{4.63}$$

$$\text{energy}_{v-s} += \sum_{fi}(\sum_{bck} g_{bk}^{fc}(2g_{ki}^{bc} - 4g_{ik}^{bc})(\sum_{aj} t_{ji}^{fa} t_j^a) \tag{4.64}$$

$$energy_{v-s} += \sum_{fa}(\sum_{bij} g_{ij}^{ab}(2t_{ij}^{fb} - t_{ji}^{fb}))\sum_{ck} t_k^c(4g_{ak}^{fc} - 2g_{ck}^{fa}) \tag{4.65}$$

$\mathbf{energy}_{o-s}$

The following contributions all have a computational cost of $\mathcal{O}(N^5)$

$$\text{energy}_{o-s} += \sum_{in}(t_{kn}^{bc}(8g_{ki}^{bc} - 4g_{ik}^{bc})(\sum_{aj} t_j^a g_{jn}^{ai}) \tag{4.66}$$

$$\text{energy}_{o-s} += \sum_{in}(t_{kn}^{cb}(2g_{ki}^{bc} - 4g_{ik}^{bc})(\sum_{aj} t_j^a g_{in}^{aj}) \tag{4.67}$$

$$\text{energy}_{o-s} += \sum_{na}(\sum_{cjk} g_{jk}^{ac} g_{kn}^{cj})(\sum_{bi} t_i^b(8t_{in}^{ba} - 4t_{in}^{ab}) \tag{4.68}$$

$$\text{energy}_{o-s} += \sum_{na}(\sum_{cjk} g_{jk}^{ac} g_{jn}^{ck})(\sum_{bi} t_i^b(2t_{in}^{ab} - 4t_{in}^{ba}) \tag{4.69}$$

The next portions creates two intermediates: 1 costs O^4V^2 to compute and the the other is reduced to $\mathcal{O}(N^5)$ using DF

$$gg_{1}{}_{an}^{ij} = \sum_{bk} g_{ik}^{ba} g_{jn}^{bk} \tag{4.70}$$

$$gg_{2}{}_{an}^{ij} = \sum_{X} B_i^{aX} \sum_{bk} B_k^{bX} g_{kn}^{bj} \tag{4.71}$$

These intermediates contribute to the energy as follows

$$\text{energy}_{o-s} += \sum_{ijan}(8gg_{2}{}_{an}^{ij} + 2gg_{1}{}_{an}^{ij} - \sum_{X} B_n^{jX} \sum_{bk} B_k^{bX} g_{ik}^{ba})(\sum_{c} t_j^c t_{in}^{ac}) \tag{4.72}$$

$$\text{energy}_{o-s} += \sum_{ijan}(-4gg_{2}{}_{an}^{ij} - 4gg_{1}{}_{an}^{ij}(\sum_{c} t_j^c t_{in}^{ca}) \tag{4.73}$$

The next contribution has a cost of O^4V^2

$$\text{energy}_{o-s} += -4 * \sum_{ijkn} (\sum_{ab} g_{ij}^{ab} t_{kn}^{ab})(\sum_{c} t_k^c g_{in}^{cj}) \tag{4.74}$$

Of the final contributions to energy_{o-s} one has a cost of O^3V^3, one has a cost of O^4V^2 and one has a cost reduced to $\mathcal{O}(N^5)$ using the DF approximation

$$\text{energy}_{o-s} += \sum_{abin} 2.0 * (\sum_{cj} g_{ij}^{ca} t_{jn}^{bc})(\sum_{k} t_k^b g_{in}^{ak}) \tag{4.75}$$

$$\text{energy}_{o-s} += \sum_{abin} (\sum_{X} B_i^{aX} (\sum_{cj} B_j^{cX} (2t_{jn}^{bc} - 4t_{jn}^{cb})))(\sum_{k} t_k^b g_{kn}^{ai}) \tag{4.76}$$

$$\text{energy}_{o-s} += \sum_{ck} t_k^c (\sum_{ijn} g_{jn}^{ci} (\sum_{ab} g_{ij}^{ab} t_{kn}^{ab})) \tag{4.77}$$

Bibliography

Bibliography

(1) Calvin, J. A.; Peng, C.; Rishi, V.; Kumar, A.; Valeev, E. F. *Chem. Rev.* **2021**, *121*, 1203–1231.

(2) Goedecker, S.; Scuseria, G. *Comput. Sci. Eng.* **2003**, *5*, 14–21.

(3) Bellman, R. E., *Adaptive Control Processes*; Princeton University Press: 1961.

(4) Beebe, N. H. F.; Linderberg, J. *Int. J. Quantum Chem.* **1977**, *12*, 683–705.

(5) Pierce, K.; Rishi, V.; Valeev, E. F. *J. Chem. Theory Comput.* **2021**, *17*, 2217–2230.

(6) Schrodinger, E. *Phys. Rev.* **1926**, *28*, 1049–1070.

(7) Sherrill, C. D.; Schaefer, H. F. *Adv. Quantum Chem.* **1999**, *34*, 143–269.

(8) Hartree, D. R. *Math. Proc. Camb. Philos. Soc.* **1928**, *24*, 89–110.

(9) Hartree, D. R. *Math. Proc. Camb. Philos. Soc.* **1928**, *24*, 111–132.

(10) Moller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618–622.

(11) Head-Gordon, M.; Pople, J. A.; Frisch, M. J. *Chem. Phys. Lett.* **1988**, *153*, 503–506.

(12) Purvis, G. D.; Bartlett, R. J. *J. Chem. Phys.* **1982**, *76*, 1910.

(13) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**, *157*, 479–483.

(14) Noga, J.; Bartlett, R. J. *J. Chem. Phys.* **1987**, *86*, 7041–7050.

(15) Born, M.; Oppenheimer, R. *Ann. Phys.* **1927**, *389*, 457–484.

(16) Roothaan, C. C. J. *Rev. Mod. Phys.* **1951**, *23*, 69–89.

(17) Slater, J. C. *Phys. Rev.* **1928**, *32*, 339–348.

(18) Szabo, A.; Ostlund, N. L., *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; Free Press ; New York, 1996.

(19) Surjan, P. R. In *Second Quantized Approach to Quantum Chem.* Springer Berlin Heidelberg: Berlin, Heidelberg, 1989, pp 87–92.

(20) Edmiston, C.; Ruedenberg, K. *Rev. Mod. Phys.* **1963**, *35*, 457–464.

(21) Strout, D. L.; Scuseria, G. E. *J. Chem. Phys.* **1995**, *102*, 8448–8452.

(22) Kutzelnigg, W. In *Explicitly Correlated Wave Functions in Chemistry and Physics*; Springer Netherlands: Dordrecht, 2003, pp 3–90.

(23) Kratochvil, M.; Engkvist, O.; Sponer, J.; Jungwirth, P.; Hobza, P. *J. Phys. Chem. A* **1998**, *102*, 6921–6926.

(24) Czyznikowska, Z.; Zalesny, R.; Ziolkowski, M.; Gora, R. W.; Cysewski, P. *Chem. Phys. Lett.* **2007**, *450*, 132–137.

(25) Leininger, M. L.; Nielsen, I. M. B.; Colvin, M. E.; Janssen, C. L. *J. Phys. Chem. A* **2002**, *106*, 3850–3854.

(26) He, X.; Fusti-Molnar, L.; Cui, G.; Merz, K. M. *J. Phys. Chem. B* **2009**, *113*, 5290–5300.

(27) Lowdin, P.-O. *Int. J. Quantum Chem.* **1995**, *55*, 77–102.

(28) Kong, L.; Bischoff, F. A.; Valeev, E. F. *Chem. Rev.* **2012**, *112*, 75–107.

(29) Brueckner, K. A.; Goldstone, J.; Levinson, C. A. *Proc. R. Soc. Math. Phys. Eng. Sci.* **1957**, *239*, 267–279.

(30) Freed, K. F. *Phys. Rev.* **1968**, *173*, 1–24.

(31) Freed, K. F. *Annu. Rev. Phys. Chem.* **1971**, *22*, 313–346.

(32) Cremer, D. *Ltd. WIREs Comput Mol Sci* **2011**, *1*, 509–530.

(33) Čížek, J. *J. Chem. Phys.* **1966**, *45*, 4256–4266.

(34) Cizek, J. In *Adv. Chem. Phys.* John Wiley and Sons, Inc.: Hoboken, NJ, USA, 2007, pp 35–89.

(35) Cizek, J.; Paldus, J. *Int. J. Quantum Chem.* **1971**, *5*, 359–379.

(36) Crawford, T. D.; Schaefer, H. F. *Rev. Comput. Chem.* **2000**.

(37) Tajti, A.; Szalay, P. G.; Császár, A. G.; Kállay, M.; Gauss, J.; Valeev, E. F.; Flowers, B. A.; Vázquez, J.; Stanton, J. F. *J. Chem. Phys.* **2004**, *121*, 11599–11613.

(38) Harding, M. E.; Vázquez, J.; Ruscic, B.; Wilson, A. K.; Gauss, J.; Stanton, J. F. *J. Chem. Phys.* **2008**, *128*, 114111.

(39) Thorpe, J. H.; Lopez, C. A.; Nguyen, T. L.; Baraban, J. H.; Bross, D. H.; Ruscic, B.; Stanton, J. F. *J. Chem. Phys.* **2019**, *150*, 224102.

(40) Bartlett, R.; Musiał, M. *Rev. Mod. Phys.* **2007**, *79*, 291–352.

(41) Urban, M.; Noga, J.; Cole, S. J.; Bartlett, R. J. *J. Chem. Phys.* **1998**, *83*, 4041.

(42) Noga, J.; Bartlett, R. J.; Urban, M. *Chem. Phys. Lett.* **1987**, *134*, 126–132.

(43) Lee, Y. S.; Kucharski, S. A.; Bartlett, R. J. *J. Chem. Phys.* **1998**, *81*, 5906.

(44) Kolda, T. G.; Bader, B. W. *SIAM Rev.* **2009**, *51*, 455–500.

(45) Doncaster, P. TIMELINE OF THE HUMAN CONDITION, 2021.

(46) Fatica, A.; Bozzoni, I. *Nat. Rev. Genet.* **2013**, *15*, 7–21.

(47) Epifanovsky, E.; Zuev, D.; Feng, X.; Khistyaev, K.; Shao, Y.; Krylov, A. I. *J. Chem. Phys.* **2013**.

(48) Zhang, T.; Liu, X.; Valeev, E. F.; Li, X. *J. Phys. Chem. A* **2021**, *125*, 4258–4265.

(49) Folkestad, S. D.; Kjønstad, E. F.; Koch, H. *J. Chem. Phys.* **2019**, *150*, 194112.

(50) Dunlap, B. I.; Connolly, J. W. D.; Sabin, J. R. *J. Chem. Phys.* **1979**, *71*, 4993.

(51) Dunlap, B. I.; Connolly, J. W.; Sabin, J. R. *Int. J. Quantum Chem.* **1977**, *12*, 81–87.

(52) Baerends, E. J.; Ellis, D. E.; Ros, P. *Chem. Phys.* **1973**, *2*, 41–51.

(53) Tucker, L. R. *Psychometrika* **1966**, *31*, 279–311.

(54) Grasedyck, L.; Kressner, D.; Tobler, C. *GAMM-Mitteilungen* **2013**, *36*, 53–78.

(55) Oseledets, I. V.; Tyrtyshnikov, E. E. *SIAM J. Sci. Comput.* **2009**, *31*, 3744–3759.

(56) Hastad, J. *Algorithms* **1990**, *11*, 644–654.

(57) Hillar, C. J.; Lim, L.-H. *J. ACM* **2013**, *60*, 1–39.

(58) Acar, E.; Dunlavy, D. M.; Kolda, T. G. *J. Chemom* **2011**, *25*, 67–86.

(59) Sorber, L.; Van Barel, M.; De Lathauwer, L. *SIAM J. Optim.* **2013**, *23*, 695–720.

(60) Phan, A. H.; Tichavsky, P.; Cichocki, A. *IEEE Trans. Signal Process* **2013**, *61*, 4834–4846.

(61) Friesner, R. A. *Chem. Phys. Lett.* **1985**, *116*, 39–43.

(62) Friesner, R. A. *J. Chem. Phys.* **1986**, *85*, 1462–1468.

(63) Friesner, R. A. *J. Phys. Chem.* **1988**, *92*, 3091–3096.

(64) Ringnalda, M. N.; Belhadj, M.; Friesner, R. A. *J. Chem. Phys.* **1990**, *93*, 3397–3407.

(65) Greeley, B. H.; Russo, T. V.; Mainz, D. T.; Friesner, R. A.; Langlois, J.-M.; Goddard, W. A.; Donnelly, R. E.; Ringnalda, M. N. *J. Chem. Phys* **1994**, *101*, 4028.

(66) Hohenstein, E. G.; Parrish, R. M.; Martínez, T. J. *J. Chem. Phys.* **2012**, *137*, 044103.

(67) Hohenstein, E. G.; Parrish, R. M.; Sherrill, C. D.; Martínez, T. J. *J. Chem. Phys.* **2012**, *137*, 221101.

(68) Parrish, R. M.; Hohenstein, E. G.; Martínez, T. J.; Sherrill, C. D. *J. Chem. Phys.* **2012**, *137*, 224106.

(69) Parrish, R. M.; Hohenstein, E. G.; Schunck, N. F.; Sherrill, C. D.; Martínez, T. J. *Phys. Rev. Lett.* **2013**, *111*, 132505.

(70) Hohenstein, E. G.; Kokkila, S. I. L.; Parrish, R. M.; Martinez, T. J. *J. Chem. Phys.* **2013**, *138*, 124111.

(71) Parrish, R. M.; Sherrill, C. D.; Hohenstein, E. G.; Kokkila, S. I. L.; Martínez, T. J. *J. Chem. Phys.* **2014**, *140*, 181102.

(72) Kokkila Schumacher, S. I. L.; Hohenstein, E. G.; Parrish, R. M.; Wang, L.-P.; Martínez, T. J. *J. Chem. Theory Comput.* **2015**, *11*, 3042–3052.

(73) Schutski, R.; Zhao, J.; Henderson, T. M.; Scuseria, G. E. *J. Chem. Phys.* **2017**, *147*, 184113.

(74) Almlof, J. *Chem. Phys. Lett.* **1991**, *181*, 319–320.

(75) Haser, M.; Almlof, J. *J. Chem. Phys* **1998**, *96*, 489.

(76) Häser, M. *Theoret. Chim. Acta* **1993**, *87*, 147–173.

(77) Wilson, A. K.; Almlof, J. *Theor. Chim. Acta* **1997**, *95*, 49–62.

(78) Constans, P.; Ayala, P. Y.; Scuseria, G. E. *J. Chem. Phys* **2000**, *113*, 10451.

(79) Whitten, J. L. *J. Chem. Phys.* **1973**, *58*, 4496–4501.

(80) Vahtras, O.; Almlof, J.; Feyereisen, M. W. *Chem. Phys. Lett.* **1993**, *213*, 514–518.

(81) Scuseria, G. E.; Ayala, P. Y. *J. Chem. Phys.* **1999**, *111*, 8330–8343.

(82) Saebo, S.; Pulay, P. *Annu. Rev. Phys. Chem* **1993**, *44*, 213–236.

(83) Hampel, C.; Werner, H.-J. *J. Chem. Phys.* **1996**, *104*, 6286–6297.

(84) Langlois, J.-M.; Muller, R. P.; Coley, T. R.; Goddard, W. A.; Ringnalda, M. N.; Won, Y.; Friesner, R. A. *J. Chem. Phys.* **1990**, *92*, 7488–7497.

(85) Friesner, R. A. *Annu. Rev. Phys. Chem.* **1991**, *42*, 341–367.

(86) Martinez, T. J.; Carter, E. A. *J. Chem. Phys.* **1995**, *102*, 7564–7572.

(87) Martinez, T. J.; Mehta, A.; Carter, E. A. *J. Chem. Phys.* **1992**, *97*, 1876–1880.

(88) Martinez, T. J.; Carter, E. A. *J. Chem. Phys.* **1994**, *100*, 3631–3638.

(89) Ko, C.; Malick, D. K.; Braden, D. A.; Friesner, R. A.; Martínez, T. J. *J. Chem. Phys.* **2008**, *128*, 104103.

(90) Martinez, T. J.; Carter, E. A. *J. Chem. Phys.* **1993**, *98*, 7081–7085.

(91) Löwdin, P.-O. *J. Math. Phys.* **1965**, *6*, 1341–1353.

(92) Lowdin, P.-O. *Int. J. Quantum Chem.* **2009**, *5*, 231–237.

(93) Folkestad, S. D.; Kjønstad, E. F.; Koch, H. *J. Chem. Phys.* **2019**, *150*, 194112.

(94) White, C. A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M. *Chem. Phys. Lett.* **1994**, *230*, 8–16.

(95) Burant, J. C.; Scuseria, G. E.; Frisch, M. J. *J. Chem. Phys.* **1996**, *105*, 8969–8972.

(96) Rudberg, E.; Salek, P. *J. Chem. Phys.* **2006**, *125*, 084106.

(97) Shenvi, N.; Van Aggelen, H.; Yang, Y.; Yang, W.; Schwerdtfeger, C.; Mazziotti, D. *J. Chem. Phys.* **2013**, *139*, 54110.

(98) Parrish, R. M.; Zhao, Y.; Hohenstein, E. G.; Martinez, T. J. *J. Chem. Phys.* **2019**, *150*, 164118.

(99) Lee, J.; Lin, L.; Head-Gordon, M. *J. Chem. Theory Comput.* **2020**, *16*, 243–263.

(100) Carroll, J. D.; Chang, J.-J. *Psychometrika* **1970**, *35*, 283–319.

(101) Harshman, R. a. *WPP* **1970**, *16*, 1–84.

(102) Benedikt, U.; Auer, A. A.; Espig, M.; Hackbusch, W. *J. Chem. Phys.* **2011**, *134*, 054118.

(103) Benedikt, U.; Bohm, K.-H.; Auer, A. A. *J. Chem. Phys.* **2013**, *139*, 224101.

(104) Benedikt, U.; Auer, H.; Espig, M.; Hackbusch, W.; Auer, A. *Mol. Phys.* **2013**, *111*, 2398–2413.

(105) Hummel, F.; Tsatsoulis, T.; Grüneis, A. *J. Chem. Phys.* **2017**, *146*, 124105.

(106) Böhm, K. H.; Auer, A. A.; Espig, M. *J. Chem. Phys.* **2016**, *144*, 244102.

(107) Chinnamsetty, S. R.; Espig, M.; Khoromskij, B. N.; Hackbusch, W.; Flad, H.-J. *J. Chem. Phys.* **2007**, *127*, 084110.

(108) Khoromskij, B.; Khoromskaia, V.; Chinnamsetty, S.; Flad, H.-J. *J. Comput. Phys.* **2009**, *228*, 5749–5762.

(109) Lewis, C. A.; Calvin, J. A.; Valeev, E. F. *J. Chem. Theory Comput.* **2016**, *12*, 5868–5880.

(110) Bischoff, F. A.; Valeev, E. F. *J. Chem. Phys.* **2011**, *134*, 104104.

(111) Füsti-Molnár, L.; Pulay, P. *J. Chem. Phys.* **2002**, *117*, 7827–7835.

(112) Dutta, A. K.; Neese, F.; Izsák, R. *J. Chem. Phys.* **2016**, *145*, 034102.

(113) Izsák, R.; Neese, F.; Klopper, W. *J. Chem. Phys.* **2013**, *139*, 094111.

(114) Izsak, R.; Hansen, A.; Neese, F. *Mol. Phys.* **2012**, *110*, 2413–2417.

(115) Izsák, R.; Neese, F. *J. Chem. Phys.* **2011**, *135*, 144105.

(116) Petrenko, T.; Kossmann, S.; Neese, F. *J. Chem. Phys.* **2011**, *134*, 054116.

(117) Kossmann, S.; Neese, F. *J. Chem. Theory Comput.* **2010**, *6*, 2325–2338.

(118) Neese, F.; Wennmohs, F.; Hansen, A.; Becker, U. *Chem. Phys.* **2009**, *356*, 98–109.

(119) Sun, X.; Pitsianis, N. P. *SIAM Rev.* **2001**, *43*, 289–300.

(120) Börm, S.; Grasedyck, L.; Hackbusch, W. *Eng. Anal. Bound. Elem.* **2003**, *27*, 405–422.

(121) Dunlap, B. I. *Phys. Chem. Chem. Phys. PCCP* **2000**, *2*, 2113–2116.

(122) Goldfarb, D.; Qin, Z. T. *SIAM J. Matrix Anal. Appl.* **2014**, *35*, 225–253.

(123) Reine, S.; Tellgren, E.; Krapp, A.; Kjargaard, T.; Helgaker, T.; Jansik, B.; Host, S.; Salek, P. *J. Chem. Phys.* **2008**, *129*, 104101.

(124) Merlot, P.; Kjærgaard, T.; Helgaker, T.; Lindh, R.; Aquilante, F.; Reine, S.; Pedersen, T. B. *J. Comput. Chem.* **2013**, *34*, 1486–1496.

(125) Kossmann, S.; Neese, F. *Chem. Phys. Lett.* **2009**, *481*, 240–243.

(126) Ten-no, S. *J. Chem. Phys.* **2004**, *121*, 117.

(127) Peng, B.; Kowalski, K. *J. Chem. Theory Comput.* **2017**, *13*, 4179–4192.

(128) Motta, M.; Shee, J.; Zhang, S.; Chan, G. K.-L. *J. Chem. Theory Comput.* **2019**, *15*, 3510–3521.

(129) Mardirossian, N.; McClain, J. D.; Chan, G. K.-L. *J. Chem. Phys.* **2018**, *148*, 044106.

(130) Kroonenberg, P. M.; de Leeuw, J. *Psychometrika* **1980**, *45*, 69–97.

(131) Beylkin, G.; Mohlenkamp, M. J. *Proc. Natl. Acad. Sci.* **2002**, *99*, 10246–10251.

(132) Uschmajew, A. *SIAM J. Matrix Anal. Appl.* **2012**, *33*, 639–652.

(133) Navasca, C.; De Lathauwer, L.; Kindermann, S. In *2008 16th European Signal Processing Conference*, 2008, pp 1–5.

(134) Řezáč, J.; Riley, K. E.; Hobza, P. *J. Chem. Theory Comput.* **2011**, *7*, 2427–2438.

(135) Rezac J.; Jurecka P.; Riley K. E.; Cerny J.; Valdes H.; Pluhackova K.; Berka K.; Rezac T.; Pitonak M.; Vondrasek J.; Hobz, P. *Collect. Czechoslov. Chem. Commun.* **2008**, *73*, 1261–1270.

(136) Helgaker, T.; Jørgensen, P.; Olsen, J., *Molecular Electronic-Structure Theory*, First; Helgaker/Molecular Electronic-Structure Theory; John Wiley & Sons, Ltd: Chichester, UK, 2000.

(137) Zhang, J.; Valeev, E. F. *J. Chem. Theory Comput.* **2012**, *8*, 3175–3186.

(138) Bates, D. M.; Smith, J. R.; Tschumper, G. S. *J. Chem. Theory Comput.* **2011**, *7*, 2753–2760.

(139) Jørgensen, P.; Simons, J. *J. Chem. Phys.* **1983**, *79*, 334–357.

(140) Wales, D. J.; Hodges, M. P. *Chem. Phys. Lett.* **1998**, *286*, 65–72.

(141) Peterson, K. A.; Adler, T. B.; Werner, H.-J. *J. Chem. Phys.* **2008**, *128*, 084102.

(142) Weigend, F.; Köhn, A.; Hättig, C. *J. Chem. Phys.* **2002**, *116*, 3175.

(143) Jr., T. H. D. *J. Chem. Phys* **1998**, *90*, 1007.

(144) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796–6806.

(145) BTAS Library, Accessed: 2019-04-09, 2021.

(146) 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. *J. Chem. Phys.* **2020**, *153*, 044120.

(147) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999**, *313*, 701–706.

(148) Li, W.; Piecuch, P.; Gour, J. R.; Li, S. *J. Chem. Phys* **2009**, *131*, 114109.

(149) Kristensen, K.; Ziolkowski, M.; Jansik, B.; Kjaergaard, T.; Jorgensen, P. *J. Chem. Theory Comput.* **2011**, *7*, 1677–1694.

(150) Kjærgaard, T.; Baudin, P.; Bykov, D.; Kristensen, K.; Jørgensen, P. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2017**, *9*, e1319.

(151) Friedrich, J.; Hanrath, M.; Dolg, M. *J. Chem. Phys.* **2007**, *126*, 154110.

(152) Watts, J. D. *Parallel Comput.* **2000**, *26*, 857–867.

(153) Peng, C.; Calvin, J. A.; Valeev, E. F. *Int. J. Quantum Chem.* **2019**, *119*, e25894.

(154) Rendell, A. P.; Lee, T. J.; Komornicki, A.; Wilson, S. *Theor. Chim. Acta* **1993**, *84*, 271–287.

(155) Kobayashi, R.; Rendell, A. P. *Chem. Phys. Lett.* **1997**, *265*, 1–11.

(156) Aprà, E.; Rendell, A. P.; Harrison, R. J.; Tipparaju, V.; deJong, W. A.; Xantheas, S. S. In *The Conference*, ACM Press: New York, New York, USA, 2009, p 1.

(157) Hirata, S. *J. Phys. Chem. A* **2003**, *107*, 9887–9897.

(158) Apra, E.; Klemm, M.; Kowalski, K. In *SC14: International Conference for High Performance Computing, Networking, Storage and Analysis*, IEEE: 2014, pp 674–684.

(159) Olson, R.; Bentz, J.; Kendall, R.; Schmidt, M.; Gordon, M. *J. Chem. Theory Comput.* **2007**, *3*, 1312.

(160) Bentz, J. L.; Olson, R. M.; Gordon, M. S.; Schmidt, M. W.; Kendall, R. A. *Comput. Phys. Commun.* **2007**, *176*, 589–600.

(161) Asadchev, A.; Gordon, M. S. *J. Chem. Theory Comput.* **2013**, *9*, 3385–3392.

(162) Janowski, T.; Pulay, P. *J. Chem. Theory Comput.* **2008**, *4*, 1585–1592.

(163) Lotrich, V.; Flocke, N.; Ponton, M.; Yau, A. D.; Perera, A.; Deumens, E.; Bartlett, R. J. *J. Chem. Phys.* **2008**, *128*, 194104.

(164) Anisimov, V. M.; Bauer, G. H.; Chadalavada, K.; Olson, R. M.; Glenski, J. W.; Kramer, W. T. C.; Aprà, E.; Kowalski, K. *J. Chem. Theory Comput.* **2014**, *10*, 4307–4316.

(165) Shen, T.; Zhu, Z.; Zhang, I. Y.; Scheffler, M. *J. Chem. Theory Comput.* **2019**, DOI: 10.1021/acs.jctc.8b01294.

(166) Gyevi-Nagy, L.; Kappay, M.; Nagy, P. R. *J. Chem. Theory Comput.* **2019**, DOI: 10.1021/acs.jctc.9b00957.

(167) Datta, D.; Gordon, M. S. *J. Chem. Theory Comput.* **2021**, acs.jctc.1c00389.

(168) Ma, W.; Krishnamoorthy, S.; Villa, O.; Kowalski, K. In *2010 IEEE International Conference on Cluster Computing*, 2010, pp 207–216.

(169) Ma, W.; Krishnamoorthy, S.; Villa, O.; Kowalski, K. *J. Chem. Theory Comput.* **2011**, *7*, 1316–1327.

(170) Kim, J.; Sukumaran-Rajam, A.; Hong, C.; Panyala, A.; Srivastava, R. K.; Krishnamoorthy, S.; Sadayappan, P. In *The 2018 International Conference*, ACM Press: New York, New York, USA, 2018, pp 96–106.

(171) Kim, J.; Panyala, A.; Peng, B.; Kowalski, K.; Sadayappan, P.; Krishnamoorthy, S. *Int. Conf. High Perform. Comput. Netw. Storage Anal. SC* **2020**, *2020-November*, DOI: 10.1109/SC41405.2020.00083.

(172) Schütz, M.; Werner, H.-J. *Chem. Phys. Lett.* **2000**, *318*, 370–378.

(173) Schütz, M. *J. Chem. Phys.* **2000**, *113*, 9986–10001.

(174) Maslen, P.; Lee, M.; Head-Gordon, M. *Chem. Phys. Lett.* **2000**, *319*, 205–212.

(175) Schütz, M. *J. Chem. Phys.* **2002**, *116*, 8772–8785.

(176) Maslen, P. E.; Dutoi, A. D.; Lee *, M. S.; Shao, Y.; Head-Gordon, M. *Mol. Phys.* **2005**, *103*, 425–437.

(177) Nagy, L.; Kailay, M.; Peter, P.; Nagy, R. *J. Chem. Theory Comput* **2021**, *17*, 878.

(178) Schutz, M.; Yang, J.; Chan, G. K.-L.; Manby, F. R.; Werner, H.-J. *J. Chem. Phys* **2013**, *138*, 054109.

(179) Riplinger, C.; Neese, F. *J. Chem. Phys.* **2013**, *138*, 034106.

(180) Riplinger, C.; Sandhoefer, B.; Hansen, A.; Neese, F. *J. Chem. Phys.* **2013**, *139*, 134101.

(181) Rolik, Z.; Szegedy, L.; Ladjanszki, I.; Ladoczki, B.; Kallay, M. *J Chem. Phys* **2013**, *139*, 094105.

(182) Nagy, P. R.; Kállay, M. *J. Chem. Phys.*

(183) Nagy, P. R.; Samu, G.; Kállay, M. *J. Chem. Theory Comput.* **2018**, *14*, 4193–4215.

(184) Ma, Q.; Werner, H.-J. *J. Chem. Theory Comput.* **2018**, *14*, 198–215.

(185) Schmitz, G.; Hattig, C. *J. Chem. Theory Comput.* **2017**, *13*, 2623–2633.

(186) Schmitz, G.; Hättig, C. *J. Chem. Phys.* **2016**, *145*, 234107.

(187) Guo, Y.; Riplinger, C.; Becker, U.; Liakos, D. G.; Minenkov, Y.; Cavallo, L.; Neese, F. *J. Chem. Phys.* **2018**, *148*, 011101.

(188) Guo, Y.; Becker, U.; Neese, F. *J. Chem. Phys.* **2018**, *148*, 124117.

(189) Calbo, J.; Sancho-Garcia, J. C.; Orti, E.; Arago, J. *J. Comput. Chem.* **2017**, *38*, 1869–1878.

(190) Werner, H.-J.; Schütz, M. *J. Chem. Phys.* **2011**, *135*, 144116.

(191) Koch, H.; de Meras, A. S. *J. Chem. Phys* **2000**, *113*, 508.

(192) Kobayashi, M.; Nakai, H. *J. Chem. Phys* **2009**, *131*, 114108.

(193) Rolik, Z.; Kallay, M. *J Chem. Phys* **2011**, *135*, 104111.

(194) Friedrich, J.; Dolg, M. *J. Chem. Theory Comput.* **2009**, *5*, 287–294.

(195) Friedrich, J.; Walczak, K. *J Chem Theory Comput* **2012**, *9*, 408–417.

(196) Eriksen, J. J.; Baudin, P.; Ettenhuber, P.; Kristensen, K.; Kjargaard, T.; Jorgensen, P. *J. Chem. Theory Comput.* **2015**, *11*, 2984–2993.

(197) Hughes, T. F.; Flocke, N.; Bartlett, R. J. *J. Phys. Chem. A* **2008**, *112*, 5994–6003.

(198) Calvin, J. A.; Lewis, C. A.; Valeev, E. F. In *IA3 '15, the 5th Workshop on Irregular Applications: Architectures and Algorithms*, ACM Press: New York, New York, USA, 2015, pp 1–8.

(199) Jung, Y. *Proc. Natl. Acad. Sci.* **2005**, *102*, 6692–6697.

(200) Watts, J. D.; Gauss, J.; Bartlett, R. J. *J. Chem. Phys.* **1993**, *98*, 8718–8733.

(201) Meyer, W. *Int. J. Quantum Chem.* **1971**, *5*, 341–348.

(202) Ahlrichs, R.; Keil, F.; Lischka, H.; Kutzelnigg, W.; Staemmler, V. *J. Chem. Phys.* **1975**, *63*, 455–463.

(203) Ahlrichs, R.; Lischka, H.; Staemmler, V.; Kutzelnigg, W. *J. Chem. Phys.* **1975**, *62*, 1225–1234.

(204) Hattig, C. *Phys. Chem. Chem. Phys.* **2005**, *7*, 59–66.

(205) Smith, D. G. A. et al. *J. Chem. Phys.* **2020**, *152*, 184108.

(206) Van Lenthe, J. H.; Duijneveldt-Van Devan Rijdt, J. G. C. M.; Van Duijneveldt, F. B. In *AB Initio Methods in Quantum Chemistry - II*; John Wiley 'I&' Sons, Ltd: 2007; Vol. 69, pp 521–566.

(207) Neese, F.; Wennmohs, F.; Hansen, A. *J. Chem. Phys.* **2009**, *130*, 114108.

(208) Nagy, P. R.; Kallay, M. *J. Chem. Theory Comput.* **2019**, *15*, 5275–5298.

(209) Murray, C. W.; Handy, N. C.; Laming, G. J. *Mol. Phys.* **1993**, *78*, 997–1014.

(210) Takatsuka, A.; Ten-no, S.; Hackbusch, W. *J. Chem. Phys* **2008**, *129*, 044112.