

Lawrence Berkeley National Laboratory

Recent Work

Title

Systematically Improvable Tensor Hypercontraction: Interpolative Separable Density-Fitting for Molecules Applied to Exact Exchange, Second- and Third-Order Møller-Plesset Perturbation Theory.

Permalink

<https://escholarship.org/uc/item/0bk5j7c3>

Journal

Journal of chemical theory and computation, 16(1)

ISSN

1549-9618

Authors

Lee, Joonho
Lin, Lin
Head-Gordon, Martin

Publication Date

2020

DOI

10.1021/acs.jctc.9b00820

Peer reviewed

**Systematically Improvable Tensor
Hypercontraction: Interpolative Separable
Density-Fitting for Molecules Applied to Exact
Exchange, Second- and Third-Order
Møller-Plesset Perturbation Theory**

Joonho Lee,^{*,†} Lin Lin,[‡] and Martin Head-Gordon^{*,†}

[†] *Department of Chemistry, University of California, Berkeley, California 94720, USA
Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California
94720, USA*

[‡] *Department of Mathematics, University of California, Berkeley, California 94720, USA
Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley,
California 94720, USA*

E-mail: linusjoonho@gmail.com; mhg@cchem.berkeley.edu

Abstract

We present a systematically improvable tensor hypercontraction (THC) factorization based on interpolative separable density fitting (ISDF). We illustrate algorithmic details to achieve this within the framework of Becke’s atom-centered quadrature grid. A single ISDF parameter c_{ISDF} controls the tradeoff between accuracy and cost. In particular, c_{ISDF} sets the number of interpolation points used in THC, $N_{\text{IP}} = c_{\text{ISDF}} \times N_{\text{X}}$ with N_{X} being the number of auxiliary basis functions. In conjunction with the resolution-of-the-identity (RI) technique, we develop and investigate the THC-RI algorithms for cubic-scaling exact exchange for Hartree-Fock and range-separated hybrids (e.g., $\omega\text{B97X-V}$) and quartic-scaling second- and third-order Møller-Plesset theory (MP2 and MP3). These algorithms were evaluated over the W4-11 thermochemistry (atomization energy) set and A24 non-covalent interaction benchmark set with standard Dunning basis sets (cc-pVDZ, cc-pVTZ, aug-cc-pVDZ, and aug-cc-pVTZ). We demonstrate the convergence of THC-RI algorithms to numerically exact RI results using ISDF points. Based on these, we make recommendations on c_{ISDF} for each basis set and method. We also demonstrate the utility of THC-RI exact exchange and MP2 for larger systems such as water clusters and C_{20} . We stress that more challenges await in obtaining accurate and numerically stable THC factorization for wavefunction amplitudes as well as the space spanned by virtual orbitals in large basis sets and implementing sparsity-aware THC-RI algorithms.

1 Introduction

The introduction of a single particle basis of atomic orbitals (AOs), $\{\omega_\mu(\mathbf{r})\}$, and the presence of $\hat{V}_{ee} = \sum_{i<j} r_{ij}^{-1}$ in the electronic Hamiltonian leads to the 4-leg two-electron repulsion integral (ERI) tensor,

$$(\mu\nu|\lambda\sigma) = \int d\mathbf{r}_1 \int d\mathbf{r}_2 \frac{\omega_\mu(\mathbf{r}_1)\omega_\nu(\mathbf{r}_1)\omega_\lambda(\mathbf{r}_2)\omega_\sigma(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}. \quad (1)$$

as a central quantity in computational quantum chemistry¹⁻⁷. Spatial locality of the N_{AO} AOs means that only a quadratic number of the $\mathcal{O}(N_{\text{AO}}^4)$ ERIs are numerically significant, which has enabled efficient AO-driven methods for Hartree-Fock (HF) and density functional theory^{8,9} as well as some post-HF methods^{10,11}. Further screening and/or collectivization of long-range quantities permits development of linear-scaling methods as well¹²⁻¹⁶. However such methods involve large prefactors because of the cost of handling the significant 4-leg ERIs.

Therefore there has been much interest in methods that entirely avoid 4-leg ERI evaluation, and instead use factorization in terms of lower rank quantities. A graphical representation of the approach reviewed below is provided in Fig. 1. The most widely used integral factorization is undoubtedly the resolution-of-the-identity (RI) (also called density-fitting (DF)) technique¹⁷⁻¹⁹. The key idea in the RI factorization is that one defines an auxiliary basis set $\{\omega_P\}$ which approximately represents electron density of system. This approximation can be made “optimal” via least-squares fit within some metric (most commonly Coulombic)²⁰. Such fitting factorizes the 4-leg tensor in Eq. (1) into 3-leg and 2-leg tensors,

$$(\mu\nu|\lambda\sigma) \approx \sum_{PQ} C_{\mu\nu}^P (P|Q) C_{\lambda\sigma}^Q \quad (2)$$

$$C_{\mu\nu}^P = \sum_Q (\mu\nu|Q) [(Q|P)]^{-1} \quad (3)$$

and thereby often greatly simplifies subsequent manipulations of the ERI tensor. The size of the auxiliary basis set, N_X , used in RI to fit the electron density is roughly 3-4 times larger than the original basis set to achieve the accuracy of 50-60 μE_h per atom^{21,22}. We note that the theoretical understanding of RI was first provided by Dunlap²³⁻²⁶. The RI technique has been combined with some local fitting strategies²⁷⁻³² to achieve linear scaling. From a similar perspective, Beebe and Linderberg applied Cholesky decomposition (CD) to the two-electron integrals which in turn achieves more or less the same speedup and compactness as those of the RI factorization³³. It is now routine to apply the RI approximation and related techniques to HF theory^{34,35}, Møller-Plesset (MP) perturbation theory^{36,37}, coupled-cluster (CC) theory^{38,39} etc.

From a somewhat different perspective, Friesner applied a pseudospectral (PS) method to approximate the ERI tensor⁴⁰⁻⁴². In PS, we define a secondary basis set as points in 3-space. The key idea is to perform one of the two integrals in Eq. (1) analytically and the rest is evaluated numerically, which avoids the Coulomb singularity and provides a numerically stable algorithm. Namely,

$$(\mu\nu|\lambda\sigma) \approx \sum_g \lambda_g \omega_\mu(\mathbf{r}_g) \omega_\nu(\mathbf{r}_g) V_{\lambda\sigma}(\mathbf{r}_g) \quad (4)$$

$$V_{\lambda\sigma}(\mathbf{r}_g) = \int d\mathbf{r} \frac{\omega_\lambda(\mathbf{r}) \omega_\sigma(\mathbf{r})}{|\mathbf{r} - \mathbf{r}_g|} \quad (5)$$

where $\{\mathbf{r}_g\}$ is a set of 3-space points and $\{\lambda_g\}$ is the weight associated with each grid point. This simple idea was successfully applied to HF achieving an overall cubic scaling algorithm, which is lower-scaling than the conventional RI-driven HF theory (quartic scaling)⁴³. This was further extended to full configuration interaction⁴⁴, configuration interaction with doubles⁴⁵, and the second- and the third-order MP (MP2 and MP3)⁴⁶ by Martínez and Carter. The major drawback of those PS methods is the lack of generality. That is, to achieve both efficiency and accuracy quoted in refs. 43,47,48 computational algorithms need to be optimized for specific basis sets. It is noteworthy that Neese and co-workers' chain-

of-sphere approximation for exact exchange (COSX) is particularly promising in terms of general applicability and simple implementation⁴⁹.

A more flexible integral factorization was proposed by Martínez and co-workers, termed tensor hypercontraction (THC)^{50–61}. It is more flexible because the 4-leg tensor in Eq. (1) is factorized into multiplications among five matrices:

$$(\mu\nu|\lambda\sigma) \approx \sum_{KL} \omega_\mu(\mathbf{r}_K)\omega_\nu(\mathbf{r}_K)M_{KL}\omega_\lambda(\mathbf{r}_L)\omega_\sigma(\mathbf{r}_L) \quad (6)$$

where $\{\mathbf{r}_K\}$ are chosen grid points and M_{KL} will be discussed in detail later. This factorization therefore requires only *quadratic* storage which is lower than the cubic storage generally needed in RI. Initial attempts to obtain such flexible and in principle systematically improvable factorization based on CANDECOMP/PARAFAC (PF) were found to be too inefficient to be competitive with conventional RI algorithms⁵⁰. Nevertheless, there have been several attempts to directly obtain the PF-THC factorization. Benedikt and co-workers studied CC with doubles with THC factorization obtained via $\mathcal{O}(N^5)$ alternating least-squares algorithm⁶². Hummel and co-workers obtained a PF-THC factorization based on $\mathcal{O}(N^4)$ alternating least-squares and the resulting factorization was applied to planewave CC with singles and doubles (CCSD)⁶³. A comparison of different strategies ($\mathcal{O}(N^{4-5})$ algorithms) for obtaining PF-THC factorization in the context of gaussian orbital based CCSD was done by Schutski and co-workers⁶⁴.

A different strategy using least-squares fitting, namely least-squares THC (LS-THC), was found to be more efficient, in principle systematically improvable, and yet still accurate⁵¹. Unlike the aforementioned approach, obtaining the THC factorization in LS-THC scales cubically with system size when sparsity is considered. Some intermediates in LS-THC have a strong resemblance with PS approaches but overall it provides more convenient factorization for subsequent calculations. Combined with the RI technique, the resulting LS-THC-RI has been applied to MP2⁵⁰, MP3⁵⁰, CCSD⁵², second-order approximate CC with singles

and doubles for ground state⁵⁴ and excited states⁵⁵, second-order complete active-space perturbation theory (CASPT2)⁶⁵, particle-particle random phase approximation (ppRPA)⁶⁶, and direct random phase approximation (dRPA)⁶⁷ for finite-sized systems (i.e., molecules). Although LS-THC-RI is powerful on its own, to realize its full potential some optimization of quadrature points and weights was needed for specific basis sets⁵⁸.

In our opinion, the major reason behind somewhat uncommon usage of LS-THC-RI by research groups other than its developers is its lack of generality in generating a compact set of grid points that works for all chemical elements and any basis set. Recently, Lu and Ying recognized that LS-THC is a form of a low-rank approximation to the density represented in 3-space⁶⁸. This unique perspective allowed for a systematic way to pick a set of 3-space points used in THC based on column-pivoted QR (QRCP) decomposition. This was termed interpolative separable density fitting (ISDF). Those points are not quadrature points unlike in the original formulation of THC. Rather, they are *interpolation* points which therefore are not associated with any quadrature weights. Subsequently, the centroidal Voronoi tessellation (CVT) technique was developed to remove the expensive QRCP step and reduce the overall cost to $\mathcal{O}(N^2)$ for obtaining interpolation points⁶⁹. The ISDF approach was then applied to HF exact exchange (HF-K)⁷⁰, dRPA⁷¹, the Bethe-Salpeter equation⁷², phonon calculations in density functional perturbation theory⁷³, and auxiliary-field quantum Monte Carlo⁷⁴ for systems under periodic boundary conditions (PBCs).

Given the utility of the ISDF point selection shown for systems under PBCs, it is valuable to apply this technique to molecular simulations where the original LS-THC techniques currently lack full generality in choosing a compact set of 3-space points. There are some technical challenges: (1) a fully numerical evaluation of Coulomb integrals is nearly impossible for finite-sized systems due to the Coulomb singularity in contrast to those for infinite systems where one can use fast Fourier transform (FFT) and (2) we use atom-centered quadrature grids⁷⁵ unlike uniform grids used for PBC simulations. We will show that (1) has already been solved in LS-THC-RI and for (2) we will present a modified $\mathcal{O}(N)$ CVT algorithm

that is well-suited for atom-centered grids. Furthermore, we will show that performing least-squares fits for each molecular orbital (MO) block in the MO transformed integrals is much more accurate than fitting the atomic orbital (AO) integrals, which was previously noted in refs. 51,57 by Martínez’s group. In passing, we note that our CVT procedure has the same goal as the radial discrete value representation (R-DVR) developed by Martínez and co-workers⁵³. R-DVR achieves generality just like CVT as it, in principle, can work with all chemical elements and basis sets without empirical fitting. However, R-DVR was shown to be less compact than hand-optimized grids for water clusters for similar accuracy⁵⁸. On the other hand, the grid points obtained from CVT were found to be as compact and accurate as the hand-optimized ones as discussed later in this work. Therefore, the CVT procedure examined here achieves both compactness and generality.

In addition to the atom-centered CVT procedure, we also test the accuracy of the factorization of the MP1 T -amplitudes when evaluating MP2 and MP3 correlation energies. The factorization of T -amplitude has already been used in THC-CCSD^{52,62,64,66}. Such factorization can speed up THC-RI-MP3 by a factor of 49 compared to the original algorithm⁵⁰ by removing the double Laplace transform associated with the algorithm. Our goal is to examine the accuracy of this factorization when using CVT grid points and the AO basis functions for least-squares fits over a wide range of small molecules. This paper is organized as follows: we will (1) review the essence of THC factorization and ISDF point selection, (2) present a modified CVT algorithm for atom-centered grids, (3) review THC algorithms for exact exchange, MP2, and MP3, (4) present careful benchmarks of the accuracy of these algorithms over the W4-11⁷⁶ and A24⁷⁷ sets (widely used thermochemistry and non-covalent interaction benchmarks, respectively) as well as water clusters and C₂₀.

2 Theory

2.1 Notation

In Table 1, we summarize the notation used in this work. We assume the spin-orbital basis

Table 1: Notation used in this work.

Notation	Description
$\mu, \nu, \lambda, \sigma$, etc.	indices for atomic orbitals (AOs)
i, j, k, l , etc.	indices for occupied molecular orbitals (MOs)
a, b, c, d , etc.	indices for virtual molecular orbitals (MOs)
P, Q, R, S	indices for resolution-of-the-identity (RI) basis functions
K, L, M, N	indices for interpolation vectors
N_{AO}	the number of AOs
N_{X}	the number of RI basis functions
N_{SP}	the number of <i>significant</i> shell-pairs
N_{IP}	the number of interpolation vectors
n_{occ}	the number of occupied orbitals
n_{vir}	the number of virtual orbitals
n_t	the number of Laplace quadrature points
$\omega_{\mu}(\mathbf{r})$	μ -th AO basis function at \mathbf{r}
$\omega_P(\mathbf{r})$	P -th auxiliary (RI) basis function at \mathbf{r}
$\phi_p(\mathbf{r})$	p -th molecular orbital (MO) at \mathbf{r}
$\xi_K(\mathbf{r})$	K -th interpolation vector at \mathbf{r}

unless specified otherwise. We define molecular orbitals (MOs) with MO coefficient matrix, \mathbf{c} ,

$$\phi_p(\mathbf{r}) = \sum_{\mu} c_{\mu p} \omega_{\mu}(\mathbf{r}) \quad (7)$$

For later use, we define *imaginary time-dependent* MOs:

$$\phi_i(t) = \phi_i e^{-t\epsilon_i} \quad (8)$$

$$\phi_a(t) = \phi_a e^{t\epsilon_a} \quad (9)$$

where ϵ_p denotes the orbital energy of the p -th MO. These time-dependent MOs naturally arise when applying the Laplace transformation to orbital energy denominators as shown

later.

2.2 Review of Tensor Hypercontraction

We review tensor hypercontraction (THC) developed by Martínez and co-workers^{50–61}. In resolution-of-the-identity (RI) or density-fitting (DF), product of AOs are represented in an auxiliary basis by the following decomposition of a 3-leg density tensor $Z_{\mu\nu}(\mathbf{r})$:

$$Z_{\mu\nu}(\mathbf{r}) = \omega_{\mu}(\mathbf{r})\omega_{\nu}(\mathbf{r}) \simeq \sum_P C_{\mu\nu}^P \omega_P(\mathbf{r}) \quad (10)$$

where $\omega_{\mu}(\mathbf{r})$ is the AO basis function on a grid point \mathbf{r} , $C_{\mu\nu}^P$ is the fit coefficient, and $\omega_P(\mathbf{r})$ is the auxiliary function. The use of the term RI is ironic in that the product-separable structure in $Z_{\mu\nu}(\mathbf{r})$ disappears with it. Namely, RI introduces an inseparable 3-leg tensor as the 3-leg fit coefficient $C_{\mu\nu}^P$. The upshot of THC factorization is that the 3-leg fit coefficient in RI is avoided by use of an approximate separable form involving a grid of points:

$$Z_{\mu\nu}(\mathbf{r}) \simeq \sum_K C_{\mu\nu}^K \xi_K(\mathbf{r}) = \sum_K \omega_{\mu}^K \omega_{\nu}^K \xi_K(\mathbf{r}) \quad (11)$$

where we write

$$\omega_{\mu}^K \equiv \omega_{\mu}(\mathbf{r}_K) \quad (12)$$

and $\xi_K(\mathbf{r})$ is a fit function (or interpolation vector) that will be specified later.

The key differences between RI and THC in representing the 4-leg integral tensor, $(\mu\nu|\lambda\sigma)$, can be also understood graphically as described in Figure 1. The RI representation evidently removes the separable structure in \mathbf{Z} whereas the THC factorization recovers such a separable form. This separable structure is the key in reducing the cost for subsequent calculations. Achieving the THC factorization takes two crucial steps: (1) choosing a set of points $\{\mathbf{r}_K\}$ in 3-space and (2) determining the fit functions $\{\xi_K(\mathbf{r})\}$. Their number, N_{IP} , must be as small as possible for efficiency, but as large as necessary to achieve usable

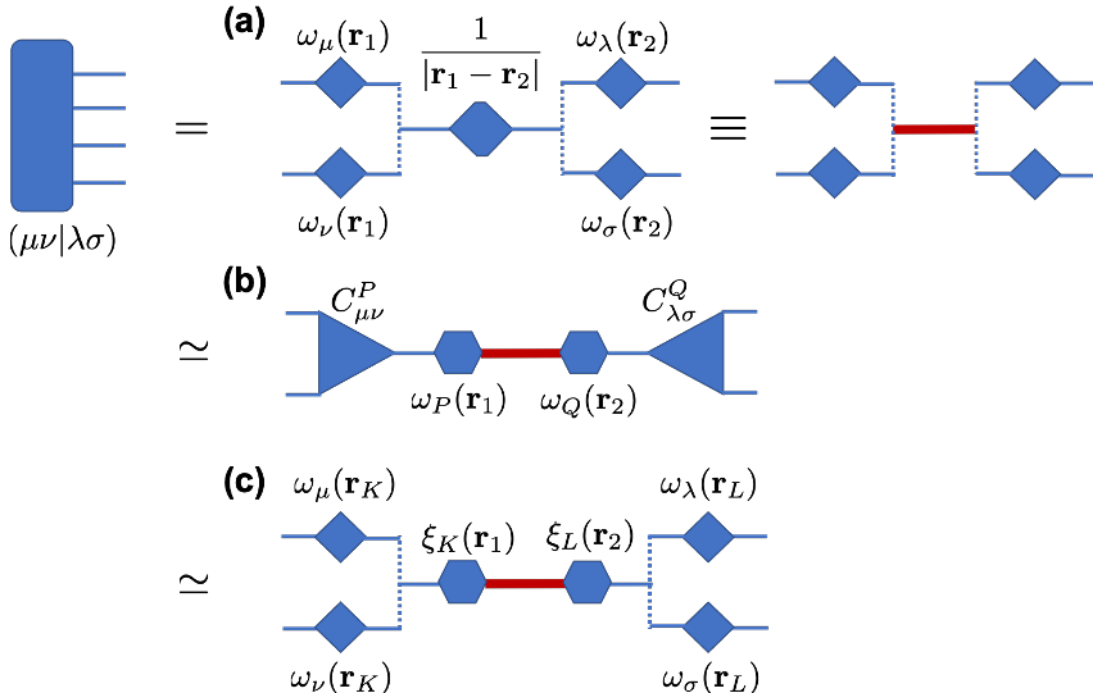


Figure 1: Graphical representation of density tensor (a) represented exactly, (b) represented using RI, and (c) represented using THC. The solid blue line is a contraction over shared indices between two tensors while the dotted blue line indicates element-wise product over connected indices (i.e., Hadamard product). The thick red bar represents a contraction (in real-space) with the Coulomb operator, $\frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|}$.

accuracy. For (1), we will follow the strategy introduced by Lu and Ying⁶⁸. Namely, we will use the interpolative separable density fitting (ISDF) technique to determine a set of *interpolation points*. Then, for (2), we will follow the least-squares THC (LS-THC) technique by Martínez and co-workers⁵¹. For the purpose of this paper, we will refer to LS-THC simply as THC.

2.3 Selecting Interpolation Points

When choosing interpolation points $\{\mathbf{r}_K\}$, Martínez and co-workers have tried to prune the Becke quadrature grid widely used in density functional theory⁷⁵. Some of their efforts have focused on further optimizing the grid representation^{53,58} and special optimization procedures were performed for each basis set⁵⁸. A simpler and automatic interpolation point generation

is possible as suggested in ISDF. We will describe how one can adapt the ISDF procedure developed in refs. 68,69 to choose a subset of Becke quadrature points. This procedure is almost automatic and also systematically improvable with a single tunable parameter.

The important observation made in ref. 68 is that Eq. (11) is in fact a form of a low-rank approximation when evaluated on a finite grid of points, N_g . If one looks at \mathbf{Z} as a N_g -by- N_{AO}^2 matrix, a singular value decomposition (SVD) of \mathbf{Z} would have at most $\min(N_g, N_{\text{AO}}^2)$ number of non-zero singular values. In most practical applications, N_{AO}^2 is greater than N_g . The goal of ISDF is then to pick a subset of N_g rows of \mathbf{Z} that closely approximates the entire matrix \mathbf{Z} . Such a low-rank approximation is called interpolative decomposition (ID) where the goal is to compress a given tensor while preserving the column or row structure. A common way to accomplish ID is to use the QR decomposition with column-pivoting (QRCP) as was done in ref. 68. Namely,

$$\mathbf{Z}^T \mathbf{\Pi} = \mathbf{Q} \mathbf{R} \tag{13}$$

where $\mathbf{\Pi}$ is the permutation matrix from the QRCP procedure, \mathbf{Q} is an orthogonal matrix, and \mathbf{R} is a upper triangular matrix. Each column of $\mathbf{\Pi}$ contains a single 1 and $(N_g - 1)$ 0's. It permutes the columns of \mathbf{Z}^T such that the diagonal entires of \mathbf{R} are non-increasing. In other words, the $\mathbf{\Pi}$ matrix orders the columns of \mathbf{Z}^T based on their relative importance. Therefore, this automatically produces an ‘‘optimal’’ subset of N_g grid points that are used as interpolation points in ISDF. The cost of this QRCP is quartic if performed exactly and can be reduced to cubic if randomized algorithms are employed^{68,70}.

Rather than employing randomized QRCP, a further cost reduction in choosing interpolation points is possible using the centroidal Voronoi tessellation (CVT) procedure proposed by Lin and co-workers⁶⁹. We will describe a modified version of the CVT procedure so that it is well-suited for Becke’s quadrature grid⁷⁵. Unlike uniform grids, the Becke grid points carry quadrature weights which also encode some geometric information of systems through

interatomic quadrature weights, $\lambda(\mathbf{r})$. In this sense, a better quantity to compress is

$$\lambda(\mathbf{r})Z_{\mu\nu}(\mathbf{r}) \simeq \sum_K C_{\mu\nu}^K \lambda(\mathbf{r}_K) \xi_K(\mathbf{r}). \quad (14)$$

The quadrature weights are necessary to account for non-uniformity of the Becke grid and it was indeed found to be much more accurate than carrying out interpolation without the weights. We note that one can trivially compute an overlap matrix element on the Becke grid:

$$S_{\mu\nu} = \sum_{\mathbf{r}} \sum_K C_{\mu\nu}^K \lambda(\mathbf{r}_K) \xi_K(\mathbf{r}) \quad (15)$$

which is a useful quantity for debugging purposes.

In Algorithm 1, we outline the CVT procedure adopted for the Becke grid. In essence, we start from an initial set of centroids, $\mathbf{c}^{(0)}$, and update these points based on the neighbors of these points as well as the weights associated with them. The resulting centroids are our interpolation points. Interested readers are referred to ref. 69. It is worthwhile to mention two modifications we made to generate the interpolation points as a subset of the Becke grid.

First, we loop over individual atoms to generate atom-specific interpolation points. This is necessary to maximally exploit the locality of Gaussian AO basis sets and atom-centered grids. Furthermore, in Algorithm 1 the distance between centroids and grid points needs to be computed every iteration and this becomes quite computationally demanding. If one were to do this for all atoms at once, this step scales quadratically with system size and carries a non-negligible prefactor. Looping over individual atoms naturally yields an algorithm that scales linearly with system size.

Second, we use the Becke quadrature weights in the weighted CVT procedure. The Becke quadrature weights reflect geometric information around a given grid point through the so-called Becke partition function⁷⁵. Namely, the weight of a grid point depends on the location of all atoms and associated grid points. Lin and co-workers suggested that electron density $\rho(\mathbf{r})$ can be used for the weight function in CVT. The use of $\rho(\mathbf{r})$ in the CVT procedure will

complicate derivative theories since the position of those 3-space points will depend on the electronic density. As such, we considered other alternatives and we found that the Becke quadrature weights are simple to use and yield interpolation points that result in accurate compression for later uses. This will not complicate derivative theories either. This weighted CVT procedure is used throughout this paper.

Algorithm 1: Weighted K-Means Algorithm For the Becke Grid. n_g and n_μ denote the number of quadrature points and the number of interpolation points (user-specified) for a given atom, respectively.

```

1 for  $A = 1$  to  $N$ ;                                     // Loop over atoms.
2 do
3   Generate  $n_g$  grid points,  $\{\mathbf{r}_A\}$ , and their weights,  $\lambda(\mathbf{r})$  for  $\mathbf{r} \in \{\mathbf{r}_A\}$ ;
4   Compute  $\omega(\mathbf{r})$  for  $\mathbf{r} \in \{\mathbf{r}_A\}$ ;
5   Select  $n_\mu$  points,  $\mathbf{c}^{(0)}$ , randomly from  $\{\mathbf{r}_A\}$ ;
6   Call the K-means subroutine described in the Algorithm I in ref. 69 with weight
   function  $\lambda(\mathbf{r})$ ;
7 end

```

2.4 Determination of Interpolation Vectors

Once we have a set of interpolation points $\{\mathbf{r}_K\}$, the hitherto undefined interpolation vectors required in Eq. (11) can be obtained via the following least-squares expression,

$$\xi_K(\mathbf{r}) = \sum_L \tilde{S}_{KL}^+ C_{\mu\nu}^L \lambda(\mathbf{r}_L) Z_{\mu\nu}(\mathbf{r}) \lambda(\mathbf{r}) \quad (16)$$

where \mathbf{Z} and \mathbf{C} are defined in Eq. (10) and Eq. (11), respectively, and the metric in 3-space is defined as

$$\tilde{S}_{KL} = \sum_{\mu\nu} \omega_\mu^K \omega_\nu^K \lambda(\mathbf{r}_K) \omega_\mu^L \omega_\nu^L \lambda(\mathbf{r}_L) \quad (17)$$

and $\tilde{\mathbf{S}}^+$ is the pseudoinverse of $\tilde{\mathbf{S}}$. $\tilde{\mathbf{S}}$ is generally ill-conditioned⁶¹ and the threshold of 10^{-12} in the drop tolerance for evaluating $\tilde{\mathbf{S}}^+$ was found to be sufficient for the purpose of our paper.

As shown by multiple authors^{51,67,70}, due to the separable form of \mathbf{C} in Eq. (11) the formation of ξ in Eq. (16) can be carried out with $\mathcal{O}(N^3)$ cost. First, the formation of $\tilde{\mathbf{S}}$ scales cubically, despite its formal $\mathcal{O}(N^4)$ scaling,

$$\tilde{S}_{KL} = \lambda(\mathbf{r}_K)\lambda(\mathbf{r}_L) \sum_{\mu} (\omega_{\mu}^K \omega_{\mu}^L) \sum_{\nu} (\omega_{\nu}^K \omega_{\nu}^L) \quad (18)$$

Similarly, the formation of $\mathbf{C}^\dagger \mathbf{Z}$ needed for Eq. (16) can be done as

$$(\mathbf{C}^\dagger \mathbf{Z})_{K,\mathbf{r}} = \sum_{\mu} (\omega_{\mu}^K \omega_{\mu}(\mathbf{r})) \sum_{\nu} (\omega_{\nu}^K \omega_{\nu}(\mathbf{r})), \quad (19)$$

which also scales cubically with system size.

With this decomposition, the two-electron four-center integral tensor can be factorized into Eq. (6) where we define \mathbf{M} to be the Coulomb integral between interpolation vectors,

$$M_{KL} = \int d\mathbf{r}_1 \int d\mathbf{r}_2 \frac{\xi_K(\mathbf{r}_1)\xi_L(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} \quad (20)$$

As pointed by Lu and Ying⁶⁸, the evaluation of each Coulomb integral needed to form \mathbf{M} for translationally invariant systems (e.g., solids) can be done by fast Fourier transform (FFT) with the $\mathcal{O}(N \log N)$. In such cases, the formation of \mathbf{M} scales only cubically with the system size. Therefore, the overall THC factorization scales cubically with the system size as long as the evaluation of Eq. (20) requires only cubic (or less than cubic) work. In passing, we note that the matrix \mathbf{M} is denoted as \mathbf{Z} in most THC literature.

However, as suggested in ref. 51, for molecular systems it is difficult to directly evaluate Eq. (20) using quadrature with a finite number of grid points. The challenge is the Coulomb singularity present at $\mathbf{r}_1 = \mathbf{r}_2$ which is an integrable singularity. Avoiding this singularity

may be possible at the expense of employing very large grids. Usually, the number of grid points required to achieve quantitative accuracy is too large to be practically useful. Instead, one could use FFT to represent ξ with planewaves and perform the Coulomb integrals in the reciprocal space assuming the molecular system of interest is in a large box under periodic boundary conditions. This has been tried in ref 69. However, this would naturally require a pseudopotential treatment of core electrons because it is difficult to represent highly localized core orbitals with planewaves. We must seek alternatives to perform all-electron calculations.

2.4.1 Atomic Orbital THC (AO-THC)

In this work, we will take an alternative route to evaluate M_{KL} which turns out to be identical to “atomic orbital” THC (AO-THC) introduced in ref. 51. Consider

$$|\xi_K\rangle = \sum_{\mu\nu L} |\mu\nu\rangle C_{\mu\nu}^L \lambda(\mathbf{r}_L) \left(\tilde{\mathbf{S}}^\dagger\right)_{LK} \quad (21)$$

It is clear that the projection of this onto $(\mathbf{r}|$ simply recovers Eq. (16) up to the weight $\lambda(\mathbf{r})$ which can be neglected if one performs an integration over \mathbf{r} analytically. Since we would like to compute M_{KL} based on analytic Coulomb integrals, we employ an RI auxiliary basis set, $|P\rangle$, and express Coulomb integrals in terms of AO and RI integrals. The quantity of interest is

$$D_{QK} = \sum_P (Q|P)^{-1/2} (P|\xi_K) = \sum_{\mu\nu PL} (Q|P)^{-1/2} (P|\mu\nu) C_{\mu\nu}^L \lambda(\mathbf{r}_L) \left(\tilde{\mathbf{S}}\right)_{LK}^+ \quad (22)$$

Using this, we have

$$M_{KL} = \sum_Q D_{QK} D_{QL} \quad (23)$$

The bottleneck in the evaluation of \mathbf{M} is computing an intermediate required for \mathbf{D} ,

$$\sum_{\mu\nu} (P|\mu\nu) C_{\mu\nu}^L = \sum_{\mu\nu} (P|\mu\nu) \omega_\mu^L \omega_\nu^L \quad (24)$$

which scales as $\mathcal{O}(N_{\text{AO}}^2 N_{\text{X}} N_{\text{IP}})$. This unfortunately exceeds cubic scaling (without any sparsity consideration).

We note that this quartic scaling can be reduced to “asymptotically” cubic-scaling since the number of shell-pairs, $|\mu\nu\rangle$, scales linearly with system size for large systems. Therefore, in this case the scaling for forming \mathbf{D} is $\mathcal{O}(N_{\text{SP}} N_{\text{X}} N_{\text{IP}})$. In addition to the shell-pair screening, a production-level implementation may screen AOs on the interpolation points \mathbf{r}_L . The number of significant basis functions at a given point, \mathbf{r}_L , should be asymptotically constant as seen in the linear-scaling density functional theory implementations⁷⁸. With this screening technique, Eq. (24) may be carried out with asymptotically quadratic cost ($\mathcal{O}(c N_{\text{IP}} N_{\text{X}})$) where c denotes the number of *significant* shell-pairs for a given grid point. For the purpose of this study, the simpler cubic-scaling algorithm is sufficient and therefore this will be used throughout this paper.

For completeness, we write the matrix \mathbf{M} in terms of exact two-electron integrals (i.e., RI is not used) as well:

$$M_{KL} = \sum_{\mu\nu\lambda\sigma} \sum_{K'L'} (\mu\nu|\lambda\sigma) C_{\mu\nu}^{K'} C_{\lambda\sigma}^{L'} \lambda(\mathbf{r}_{K'}) \lambda(\mathbf{r}_{L'}) \left(\tilde{\mathbf{S}}^+\right)_{K'K} \left(\tilde{\mathbf{S}}^+\right)_{L'L}. \quad (25)$$

Without considering sparsity, Eq. (25) scales as $\mathcal{O}(N_{\text{AO}}^4 N_{\text{IP}})$ and this is not effective for large-scale applications. With sparsity, the scaling of the construction of \mathbf{M} can be shown to be cubic. Namely, the contraction over μ and ν to form

$$G_{\lambda\sigma}^K = \sum_{K'} \left(\tilde{\mathbf{S}}^+\right)_{K'K} \left(\sum_{\mu\nu} (\mu\nu|\lambda\sigma) C_{\mu\nu}^{K'} \lambda(\mathbf{r}_{K'})\right) \quad (26)$$

scales as $\mathcal{O}(N_{\text{SP}}^2 N_{\text{IP}})$. Similarly, the second step where we contract over λ and σ ,

$$M_{KL} = \sum_{L'} \left(\tilde{\mathbf{S}}^+\right)_{L'L} \left(\sum_{\lambda\sigma} G_{\lambda\sigma}^K C_{\lambda\sigma}^{L'}\right), \quad (27)$$

scales as $\mathcal{O}(N_{\text{SP}}N_{\text{IP}}^2)$. With THC, the generation of the tensor \mathbf{M} scales worse than the generation of the ERI tensor $(\mu\nu|\lambda\sigma)$, which asymptotically scales as $\mathcal{O}(N_{\text{SP}}^2)$, and the hope is that the prefactor may be smaller with THC for some system size.

We summarize the discussion of computational cost to form \mathbf{M} in Table 2.

Table 2: Computational cost for forming \mathbf{M} (Eq. (23), Eq. (25)). Just with shell-pair sparsity, one can achieve a cubic-scaling algorithm to form \mathbf{M} .

Eq.	RI	Shell-pair sparsity	Computational cost
(25)	No	No	$\mathcal{O}(N_{\text{AO}}^4 N_{\text{IP}})$
(23)	Yes	No	$\mathcal{O}(N_{\text{AO}}^2 N_{\text{X}} N_{\text{IP}})$
(25)	No	Yes	$\mathcal{O}(N_{\text{SP}}^2 N_{\text{IP}}) + \mathcal{O}(N_{\text{SP}} N_{\text{IP}}^2)$
(23)	Yes	Yes	$\mathcal{O}(N_{\text{SP}} N_{\text{X}} N_{\text{IP}}) + \mathcal{O}(N_{\text{X}}^3) + \mathcal{O}(N_{\text{X}} N_{\text{IP}}^2) + \mathcal{O}(N_{\text{X}}^2 N_{\text{IP}}) + \mathcal{O}(N_{\text{IP}}^3)$

2.4.2 Molecular Orbital THC (MO-THC)

In some applications of THC, one may not need the entire set of MO integrals. For instance, in the case of RI-MP2, the only required MO block is the occupied-virtual (OV) block. It is then possible to perform a least-squares fit on the OV block, which is easier to represent than the entire MO integrals. We denote this MO-block-specific THC factorization as MO-THC⁶¹. For later use, we define an MO-block-specific quantity $\mathbf{D}^{[xy]}$ (x and y denote MO blocks which will be o (occupied), v (virtual) or n (AO)),

$$D_{QK}^{[xy]} = \sum_P (Q|P)^{-1/2} (P|\xi_K^{[xy]}) \quad (28)$$

where

$$|\xi_K^{[xy]}\rangle = \sum_{p_x q_y L} |p_x q_y\rangle C_{p_x q_y}^L \lambda(\mathbf{r}_L) \left(\tilde{\mathbf{S}}^{[xy]}\right)_{LK}^+ \quad (29)$$

$$= \sum_{p_x q_y L} \sum_{\mu\nu\lambda\sigma} |\mu\nu\rangle c_{\mu p_x} c_{\nu q_y} c_{\lambda p_x} c_{\sigma q_y} C_{\lambda\sigma}^L \lambda(\mathbf{r}_L) \left(\tilde{\mathbf{S}}^{[xy]}\right)_{LK}^+ \quad (30)$$

$$= \sum_L \sum_{\mu\nu\lambda\sigma} |\mu\nu\rangle P_{\mu\lambda}^{[x]} P_{\nu\sigma}^{[y]} \omega_\lambda^L \omega_\sigma^L \lambda(\mathbf{r}_L) \left(\tilde{\mathbf{S}}^{[xy]}\right)_{LK}^+ \quad (31)$$

$$= \sum_L \left(\sum_{\mu\nu} |\mu\nu\rangle \tilde{\omega}_\mu^L[x] \tilde{\omega}_\nu^L[y] \right) \lambda(\mathbf{r}_L) \left(\tilde{\mathbf{S}}^{[xy]}\right)_{LK}^+, \quad (32)$$

with $C_{\mu p_x}$ being the MO coefficient for the MO block x and $\left(\tilde{\mathbf{S}}^{[xy]}\right)^+$ is the pseudoinverse of the metric for the pertinent MO block $[xy]$. The bottleneck in forming $\mathbf{D}^{[xy]}$ scales as $\mathcal{O}(N_{\text{SP}} N_{\text{X}} N_{\text{IP}})$ (i.e., the same as forming \mathbf{D} as expected). With this, \mathbf{M} becomes also MO-block-dependent,

$$M_{KL}^{[wx][yz]} = \sum_Q D_{QK}^{[wx]} D_{QL}^{[yz]} \quad (33)$$

and if $[wx] = [yz]$

$$M_{KL}^{[wx]} = \sum_Q D_{QK}^{[wx]} D_{QL}^{[wx]} \quad (34)$$

We will specify MO-blocks whenever appropriate.

3 Hartree-Fock Theory

The exact exchange matrix reads

$$K_{\mu\nu} = -\frac{1}{2} \sum_{\lambda\sigma} (\mu\lambda|\sigma\nu) P_{\sigma\lambda} \quad (35)$$

where the one-particle density matrix, \mathbf{P} , is

$$P_{\sigma\lambda} = \sum_i c_{\sigma i} c_{\lambda i} \quad (36)$$

The complexity of evaluating $K_{\mu\nu}$ scales as $\mathcal{O}(N_{\text{AO}}^4)$ without any screening. With proper integral screening, this costs only a quadratic ($\mathcal{O}(N_{\text{SP}}^2)$) amount of work. We will refer to this as the AO-K algorithm. The RI implementation of the K -build asymptotically costs more:

$$K_{\mu\nu}^{\text{RI}} = -\frac{1}{2} \sum_{iPQ} (\mu i|P)(P|Q)^{-1}(Q|i\nu) \quad (37)$$

where the final contraction to form \mathbf{K} scales as $\mathcal{O}(N_{\text{AO}}^2 n_{\text{occ}} N_{\text{X}})$ (i.e., quartic scaling).

In terms of \mathbf{M} , the evaluation of THC-RI-K is straightforward:

$$K_{\mu\nu}^{\text{THC}} = -\frac{1}{2} \sum_{iKL} \omega_{\mu}^K \phi_i^K M_{KL} \phi_i^L \omega_{\nu}^L \quad (38)$$

where ϕ_i is defined in Eq. (7). Since every operation in Eq. (38) is a matrix-matrix multiplication (with a carefully chosen contraction ordering), the THC-RI-K algorithm scales cubically with system size (i.e., $\mathcal{O}(N_{\text{AO}} N_{\text{IP}}^2)$). In comparison to RI-K, ISDF improves the asymptotic scaling and the hope is that it will show this favorable scaling starting from reasonably small molecules. When evaluating, Eq. (38) we have two choices for the input \mathbf{M} : MO-THC and AO-THC. MO-THC uses $\mathbf{M}^{[\text{on}]}$ where o denotes the occupied MO block and n denotes the entire AO block. AO-THC uses $\mathbf{M}^{[nn]}$ which does not depend on MOs unlike the first one.

The first one ($\mathbf{M}^{[\text{on}]}$) turns out to closely approximate RI-K for a fixed density matrix \mathbf{P} . From a low-rank perspective, it is much easier to represent $(\mu i|i\nu)$ than $(\mu\lambda|\sigma\nu)$. This is simply due to the fact that the former has a lower rank to begin with (due to its lower dimension). As such, it is expected that the THC factorization works more efficiently with the former integral. Indeed, this was found to be the case in the evaluation of the local energy in

auxiliary-field quantum Monte Carlo⁷⁴ and solving the Bethe-Salpeter equation⁷². However, one must consider additional complications arising from the use of MO-dependent \mathbf{M} in Eq. (38). Since $\mathbf{M}^{[\text{on}]}$ is a function of \mathbf{P} , it is necessary to re-derive the Fock matrix properly as a derivative of the THC-RI-K energy expression with respect to \mathbf{P} . This subtlety was not explored in the THC-RI-K implementation paper for planewaves⁷⁰. When using orbital gradient driven self-consistent field (SCF) solvers such as geometric direct minimization⁷⁹, ignoring this subtlety leads to numerically unstable behavior and SCF calculations often do not even converge. On the other hand, in our experience, the use of other SCF solvers such as Pulay’s direct inversion of the iterative subspace (DIIS)^{80,81} can converge provided that the stopping criterion is relatively loose.

Given this, we adopted the AO-THC-RI-K approach which uses MO-independent $\mathbf{M} = \mathbf{M}^{[\text{nn}]}$. This was found to be generally less accurate than AO-THC for given number of interpolation vectors but it has a few notable strengths: (1) it is numerically stable and SCF solves can converge tightly and (2) the (relatively expensive) formation of \mathbf{M} needs to occur only once at the beginning of the SCF.

For completeness, we briefly discuss the evaluation of \mathbf{J} . The exact J-build is

$$J_{\mu\nu} = \sum_{\lambda\sigma} (\mu\nu|\lambda\sigma) P_{\sigma\lambda}, \quad (39)$$

with the asymptotic scaling of $\mathcal{O}(N_{\text{SP}}^2)$. The RI-J expression reads

$$J_{\mu\nu}^{\text{RI}} = \sum_{\mu\nu\lambda\sigma} \sum_{PQ} (\mu\nu|P)(P|Q)^{-1}(Q|\lambda\sigma) P_{\lambda\sigma}, \quad (40)$$

which asymptotically scales as $\mathcal{O}(N_{\text{SP}}N_X) + \mathcal{O}(N_X^2)$. There is also a single cubic scaling step which arises from the formation of $(P|Q)^{-1}$ (i.e., $\mathcal{O}(N_X^3)$). With THC-RI, we have

$$J_{\mu\nu}^{\text{THC-RI}} = \sum_{iKL} \omega_{\mu}^K \omega_{\nu}^K M_{KL} \phi_i^L \phi_i^L, \quad (41)$$

which asymptotically scales as $\mathcal{O}(N_{\text{SP}}N_{\text{IP}}) + \mathcal{O}(N_{\text{IP}}^2) + \mathcal{O}(N_{\text{IP}}n_{\text{occ}})$. There is also a cubic scaling step required to form \mathbf{M} . As N_{IP} is expected to be larger than N_{X} , the THC-RI-J build is asymptotically more expensive than the RI-J build. We note that just as in THC-RI-K, THC-RI-J can also be done with either AO-THC ($\mathbf{M} = \mathbf{M}^{[nn]}$) or MO-THC ($\mathbf{M} = \mathbf{M}^{[nn][oo]}$).

Using RI-J and RI-K at the same time can result in non-variational HF energies⁸² because the HF energy with both RI-J and RI-K is no longer an upper bound to the exact HF energy. The origin of this is that \mathbf{J} is positive-definite whereas \mathbf{K} is negative-definite. The RI approximation makes a non-positive error for the energy from \mathbf{J} but a non-negative error for the energy from \mathbf{K} . These two errors do not cancel out and often cause non-positive error overall. This is the origin of potential non-variationality in RI-HF (and also RI-J itself). As such, we investigated THC-RI-K with exact formation of \mathbf{J} and will present numerical data for it later in this paper.

We present a summary of computational scaling for computing \mathbf{J} and \mathbf{K} with different approximations in Table 3.

Table 3: Computational scaling for forming \mathbf{J} and \mathbf{K} in HF. M here denote a measure for system size that scales linearly with the number of atoms. Note that RI-J is cubic scaling due to the formation of $(P|Q)^{-1}$ and THC-RI-J is cubic scaling due to the formation of \mathbf{M} .

Algorithm	RI	Sparsity	THC	Scaling for \mathbf{J}	Scaling for \mathbf{K}
AO	No	No	No	$\mathcal{O}(M^4)$	$\mathcal{O}(M^4)$
AO	No	Yes	No	$\mathcal{O}(M^2)$	$\mathcal{O}(M^2)$
RI	Yes	Yes	No	$\mathcal{O}(M^3)$	$\mathcal{O}(M^4)$
THC-RI	Yes	Yes	Yes	$\mathcal{O}(M^3)$	$\mathcal{O}(M^3)$

4 Second-Order Møller-Plesset Perturbation Theory

The second-order Møller-Plesset perturbation theory (MP2) correlation energy with the RI approximation reads

$$E_{\text{MP2}}^{\text{RI}} = -\frac{1}{4} \sum_{ijab} \frac{|\langle ij||ab \rangle_{\text{RI}}|^2}{\Delta_{ij}^{ab}} \quad (42)$$

where Δ_{ij}^{ab} is the energy denominator,

$$\Delta_{ij}^{ab} = \epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j, \quad (43)$$

with ϵ_p being the MO energy of the p -th MO and the RI integrals are

$$\langle ij||ab \rangle_{\text{RI}} = \sum_P \left(B_{ia}^P B_{jb}^P - B_{ib}^P B_{ja}^P \right) \quad (44)$$

where

$$B_{ia}^P = \sum_Q C_{ia}^Q (Q|P)^{1/2} = \sum_Q (ia|Q)(Q|P)^{-1/2} \quad (45)$$

We then define J-like and K-like terms,

$$E_{\text{MP2-J}}^{\text{RI}} = -\frac{1}{2} \sum_{ijab} \frac{\sum_P (B_{ia}^P B_{jb}^P) \sum_Q (B_{ia}^Q B_{jb}^Q)}{\Delta_{ij}^{ab}} \quad (46)$$

$$E_{\text{MP2-K}}^{\text{RI}} = \frac{1}{2} \sum_{ijab} \frac{\sum_P (B_{ia}^P B_{jb}^P) \sum_Q (B_{ib}^Q B_{ja}^Q)}{\Delta_{ij}^{ab}} \quad (47)$$

and evidently $E_{\text{MP2}} = E_{\text{MP2-J}} + E_{\text{MP2-K}}$. The computational scaling of both RI-MP2-J and RI-MP2-K correlation energy evaluation is $\mathcal{O}(n_{\text{occ}}^2 n_{\text{vir}}^2 N_X)$.

With the Laplace transformation (LT) applied to the energy denominator⁸³, the correlation energy can be rewritten as

$$E_{\text{MP2}}^{\text{RI}} = -\frac{1}{4} \int_0^\infty dt \sum_{ijab} |\langle ij||ab \rangle_{\text{RI}}|^2 e^{-t\Delta_{ij}^{ab}} \quad (48)$$

where in addition to the sum over MO indices there is also a one-dimensional (1D) quadrature over t . The number of quadrature points does not scale with system size so this will be left out in the cost analysis below⁸³. With this trick, an interesting cost reduction occurs in the

evaluation of the RI-MP2-J correlation energy, (with explicit spin-sum)

$$E_{\text{MP2-J}}^{\text{RI}} = -\frac{1}{2} \int_0^\infty dt \sum_{\sigma_1 \sigma_2} \sum_{PQ} X_{PQ}^{\sigma_1}(t) X_{PQ}^{\sigma_2}(t) \quad (49)$$

where we performed the sum over MO indices first, and

$$X_{PQ}^\sigma(t) = \sum_{i_\sigma a_\sigma} B_{i_\sigma a_\sigma}^P B_{i_\sigma a_\sigma}^Q e^{-t(\epsilon_{a_\sigma} - \epsilon_{i_\sigma})}. \quad (50)$$

The formation of \mathbf{X}^σ is the bottleneck in evaluating the RI-MP2-J correlation energy, which scales as $\mathcal{O}(n_{\text{occ}} n_{\text{vir}} N_X^2)$. This quartic scaling algorithm for the evaluation of RI-MP2-J correlation energy has been used in the scaled opposite-spin (SOS)-MP2 method⁸⁴. The opposite-spin correlation energy involves only the RI-MP2-J term and therefore can be evaluated with this algorithm at a reduced scaling.

The RI-MP2-K correlation energy, which appears only in the same-spin correlation energy, reads (with explicit spin-sum)

$$E_{\text{MP2-K}}^{\text{RI}} = \frac{1}{2} \int_0^\infty dt \left[\sum_{\sigma} \sum_{i_\sigma a_\sigma} \sum_{j_\sigma b_\sigma} \sum_P (B_{i_\sigma a_\sigma}^P B_{j_\sigma b_\sigma}^P) \sum_Q (B_{i_\sigma b_\sigma}^Q B_{j_\sigma a_\sigma}^Q) e^{-t\Delta_{i_\sigma j_\sigma}^{a_\sigma b_\sigma}} \right] \quad (51)$$

Summing over the MO indices first results in a worse computational scaling compared to the usual RI-MP2-K algorithm without the Laplace trick. Therefore, there is not much computational benefit for performing LT-RI-MP2. One may work directly with AO quantities and exploit sparsity of integrals in the AO basis set to lower the asymptotic scaling as is done in LT-AO-MP2⁸⁵ but this involves large prefactors and we do not discuss this here.

With THC-RI, further cost reductions can be achieved for both RI-MP2-J and RI-MP2-K. The THC-RI-MP2-J correlation energy reads

$$E_{\text{MP2-J}}^{\text{THC-RI}} = -\frac{1}{2} \int_0^\infty dt \sum_{KLMN} M_{KL}^{[ov]} M_{MN}^{[ov]} W_{KM}(t) \tilde{W}_{KM}(t) W_{LN}(t) \tilde{W}_{LN}(t) \quad (52)$$

where

$$W_{KM}(t) = \sum_i \phi_i^K \left(\frac{t}{2} \right) \phi_i^M \left(\frac{t}{2} \right) = \sum_i \phi_i^K(0) \phi_i^M(t) \quad (53)$$

and

$$\tilde{W}_{KM}(t) = \sum_a \phi_a^K \left(\frac{t}{2} \right) \phi_a^M \left(\frac{t}{2} \right) = \sum_a \phi_a^K(0) \phi_a^M(t) \quad (54)$$

The cost of evaluating Eq. (52) is $\mathcal{O}(N_X N_{\text{IP}}^2)$ (i.e., cubic-scaling with system size). This cubic-scaling is due to the fact that Eq. (52) involves only matrix multiplications and the formation of \mathbf{W} and $\tilde{\mathbf{W}}$ scales also cubically with system size. Therefore, with THC, the evaluation of RI-MP2-J can be completed with a cubic cost⁵⁹ in contrast to its original quartic scaling in Eq. (49).

The THC-RI-MP2-K correlation energy is given as

$$E_{\text{MP2-K}}^{\text{THC-RI}} = -\frac{1}{2} \int_0^\infty dt \sum_{KLMN} \sum_{iab} \left(M_{KL}^{[\text{ov}]} M_{MN}^{[\text{ov}]} \phi_i^K \left(\frac{t}{2} \right) \phi_i^M \left(\frac{t}{2} \right) \phi_a^K \left(\frac{t}{2} \right) \phi_a^N \left(\frac{t}{2} \right) \phi_b^L \left(\frac{t}{2} \right) \phi_b^M \left(\frac{t}{2} \right) W_{LN}(t) \right). \quad (55)$$

Assuming that $N_{\text{IP}} > N_X > n_{\text{vir}} > n_{\text{occ}}$ and no sparsity is considered, a different contraction strategy (from that of MP2-J) leads to a potentially more efficient (in terms of both storage and operations) algorithm. Starting from Eq. (55), we introduce another intermediate,

$$\tilde{B}_{ib}^{\hat{Q}}(t) = \sum_K M_{KL}^{[\text{ov}]} \phi_i^K(t) \phi_b^K(t) \quad (56)$$

where \mathbf{M} is given in Eq. (23). The formation of this intermediate scales as $\mathcal{O}(n_{\text{occ}} n_{\text{vir}} N_{\text{IP}}^2)$ and the storage requirement scales as $\mathcal{O}(n_{\text{occ}} n_{\text{vir}} N_{\text{IP}})$. Furthermore, we define a second intermediate that can be obtained from $\tilde{\mathbf{B}}$. It is defined as

$$\tilde{C}_{iN}^L(t, s) = \sum_b \tilde{B}_{ib}^L(t) \phi_b^N(s), \quad (57)$$

which requires $\mathcal{O}(n_{\text{occ}}n_{\text{vir}}N_{\text{IP}}^2)$ operations and $\mathcal{O}(n_{\text{occ}}N_{\text{IP}}^2)$ storage (potentially $\mathcal{O}(N_{\text{IP}}^2)$ with a direct algorithm). With this, we rewrite the THC-RI-MP2-K correlation energy:

$$E_{\text{MP2-K}}^{\text{THC-RI}} = -\frac{1}{2} \int_0^\infty dt \sum_{iNL} \tilde{C}_{iN}^L \left(\frac{t}{2}, \frac{t}{2} \right) \tilde{C}_{iL}^N \left(\frac{t}{2}, \frac{t}{2} \right) W_{LN}(t) \quad (58)$$

which scales cubically with system size. For the numerical simulations presented below, we adopted this algorithm. We note that summing over all the MO indices in MP2-K leads to a triple sum over interpolation points with $\mathcal{O}(N_{\text{IP}}^4)$ scaling. This alternative algorithm could be more efficient with sparsity. It will be interesting to compare these two algorithms with careful sparsity consideration in the future. We note that this algorithm is more or less the same as the THC-RI-MP2 algorithm with restricted orbitals proposed by Martínez and co-workers^{50,58}. In Algorithm 2, we illustrate step-by-step our algorithm for evaluating the THC-RI-MP2 same-spin correlation energies. The resulting algorithm scales as $\mathcal{O}(N_{\text{IP}}^3)$ and $\mathcal{O}(N_{\text{IP}}^3n_{\text{occ}})$ for THC-RI-MP2-J and THC-RI-MP2-K, respectively. Furthermore, the storage

requirement is only quadratic with system size for both energy contributions.

Algorithm 2: THC-RI-MP2 algorithm. w_t denotes a Laplace quadrature weight and E_J and E_K denote the THC-RI-MP2-J and THC-RI-MP2-K same-spin correlation energies, respectively.

```

1 for  $t = 1$  to  $n_t$ ;                                     // Loop over Laplace points.
2 do
3   Form  $\omega_i^K(t/2) = \omega_i^K e^{t/2\epsilon_i}$ ;           //  $\mathcal{O}(n_{\text{occ}}N_{\text{IP}})$ 
4   Form  $\omega_a^K(t/2) = \omega_a^K e^{-t/2\epsilon_a}$ ;         //  $\mathcal{O}(n_{\text{vir}}N_{\text{IP}})$ 
5   Form  $W_{KL}(t)$  (Eq. (53));                          //  $\mathcal{O}(n_{\text{occ}}N_{\text{IP}}^2)$ 
6   Form  $\tilde{W}_{KL}(t)$  (Eq. (54));                       //  $\mathcal{O}(n_{\text{vir}}N_{\text{IP}}^2)$ 
7   Compute  $A_{KL} = \sum_M W_{KM}\tilde{W}_{KM}M_{ML}$ ;         //  $\mathcal{O}(N_{\text{IP}}^3)$ 
8   Accumulate  $E_J -= 0.5 * w_t * \sum_{KL} A_{KL}A_{LK}$ ;     //  $\mathcal{O}(N_{\text{IP}}^2)$ 
9   for  $i = 1$  to  $n_{\text{occ}}$ ;                               // (Parallel) loop over occupied orbitals.
10  do
11    Compute  $\tilde{C}_{KL} = \sum_M \tilde{W}_{KM}\omega_i^M(t/2)M_{RL}$ ; //  $\mathcal{O}(N_{\text{IP}}^3n_{\text{occ}})$ 
12    Accumulate  $E_K += 0.5 * w_t * \sum_{KL} \tilde{C}_{KL}\tilde{C}_{LK}W_{KL}$ ; //  $\mathcal{O}(N_{\text{IP}}^2n_{\text{occ}})$ 
13  end
14 end

```

In passing we note that we used $\mathbf{M}^{[ov]}$ as opposed to $\mathbf{M}^{[nn]}$. $\mathbf{M}^{[ov]}$ is, due to its lower rank, easier to represent and fit. As such, the accuracy of MO-THC is usually higher than that of AO-THC⁶¹ for a fixed number of interpolation points compared to when compressing $\mathbf{M}^{[nn]}$. This higher accuracy comes with more difficult derivative expression and implementation of response theory.

Lastly, we briefly discuss the use of THC-RI for recently developed regularized MP2 (κ -MP2). Along with orbital optimization (OOMP2), κ -OOMP2 has been shown to be useful for thermochemistry, non-covalent interaction, and barrier heights⁸⁶. More recently, it was applied to distinguishing artificial and essential symmetry breaking⁸⁷ and singlet birad-

icaloids⁸⁸. Furthermore, using κ -OOMP2 orbitals was shown to improve the performance of MP3 greatly⁸⁹. As such, κ -OOMP2 (and κ -MP2) is a useful electron correlation model and it is worthwhile to see if the same scaling reduction can be applied to it. The RI- κ -MP2 correlation energy is

$$E_{\kappa\text{-MP2}}^{\text{RI}} = -\frac{1}{4} \sum_{ijab} \frac{|\langle ij||ab \rangle_{\text{RI}}|^2}{\Delta_{ij}^{ab}} (1 - e^{-\kappa \Delta_{ij}^{ab}})^2 \quad (59)$$

$$= -\frac{1}{4} \int_0^\infty dt \sum_{ijab} |\langle ij||ab \rangle_{\text{RI}}|^2 e^{-t \Delta_{ij}^{ab}} (1 - e^{-\kappa \Delta_{ij}^{ab}})^2 \quad (60)$$

$$= -\frac{1}{4} \int_0^\infty dt \sum_{ijab} |\langle ij||ab \rangle_{\text{RI}}|^2 (e^{-t \Delta_{ij}^{ab}} - 2e^{-(t+\kappa) \Delta_{ij}^{ab}} + e^{-(t+2\kappa) \Delta_{ij}^{ab}}) \quad (61)$$

The last expression merely suggests that the RI- κ -MP2 correlation energy may be computed as three LT-RI-MP2 energy evaluations with shifted t values and a single MO-integral evaluation. The application of ISDF to this expression is essentially identical to what was discussed above for THC-RI-MP2. Therefore, the scalings of THC-RI- κ -MP2-J and THC-RI- κ -MP2-K are cubic and quartic, respectively.

A summary of computational scaling and storage requirement for evaluating MP2-J and MP2-K correlation energies is given in Table 4.

Table 4: Computational scaling for evaluating MP2-J and MP2-K correlation energies. M here denote a measure for system size that scales linearly with the number of atoms.

RI	LT	THC	Scaling for MP2-J	Scaling for MP2-K
No	No	No	$\mathcal{O}(M^5)$	$\mathcal{O}(M^5)$
Yes	Yes	No	$\mathcal{O}(M^4)$	$\mathcal{O}(M^5)$
Yes	Yes	Yes	$\mathcal{O}(M^3)$	$\mathcal{O}(M^4)$

5 Third-Order Møller-Plesset Perturbation Theory

As noted in ref. 50, a more drastic scaling reduction is possible for the third-order Møller-Plesset perturbation theory (MP3) correlation energy evaluation. The MP3 correlation en-

ergy reads

$$E_{\text{MP3}} = E_{\text{vv}}^{(3)} + E_{\text{oo}}^{(3)} + E_{\text{ov}}^{(3)} \quad (62)$$

where the spin-orbital expressions for each term are:

$$E_{\text{vv}}^{(3)} = \frac{1}{8} \sum_{ijklcd} t_{ij}^{ab} \langle ab || cd \rangle t_{ij}^{cd} \quad (63)$$

$$E_{\text{oo}}^{(3)} = \frac{1}{8} \sum_{ijklab} t_{ij}^{ab} \langle ij || kl \rangle t_{kl}^{ab} \quad (64)$$

$$E_{\text{ov}}^{(3)} = - \sum_{ijabkc} t_{ij}^{ab} \langle ic || kb \rangle t_{kj}^{ac} \quad (65)$$

which scales as $\mathcal{O}(n_{\text{occ}}^2 n_{\text{vir}}^4)$, $\mathcal{O}(n_{\text{occ}}^4 n_{\text{vir}}^2)$, and $\mathcal{O}(n_{\text{occ}}^3 n_{\text{vir}}^3)$ with system size (i.e., sextic scaling), respectively. The application of RI does not reduce the asymptotic computational cost but may help to reduce the storage requirement.

From the THC point of view, the MP3 correlation energy poses an additional challenge that was not present in HF and MP2. Namely, the RI-MP3 correlation energy requires the representation of oo, vv, and ov blocks of the RI MO-integrals. Based on rank considerations, it is expected that the oo-block is the easiest to compress and the vv-block is the most difficult to compress. In passing, we note that $E_{\text{vv}}^{(3)}$, $E_{\text{oo}}^{(3)}$, and $E_{\text{ov}}^{(3)}$ have a total of two, two, and four, respectively, unique contributions depending on the spin-blocks of the first \mathbf{t} , the middle integral, and the second \mathbf{t} . Such unique contributions are summarized in Table 5.

Applying LT twice to Eq. (63), Eq. (64), and Eq. (65) and using the RI approximation and following the same THC factorization described in ref. 50, it is straightforward to show that every energy contribution can be computed with a quartic cost and quadratic storage. As an example, we discuss the application of THC-RI to $E_{\text{vv}}^{(3)}$. As described in Table 5, we note that $E_{\text{vv}}^{(3)}$ and $E_{\text{oo}}^{(3)}$ have only two distinct spin-block contributions: same-spin (SS) and opposite-spin (OS). Namely,

$$E_{\text{vv}}^{(3)} = E_{\text{vv-SS}}^{(3)} + E_{\text{vv-OS}}^{(3)} \quad (66)$$

Table 5: Individual energy contributions in E_{MP3} categorized by the spin-blocks of the first \mathbf{t} , the middle integral, and the second \mathbf{t} . SS means “same-spin” while OS means “opposite-spin”.

Contribution	First \mathbf{t}	Middle integral	Second \mathbf{t}
$E_{\text{vv-SS}}^{(3)}$	SS	SS	SS
$E_{\text{vv-OS}}^{(3)}$	OS	OS	OS
$E_{\text{oo-SS}}^{(3)}$	SS	SS	SS
$E_{\text{oo-OS}}^{(3)}$	OS	OS	OS
$E_{\text{ov-SS-SS-SS}}^{(3)}$	SS	SS	SS
$E_{\text{ov-OS-OS-OS}}^{(3)}$	OS	OS	OS
$E_{\text{ov-OS-OS-SS}}^{(3)}$	OS	OS	SS
$E_{\text{ov-OS-SS-OS}}^{(3)}$	OS	SS	OS

where

$$E_{\text{vv-SS}}^{(3)} = \frac{1}{8} \sum_{\sigma \in \{\alpha, \beta\}} \sum_{\substack{i_\sigma j_\sigma \\ a_\sigma b_\sigma c_\sigma d_\sigma}} t_{i_\sigma j_\sigma}^{a_\sigma b_\sigma} \langle a_\sigma b_\sigma | | c_\sigma d_\sigma \rangle t_{i_\sigma j_\sigma}^{c_\sigma d_\sigma} \quad (67)$$

$$E_{\text{vv-OS}}^{(3)} = \frac{1}{4} \sum_{\substack{\sigma, \sigma' \in \{\alpha, \beta\} \\ \sigma \neq \sigma'}} \sum_{\substack{i_\sigma j_{\sigma'} \\ a_\sigma b_{\sigma'} c_\sigma d_{\sigma'}}} t_{i_\sigma j_{\sigma'}}^{a_\sigma b_{\sigma'}} \langle a_\sigma b_{\sigma'} | | c_\sigma d_{\sigma'} \rangle t_{i_\sigma j_{\sigma'}}^{c_\sigma d_{\sigma'}} \quad (68)$$

We take an approach where we (1) sum over all the MO indices first, (2) loop over an interpolation point index, (3) store everything for a given index, and (4) accumulate the energy contributions. We loop over \hat{U} in Eq. (69) and compute intermediates accordingly:

$$E_{\text{vv-SS}}^{(3)} = \int_t \int_s \sum_{KLM} \left(A_{KL}^M(t) B_{KL}^M(t, s) - (A1)_{KL}^M(t, s) (A2)_{KL}^M(t, s) (A3)_{KL}^M(t, s) \right) \quad (69)$$

where

$$A1_{KL}^M(t, s) = \sum_M W_{KM}(t + s) M_{M\hat{U}}^{[\text{ov}]} \tilde{W}_{ML}(s) \quad (70)$$

$$A2_{KL}^M(t, s) = \sum_M M_{KM}^{[\text{ov}]} W_{M\hat{U}}(t + s) \tilde{W}_{ML}(t) \quad (71)$$

$$A3_{KL}^M(t, s) = \sum_M \tilde{W}_{KM}(t) \tilde{W}_{M\hat{U}}(s) M_{ML}^{[\text{vv}]} \quad (72)$$

This strategy ensures the quartic computational cost and quadratic storage cost without considering sparsity. More or less identical strategies can be applied to $E_{\text{oo-SS}}^{(3)}$ and $E_{\text{oo-OS}}^{(3)}$ and the resulting expressions exhibit quartic scaling. We will not explicitly write down $E_{\text{oo-SS}}^{(3)}$ and $E_{\text{oo-OS}}^{(3)}$ because they may be obtained by replacing virtuals with occupieds and vice versa in $E_{\text{vv-SS}}^{(3)}$ and $E_{\text{vv-OS}}^{(3)}$. Likewise, the ov contribution can be obtained with THC-RI using similar strategies described here.

This strategy, however, comes with a high prefactor due to the double LT (i.e., a prefactor of n_t^2). With 7 Laplace quadrature points⁸⁴, this prefactor becomes 49. As such, we found it to be difficult to observe computational benefits for medium-sized molecules this way. The large prefactor from double LT can be *completely* removed if one starts from THC-factorized first-order MP amplitudes:

$$t_{ij}^{ab} = \sum_{KL} \left(\tau_i^K \tau_a^K T_{KL} \tau_j^L \tau_b^L - \tau_i^K \tau_b^K T_{KL} \tau_j^L \tau_a^L \right) \quad (73)$$

where τ_p^K is some basis functions evaluated on a 3-space point \mathbf{r}_K and T_{KL} is defined to make Eq. (73) true in the least-squares sense. Factorizing amplitudes has been explored by several groups in the context of coupled-cluster methods^{52,62,64,66}. Our approach is essentially the same as that of ref. 52, and we apply it to the evaluation of MP3 energy. The choice for the τ basis functions may not seem immediately obvious and for simplicity we used the MOs ϕ_p^K and the same set of interpolation points as for the integral factorization. This turns out to work reasonably well for small basis sets although it takes a few more interpolation points

than the double LT algorithm described above for a fixed target accuracy.

The factorization in Eq. (73) seems more reasonable when considering

$$t_{ij}^{ab} = \int_t \left(\phi_i^K(t) \phi_a^K(t) \lambda_K M_{KL}^{[ov]} \lambda_L \phi_j^L(t) \phi_b^L(t) - \phi_i^K(t) \phi_b^K(t) \lambda_K M_{KL}^{[ov]} \lambda_L \phi_j^L(t) \phi_a^L(t) \right). \quad (74)$$

The structure of t_{ij}^{ab} is already factorized in Eq. (74) with a summation over the quadrature points. Our goal is then to find \mathbf{T} which best satisfies

$$\sum_{KL} \phi_i^K \phi_a^K \lambda_K T_{KL} \lambda_L \phi_j^L \phi_b^L = \int_t \sum_{KL} \left(\phi_i^K(t) \phi_a^K(t) \lambda_K M_{KL}^{[ov]} \lambda_L \phi_j^L(t) \phi_b^L(t) \right) \quad (75)$$

We then define a time-dependent version of Eq. (11),

$$C_{ia}^K(t) = \phi_i^K(t) \phi_a^K(t) \quad (76)$$

which is an $n_{\text{occ}} n_{\text{vir}}$ -by- N_{IP} matrix. Then, the least-squares solution for \mathbf{T} is

$$\mathbf{T} = (\tilde{\mathbf{S}}^{[\text{ov}]}(0))^+ \left(\int_t \tilde{\mathbf{S}}^{[\text{ov}]} \left(\frac{t}{2} \right) \mathbf{M}^{[\text{ov}]} \tilde{\mathbf{S}}^{[\text{ov}]} \left(\frac{t}{2} \right) \right) (\tilde{\mathbf{S}}^{[\text{ov}]}(0))^+ \quad (77)$$

where the time-dependent metric is given as (analogously to Eq. (17)),

$$\tilde{S}_{KL}^{[\text{ov}]}(t) = \sum_{ia} \lambda_K C_{ia}^K(t) C_{ia}^L(t) \lambda_L \quad (78)$$

Similar to Eq. (18), the evaluation of $\tilde{\mathbf{S}}^{[\text{ov}]}(t)$ can be done at a cubic cost. Furthermore, it is also generally a ill-conditioned, singular matrix so a threshold of 10^{-12} was used in the drop tolerance for calculating its pseudoinverse. Therefore, obtaining the THC-factorized first-order MP amplitudes scales only cubically with system size. With \mathbf{T} , one can remove the double quadrature loop in Algorithm 3 and thereby the large prefactor of n_t^2 no longer exists. Our numerical results are obtained with this additional factorization of \mathbf{t} .

The detailed algorithm for the MP3-VV correlation energy in our final THC-RI-MP3

formulation is presented in Algorithm 3.

Algorithm 3: THC-RI-MP3 algorithm for evaluating $E_{\text{vv-SS}}^{(3)}$.

```

1 Call Subroutine FormT ;                               // Form the core tensor T
2 for  $\hat{U} = 1$  to  $N_{\text{IP}}$ ;                               // Loop over interpolation points.
3 do
4   Form  $A_{KL} = \sum_M \tilde{W}_{KM} M_{M\hat{U}}^{[\text{vv}]} \tilde{W}_{ML}$ ;           //  $\mathcal{O}(N_{\text{IP}}^4)$ 
5   Form  $B_{KL} = \sum_{MN} T_{KM} \tilde{W}_{M\hat{U}} \tilde{W}_{N\hat{U}} W_{MN} T_{NL}$ ;       //  $\mathcal{O}(N_{\text{IP}}^4)$ 
6    $E += \sum_{KL} A_{KL} B_{KL}$ ;                                       //  $\mathcal{O}(N_{\text{IP}}^3)$ 
7   Form  $A1_{KL} = \sum_M W_{KM} T_{M\hat{U}} \tilde{W}_{ML}$ ;                   //  $\mathcal{O}(N_{\text{IP}}^4)$ 
8   Form  $A2_{KL} = \sum_M T_{KM} W_{M\hat{U}} \tilde{W}_{ML}$ ;                   //  $\mathcal{O}(N_{\text{IP}}^4)$ 
9   Form  $A3_{KL} = \sum_M \tilde{W}_{KM} \tilde{W}_{M\hat{U}} M_{ML}^{[\text{vv}]}$ ;           //  $\mathcal{O}(N_{\text{IP}}^4)$ 
10   $E -= \sum_{KL} A1_{KL} A2_{KL} A3_{KL}$ ;                             //  $\mathcal{O}(N_{\text{IP}}^3)$ 
11 end

```

We present a summary of computational scaling and storage for MP3 in Table 6.

Table 6: Computational scaling and storage cost for evaluating MP3 correlation energies. M here denotes a measure of molecule size that scales linearly with the number of atoms.

RI	LT	THC	Scaling for MP3
No	No	No	$\mathcal{O}(M^6)$
Yes	Yes	No	$\mathcal{O}(M^6)$
Yes	Yes	Yes	$\mathcal{O}(M^4)$

6 Computational Details

We consider two prototypical quantum chemistry benchmark sets in this work: (1) the W4-11 set for total atomization energies⁷⁶ and (2) the A24 set for non-covalent interaction energies⁷⁷. W4-11 contains a total of 140 data points and spans an energy range from 2.67 kcal/mol to 1007.9 kcal/mol. A24 contains 24 data points which accounts for energies ranging from 0.37 kcal/mol to 6.53 kcal/mol. It is crucial to examine the utility of THC on

both of the energy scales. MP2 and MP3 calculations on the A24 set were all counterpoise corrected⁹⁰.

We examine the effect of the basis set on the accuracy of THC-RI calculations for HF, ω B97X-V⁹¹ (a range-separated hybrid functional), MP2, and MP3. This was done by considering commonly used Dunning’s cc-pVDZ, cc-pVTZ, aug-cc-pVDZ, and aug-cc-pVTZ basis sets⁹²⁻⁹⁴. The pertinent RI basis sets for each AO basis set were used⁹⁵. Although there are different RI basis sets available for RI-K, we used RI basis set made for RI-MP2 for all methods. This was sufficient for the purpose of this study. We note that there has been no report of THC calculations using augmented bases previously.

In every calculation, we allowed for spin-unrestriction. For HF and ω B97X-V calculations, we performed wavefunction internal stability analysis⁹⁶ to ensure the local stability of obtained solutions. For RI-K and THC-RI-K simulations combined with HF and ω B97X-V, we use converged exact SCF solutions as a guess and did not perform the stability analysis for those. We used 99 radial points and 590 Lebedev angular points for all ω B97X-V calculations. Geometric direct minimization (GDM)⁷⁹ was used for the SCF optimizer and 10^{-5} to 10^{-6} for the root-mean-square of the orbital gradient was used for the convergence criterion. This is sufficient for single-point calculations. All correlated wavefunction calculations are based on exact SCF solutions and correlate all electrons.

The CVT procedure was performed with the parent grid of SG-1⁹⁷, which contains 50 radial points and 194 Lebedev angular points for each atom. This was found to be sufficient for the systems considered in this paper. We used the standard 7-point Laplace quadrature grid as suggested in ref. 84.

All calculations were performed with a development version of Q-Chem⁹⁸ which involves the use of (1) an integral library (`libqints`) which computes AO integrals for gaussian-type-orbitals, (2) an integral transformation library (`libposthf`) which produces RI-MO integrals, (3) a general quadrature library (`libgrid`) which takes care of the Becke grid generation, (4) an SCF library (`libgscf`) which performs an SCF calculation for arbitrary

types of orbitals, (5) a generalized many-body perturbation theory library (`libgmbpt`) which performs MP2 and MP3, and (6) an ISDF/THC library (`libisdf`) which implements THC-RI algorithms for various methods. Except `libqints`, those libraries were developed over a number of publications including the present work and interested readers are referred to refs. 86–88,99–103 for further details.

7 Results and Discussion

For various basis sets, we will investigate the utility of the THC factorization in conjunction with the ISDF point selection. We assess the accuracy of THC-RI methods by studying the root-mean-square-deviation (RMSD) and maximum unsigned deviation (MAX) from the exact RI results for both W4-11 and A24 sets. RMSD shows whether the underlying distribution of relative energies is changed by THC and MAX highlights a chemical example that is least accurately described by THC. We note that for MP2 and MP3 the exact results employ both RI and LT. The error from LT was found to be negligible in the systems studied in this paper.

We use c_{ISDF} to denote the number of interpolation points (i.e., N_{IP}) such that they are a multiple of the number of auxiliary basis functions, N_{X} :

$$N_{\text{IP}} = c_{\text{ISDF}} \times N_{\text{X}} \tag{79}$$

We scanned over $c_{\text{ISDF}} \in [2.0, 10.0]$ in the case of THC-RI-K and $c_{\text{ISDF}} \in [2.0, 5.0]$ in the case of MP2 and MP3 with an increment of 0.5.

7.1 THC-RI-K for HF and ω B97X-V

As mentioned before, the evaluation of exact exchange with MO-THC requires modification to the Fock matrix evaluation. Because of this, we chose to use AO-THC for exact exchange although a more compact THC factorization should be possible with MO-THC. As we will see

from numerical results, AO-THC still can be converged to numerically exact RI-K energies with a sufficiently large number of interpolation points.

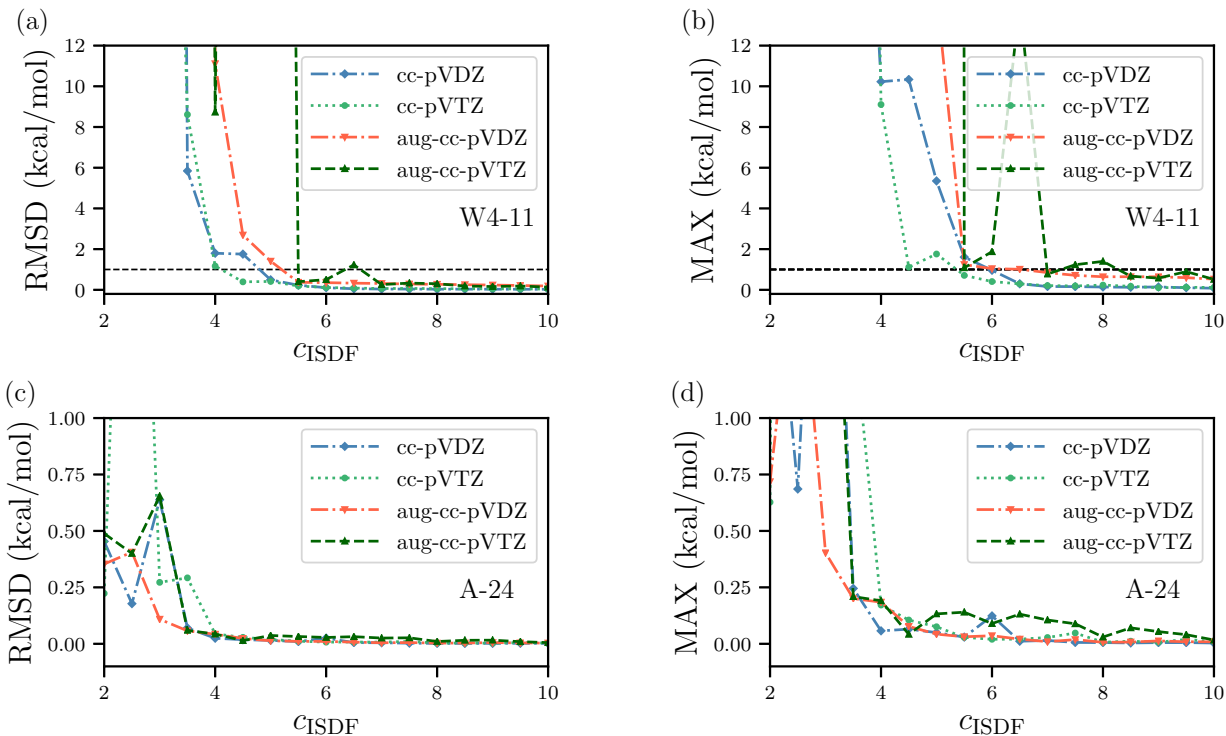


Figure 2: THC-RI-K HF errors (kcal/mol) with respect to exact RI-K HF for various basis sets: (a) root-mean-square-deviation (RMSD) for W4-11 (b) maximum unsigned deviation (MAX) for W4-11, (c) RMSD for A24, and (d) MAX for A24. Note that we used AO-THC for all calculations presented here. The dotted lines in (a) and (b) indicate 1 kcal/mol. A non-variational collapse was observed when there is a sudden jump in RMSD and/or MAX.

In Fig. 2, we assess the accuracy of THC-RI-K when used in HF calculations. Unfortunately, THC-RI-K does not guarantee variationality in the energy evaluation so we observed striking non-variational failures for small c_{ISDF} . Comparing W4-11 ((a) and (b)) and A24 ((c) and (d)) in Fig. 2, it is obvious that THC-RI-K works better in an absolute sense for non-covalent interaction energies. This is important because of the small magnitude of the A24 energy differences. It is also important to note that the factorization becomes more inaccurate with increasing size of basis set. In particular, for the largest basis set examined here (aug-cc-pVTZ) THC-RI-K exhibits a more inaccurate compression for a given c_{ISDF} than other basis sets. From rank considerations, it is not surprising that THC becomes

more inaccurate with a larger basis set. Based on (b) and (d), we conclude that $c_{\text{ISDF}} = 8.5$ approaches the accuracy of exact RI-K with an error less than 1 kcal/mol. Furthermore, this value was large enough to have an overall numerically stable SCF calculation in both W4-11 and A24.

We further assess the utility of THC-RI-K in the context of an RSH functional, $\omega\text{B97X-V}$. The exact exchange is now evaluated with 100% of the long-range Coulomb operator, $\text{erf}(\omega r_{12})/r_{12}$ along with 16.7% of the short-range Coulomb operator, $\text{erfc}(\omega r_{12})/r_{12}$ where the range-separation parameter ω is 0.3. Since the short-range contribution is substantially reduced, it is possible to see whether the accuracy of THC-RI-K varies for each component of the Coulomb operator. The assessment of accuracy of $\omega\text{B97X-V}$ with THC-RI-K is presented

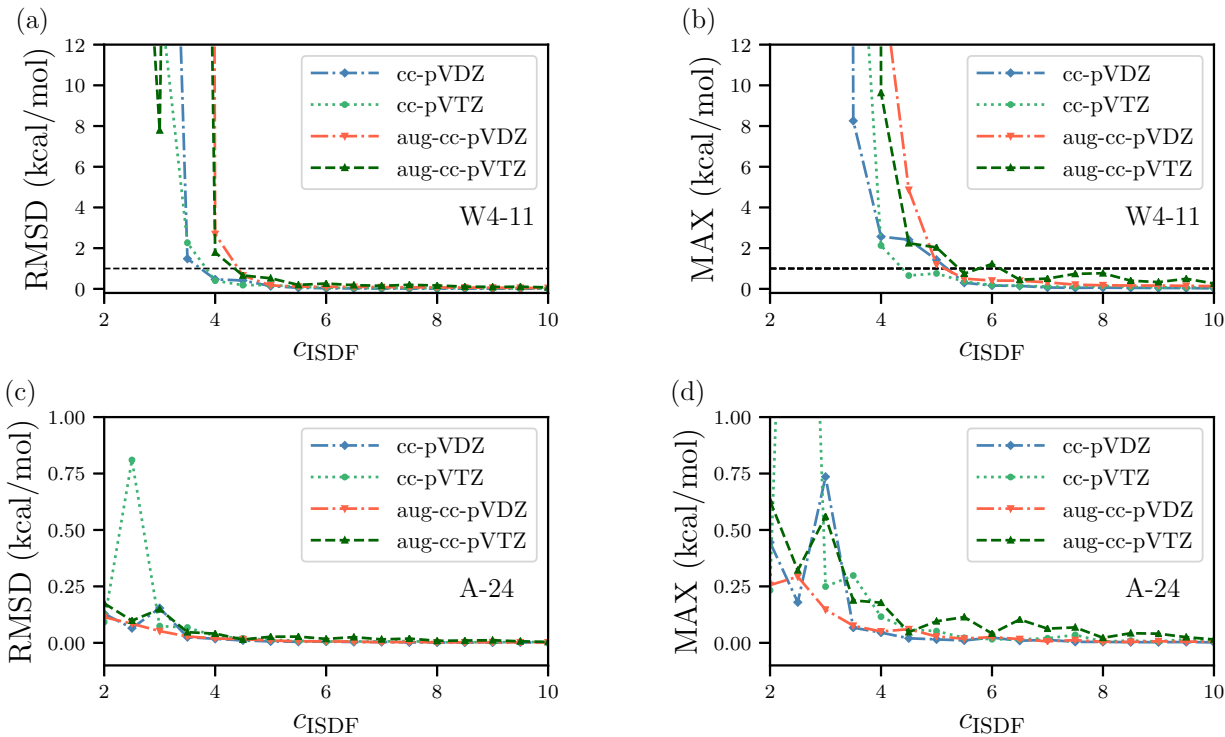


Figure 3: Same as Fig. 2 except that the results here are for $\omega\text{B97X-V}$.

in Fig. 3. Similarly to HF, variational collapse was observed in some systems for small c_{ISDF} . With sufficiently large c_{ISDF} , such instability did not occur anymore. We also note that it was easier to achieve numerical stability in THC-RI-K $\omega\text{B97X-V}$ with a smaller c_{ISDF} than that

of THC-RI-K HF. This indicates that the variational collapse may be caused by mainly the short-range component of exact exchange. Furthermore, a faster convergence with respect to c_{ISDF} is achieved with $\omega\text{B97X-V}$. This is most obvious from comparing Fig. 2 (b) and (d) with Fig. 3 (b) and (d). This suggests that it is possible that long-range exact exchange may be easier to compress than the short-range one. More thorough analysis for applying THC to range-separated Coulomb operators will be studied in the future. Our recommendation for c_{ISDF} in the case of $\omega\text{B97X-V}$ is 6.5 for all basis sets considered in this work.

7.2 THC-RI-MP2

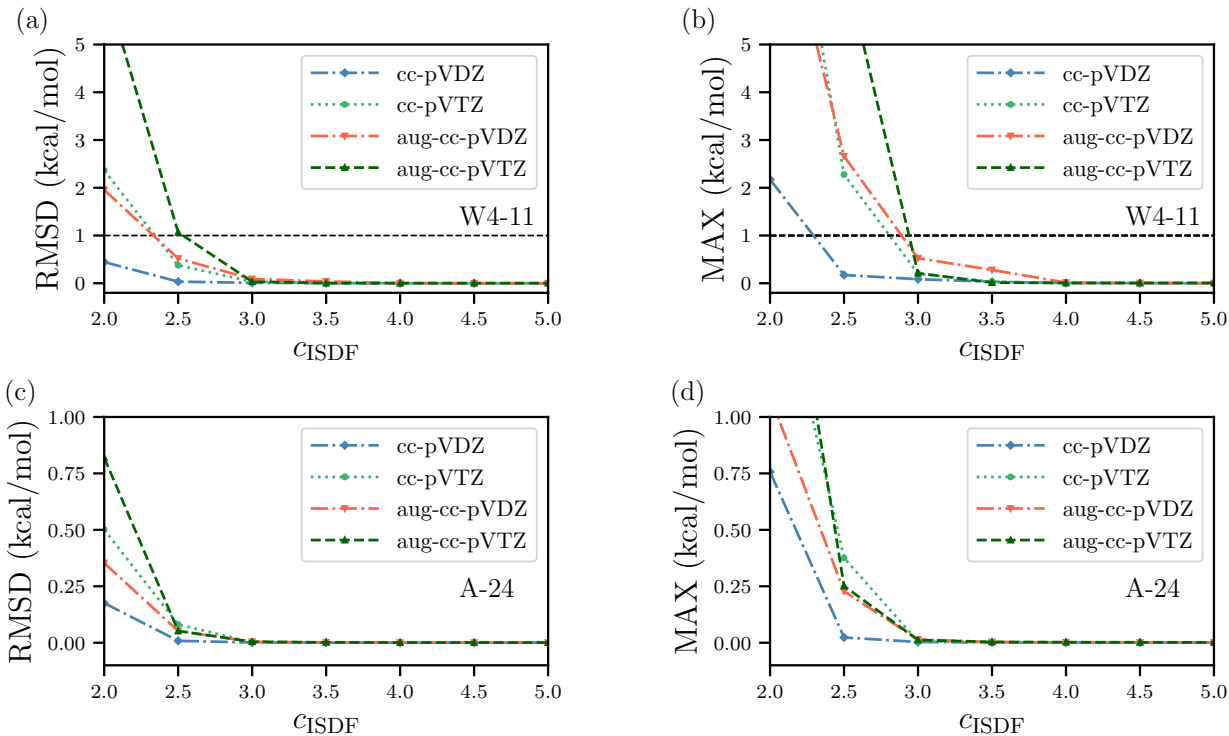


Figure 4: Same as Fig. 2 except that the results here are for MP2. Note that we used MO-THC (i.e., occupied-virtual block only) for all calculations presented here.

THC-RI-MP2 is an easier example than THC-RI-K because of the fact that one can employ MO-THC without causing numerical instability. The resulting performance of THC-RI-MP2 is exceptionally accurate for both W4-11 and A24 sets as shown in Fig. 4. With

$c_{\text{ISDF}} = 3.0$, based on RMSD THC-RI-MP2 is statistically indistinguishable from RI-MP2. As before, we observe more difficulties in the W4-11 set than in the A24 set and also with larger basis sets. Nevertheless, with $c_{\text{ISDF}} = 3.5$, regardless of the underlying AO basis sets it is possible to achieve 1 kcal/mol accuracy (W4-11 MAX) as shown in (b) and < 0.05 kcal/mol (A24 MAX) as shown in (d) of Fig. 4. We note that cc-pVDZ shows a faster convergence in both data sets compared to other larger basis sets. This is consistent with studies by Martínez and co-workers where they focused on THC-RI-MP2 with cc-pVDZ^{51,58,59}.

Based on these studies, we recommend $c_{\text{ISDF}} = 2.5$ for cc-pVDZ and $c_{\text{ISDF}} = 3.0$ for other larger basis sets for THC-RI-MP2. THC-RI-MP2 achieves remarkably accurate factorization with only about three times more interpolation points than the number of auxiliary basis functions. We emphasize that for cc-pVDZ we obtain a grid representation for THC-RI-MP2 as compact as what was obtained from hand-optimized grids for cc-pVDZ developed by Martínez and co-workers⁵⁸. In particular, for water clusters, they obtained a total of 219 optimized grid points per water molecule that achieves more or less exact accuracy. In our case, for water molecules included in the A24 set, 210 grid points per water molecule were enough to converge to RI-MP2. The real strength of ISDF point selection is that we do not have to hand-optimize the grid representation for each basis set to achieve similar compactness. The general procedure discussed in this paper can be combined with *any* available AO basis sets. Given the timing benchmarks provided by Martínez’s group⁵⁸, we expect that THC-RI-MP2 combined with ISDF will be an accurate and efficient electronic structure tool even without considering any sparsity.

In passing, we note that with AO-THC we were not able to reach the same accuracy as MO-THC with the same c_{ISDF} . Nonetheless, we were able to achieve chemical accuracy (1 kcal/mol) with $c_{\text{ISDF}} = 5.0$ for all basis sets considered in this work. In the presence of diffuse basis functions, AO-THC RI-MP2 exhibits much slower convergence to RI-MP2 with respect to c_{ISDF} . We strongly recommend MO-THC over AO-THC for THC-RI-MP2 although nuclear gradient theory can be more complicated⁶¹.

7.3 THC-RI-MP3

As mentioned in Section 5, THC-RI-MP3 is more challenging than THC-RI-MP2 and THC-RI-HF. This is because it requires an accurate representation for all of the MO blocks in the two-electron tensor.

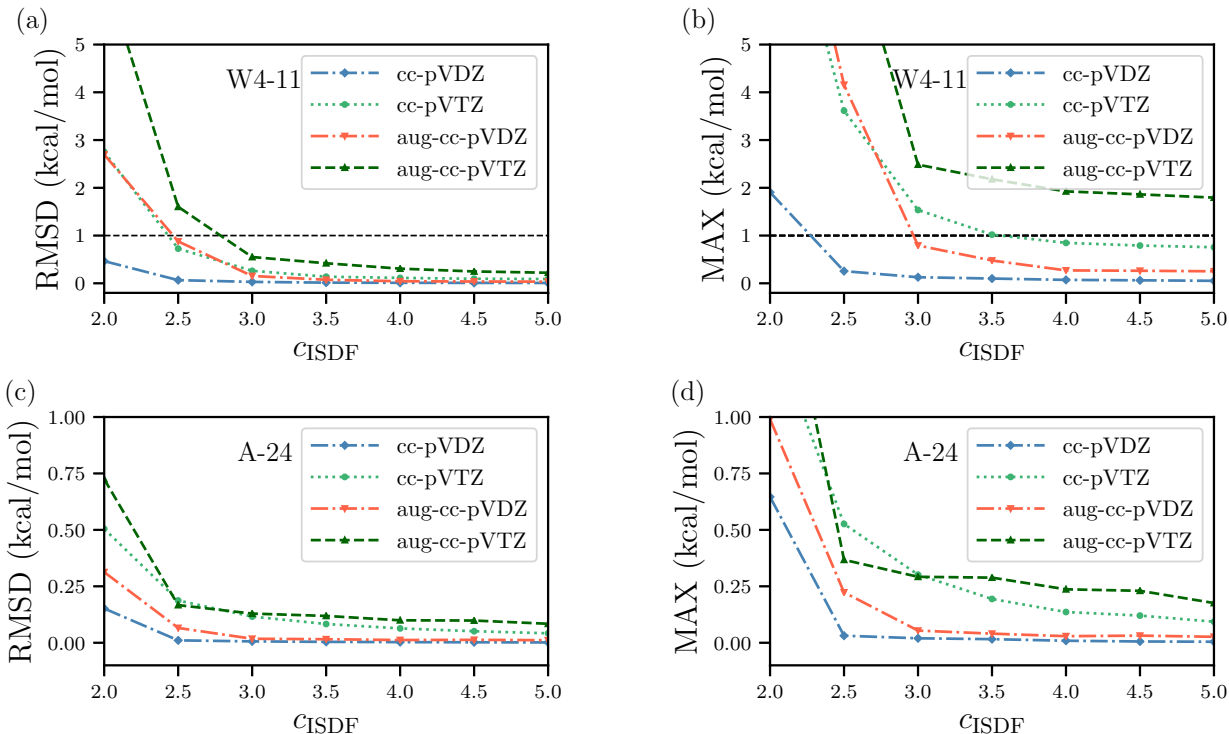


Figure 5: Same as Fig. 2 except that the results here are for MP3. Note that we used MO-THC for all calculations presented here. The slow convergence for large basis sets (cc-pVTZ and aug-cc-pVTZ) is likely due to the amplitude factorization as indicated by our investigation of THC-RI-MP2 with both factorized integrals and amplitude. See the Supporting Information for details.

In Fig. 5, we present the THC-RI-MP3 results for W4-11 and A24 sets. Based on RMSDs, with $c_{\text{ISDF}} = 3.0$ THC-RI-MP3 invokes an error below 1 kcal/mol for all basis sets. However, in its worst-performing chemical system of W4-11 (see Fig. 5 (b)), for cc-pVTZ and aug-cc-pVTZ basis sets we observe remarkably slow convergence of the energy with respect to c_{ISDF} . This data point corresponds to the atomization energy of P_4 . We performed a THC-RI-MP3 calculation with $c_{\text{ISDF}} = 10.0$ for this single data point and the error becomes 1.33 kcal/mol. A similar slow convergence was observed in A24 for ethyne dimer although the error is not

as large as that of P_4 . This is partly due to its much smaller energy scale (1.45 kcal/mol) compared to that of P_4 (290.58 kcal/mol).

We found that the amplitude factorization defined in Eq. (77) fails to faithfully represent amplitude for larger basis sets. Since we cannot perform THC-RI-MP3 without factorized amplitude due to the steep computational cost, we assessed the accuracy of amplitude factorization by evaluating the THC-RI-MP2 with both factorized integrals and amplitudes. It turns out that the combination of integral and amplitude factorizations results in far worse energies particularly for larger basis sets. Interested readers are referred to the Supporting Information for further details (see Fig. S1). It appears that direct factorization of amplitude as proposed in ref. 64 should be able to achieve greater accuracy at an increased cost via potentially difficult non-linear optimization.

The slow convergence is also associated with the metric matrix $\tilde{\mathbf{S}}$ becoming more singular with increasing size of basis set. This singularity then affects the accuracy of the least-squares fit due to thresholding in the formation of the pseudoinverse. This was not a serious problem for MP2 because the occupied-virtual metric is relatively better conditioned than that of the virtual-virtual block. It turns out that the singularity problem is not as serious for the uniform grid as is done in planewave codes^{70,74}. It will be interesting to explore a mixed grid representation between Becke’s and uniform grids to see if it helps to cope with the numerical instability.

Given these results, we recommend $c_{\text{ISDF}} = 2.5$ for cc-pVDZ and $c_{\text{ISDF}} = 4.0$ for aug-cc-pVDZ. For the larger basis sets, cc-pVTZ and aug-cc-pVTZ, we recommend $c_{\text{ISDF}} = 5.0$ based on statistical performance. However, use of THC-RI-MP3 with these basis sets requires caution due to the remarkably slow convergence rate observed in some chemical systems such as P_4 .

7.4 Summary of Recommendations

We summarize the recommendations made for each method throughout the previous sections in Table 7. The most successful application of THC-RI combined with the ISDF point

Table 7: Summary of recommendations for c_{ISDF} in THC-RI to achieve the same accuracy as various exact RI methods. For HF and $\omega\text{B97X-V}$, AO-THC was employed while MO-THC was employed for MP2 and MP3.

Basis Set	HF	$\omega\text{B97X-V}$	MP2	MP3
cc-pVDZ	8.5	6.5	2.5	2.5
cc-pVTZ	8.5	6.5	3.0	5.0
aug-cc-pVDZ	8.5	6.5	3.0	4.0
aug-cc-pVTZ	8.5	6.5	3.0	>5.0

selection was THC-RI-MP2. We shall further provide practical information for THC-RI methods. In particular, we present N_{AO} , N_{X} , and N_{IP} (of each THC-RI method) of the oxygen atom (Table 8) and the hydrogen atom (Table 9) for the basis sets considered here.

Table 8: The number of AO basis functions, auxiliary basis functions, and (recommended) interpolation points for the oxygen atom for performing THC-RI-MP2 with various basis sets.

Basis Set	N_{AO}	N_{X}	N_{IP}			
			HF	$\omega\text{B97X-V}$	MP2	MP3
cc-pVDZ	14	70	593*	455	175	175
cc-pVTZ	30	81	684*	525*	243	405
aug-cc-pVDZ	23	86	730*	559	258	344
aug-cc-pVTZ	46	106	898*	688*	318	530

* Some redundancies in the CVT process were removed so that we have fewer points than $c_{\text{ISDF}} \times N_{\text{X}}$.

Table 9: Same as Table 8 but for the hydrogen atom.

Basis Set	N_{AO}	N_{X}	N_{IP}			
			HF	$\omega\text{B97X-V}$	MP2	MP3
cc-pVDZ	5	14	119	91	35	35
cc-pVTZ	14	30	255	195	90	150
aug-cc-pVDZ	9	23	195	149	69	92
aug-cc-pVTZ	23	46	391	299	138	230

The molecules included in W4-11 and A24 are relatively small in size so it may be unclear how much compression one is achieving with those recommended values. We provide N_{AO} , N_{X} , n_{occ} , and N_{IP} for $c_{\text{ISDF}} \in [2.0, 10.0]$ with an increment of 0.5. Interested readers are referred to the Supporting Information for these parameters. We shall briefly discuss those parameters for the largest system in each of W4-11 and A24 sets. The largest molecule in W4-11 is propane (Table 10) and the one in A24 is ethene dimer (Table 11). As shown in these tables, except for the case of MP2/cc-pVDZ for propane, THC always achieves compression for the pertinent integrals even for these small molecules. In the next section, we present more numerical data for larger systems to demonstrate the power of THC combined with ISDF grids further.

Table 10: N_{AO} , N_{X} , n_{occ} , $n_{\text{occ}}n_{\text{vir}}$, N_{AO}^2 , and N_{IP} of various methods for propane in W4-11. Note that for N_{IP} we present the corresponding recommended values for each method.

Basis	N_{AO}	N_{X}	n_{occ}	$n_{\text{occ}}n_{\text{vir}}$	N_{AO}^2	N_{SP}	N_{IP}			
							HF	$\omega\text{B97X-V}$	MP2	MP3
cc-pVDZ	82	394	69	897	6724	902	3339	2559	984	984
cc-pVTZ	202	483	189	2457	40804	3061	4090	3134	1445	2410
aug-cc-pVDZ	141	514	128	1664	19881	2278	4354	3333	1537	2053
aug-cc-pVTZ	322	686	309	4017	103684	6542	5816	4451	2058	3430

Table 11: Same as Table 10, but for ethene dimer in A-24.

Basis	N_{AO}	N_{X}	n_{occ}	$n_{\text{occ}}n_{\text{vir}}$	N_{AO}^2	N_{SP}	N_{IP}			
							HF	$\omega\text{B97X-V}$	MP2	MP3
cc-pVDZ	96	336	80	1280	9216	1152	2848	2176	832	832
cc-pVTZ	232	564	216	3456	53824	3642	4776	3664	1688	2816
aug-cc-pVDZ	164	472	148	2368	26896	2920	4007	3064	1408	1880
aug-cc-pVTZ	368	792	352	5632	135424	8062	6711	5136	2372	3956

7.5 Application to Larger Systems: $(\text{H}_2\text{O})_N$ and C_{20}

With the values determined through benchmarking over the W4-11 and A24 sets, we applied $\omega\text{B97X-V}$ with THC-RI-K and THC-RI-MP2 to larger systems than those in W4-11 and A24.

The basis set used here is cc-pVTZ which should be a good compromise between accuracy and cost for DFT and MP2 calculations.

Table 12: Error per water molecule (μE_h) in the absolute energies of THC-RI-K ω B97X-V compared to RI-K ω B97X-V for $(\text{H}_2\text{O})_N$. The cc-pVTZ basis set and its corresponding auxiliary basis set are used and $c_{\text{ISDF}} = 6.5$ as determined in Section 7.4.

N	Error/ N
4	12.6
8	9.2
12	14.8
16	2.7
20	19.7

Table 13: Error per water molecule (μE_h) in the absolute energies of THC-RI-MP2 compared to RI-MP2 for $(\text{H}_2\text{O})_N$. The cc-pVTZ basis set and its corresponding auxiliary basis set are used and $c_{\text{ISDF}} = 3.0$ as determined in Section 7.4.

N	Error/ N
8	22.8
16	25.2
32	29.8

First, we applied THC to water clusters ranging from $N = 2$ to $N = 20$ for ω B97X-V and $N = 8, 16, 32$ for MP2. The geometries for $N = 8, 16$ were obtained by truncating geometries reported in ErgoSCF¹⁰⁴. Other geometries were taken from the Cambridge Cluster Database¹⁰⁵. It is shown in Table 12 and Table 13 that when growing the number of water molecules, the error per water molecule does not increase significantly. They are on the order of 10 - 30 μE_h which provides enough margin for error cancellation to achieve chemical accuracy.

Next, we also apply these methods to compute the relative energetics of four different Jahn-Teller distorted cage structures of C_{20} (C_{2h} , D_{2h} , C_i , and D_{3h}) taken from Manna and Martin’s work¹⁰⁶. As noted in refs. 106 and 87, those four structures differ only slightly in energy. In fact, at the level of MP2, there is near-degeneracy between C_i and D_{3h} cages and also between C_{2h} and D_{2h} cages. Achieving accurate integral factorization consistently over all four structures is crucial to capture this qualitative feature. In Table 14, we present the

Table 14: Error per atom (μE_h) in the absolute energies of THC-RI-K ω B97X-V and THC-RI-MP2 compared to corresponding reference methods for different structures of C_{20} . The cc-pVTZ basis set and its corresponding auxiliary basis set are used and $c_{\text{ISDF}} = 6.5$ for THC-RI-K and $c_{\text{ISDF}} = 3.0$ for THC-RI-MP2 as determined in Section 7.4.

Structure	THC-RI-K	THC-RI-MP2
C_{2h}	10.95	9.54
D_{2h}	11.35	18.94
C_i	12.88	20.02
D_{3h}	14.33	61.05

Table 15: Relative energies (kcal/mol) of four cage structures of C_{20} obtained from RI-K ω B97X-V, THC-RI-K ω B97X-V, RI-MP2, and THC-RI-MP2. The cc-pVTZ basis set and its corresponding auxiliary basis set are used and $c_{\text{ISDF}} = 6.5$ for THC-RI-K and $c_{\text{ISDF}} = 3.0$ for THC-RI-MP2 as determined in Section 7.4.

Structure	RI-K	THC-RI-K	RI-MP2	THC-RI-MP2
C_{2h}	0.17	0.14	0.83	0.70
D_{2h}	0.02	0.00	0.83	0.82
C_i	0.00	0.00	0.00	0.00
D_{3h}	0.22	0.24	0.01	0.52

absolute energy error per atom for THC-RI-K and THC-RI-MP2. THC-RI-K shows 10-15 μE_h per atom errors whereas THC-RI-MP2 error can be as large as 61 μE_h per atom. This yields the relative energies in Table 14. Both RI and THC-RI predict C_i to be the lowest structure. The error of THC-RI-K ranges from 0.02 kcal/mol to 0.03 kcal/mol which is essentially chemically indistinguishable from RI-K. On the other hand, THC-RI-MP2 does not capture the near-degeneracy between C_i and D_{3h} . Nevertheless, the difference here is on the order of 0.5 kcal/mol.

Based on these larger applications, we conclude that the numerically observed error caused by THC combined with ISDF grid points grows only linearly with system size and can achieve chemical accuracy for larger systems such as C_{20} . This suggests that one can set c_{ISDF} independent of system size to achieve a fixed accuracy per atom or molecule.

Table 16: Estimated cost ratio between conventional RI methods and THC-RI methods for a range of number of electrons (N_e) assuming stoichiometry of $(\text{H}_2\text{O})_n$ and c_{ISDF} in Table 8 and Table 9. If the ratio is greater than 1, then the THC algorithm is faster than the RI algorithm and vice versa. Note that these are based on computational bottlenecks of each algorithm and no special treatment for sparsity was assumed. Non-trivial prefactors associated with each method were neglected or simplified. For instance, the cost of THC-RI-MP3 was assumed to be $\mathcal{O}(16N_{\text{IP}}^4)$ whereas THC-RI-HF was simply assumed to be $\mathcal{O}(N_{\text{AO}}N_{\text{IP}}^2+n_{\text{occ}}N_{\text{IP}}^2)$. Similarly, we used $\mathcal{O}(7\times n_{\text{occ}}n_{\text{vir}}N_{\text{IP}}^2)$ for THC-RI-MP2, $\mathcal{O}(n_{\text{occ}}^2n_{\text{vir}}^2N_X)$ for RI-MP2, $\mathcal{O}(N_{\text{AO}}^2n_{\text{occ}}N_X)$ for RI-K, and $\mathcal{O}(n_{\text{occ}}^2n_{\text{vir}}^4+n_{\text{occ}}^4n_{\text{vir}}^2+4\times n_{\text{occ}}^3n_{\text{vir}}^3)$ for RI-MP3.

N_e	HF	$\omega\text{B97X-V}$	MP2	MP3
16	0.08	0.14	0.01	0.00
32	0.16	0.27	0.02	0.00
64	0.32	0.55	0.04	0.01
128	0.65	1.10	0.09	0.04
256	1.29	2.20	0.17	0.17
512	2.58	4.40	0.34	0.69
1024	5.16	8.79	0.68	2.76
2048	10.32	17.58	1.36	11.05
4096	20.65	35.16	2.73	44.19
8192	41.30	70.32	5.45	176.75
16384	82.59	140.64	10.90	706.99

7.6 Estimated Crossovers

RI methods are well-known to be effective for molecules up to moderately large in size, yielding shorter times to solution than conventional AO quantum chemistry, despite poorer formal scaling for HF and DFT exact exchange. THC methods, as we have explored in detail, reduce the formal scaling at the cost of quite large prefactors. It is then fairly straightforward to estimate rough crossovers for HF exchange, exact exchange in DFT ($\omega\text{B97X-V}$), MP2 and MP3. These estimates for the medium-sized cc-pVTZ basis are computed in Table 16 based on Table 8 and Table 9, using the assumption of $(\text{H}_2\text{O})_n$ stoichiometry, the recommended c_{ISDF} values already given, and operation counts for the rate-determining steps via RI and THC-RI. The main point is that in the absence of any special treatment for sparsity the crossovers are on the order of 100–200 electrons for exchange, 1000–2000 electrons for MP2, and 500–1000 electrons for MP3, which are large but not unreachable, especially for exchange and MP2.

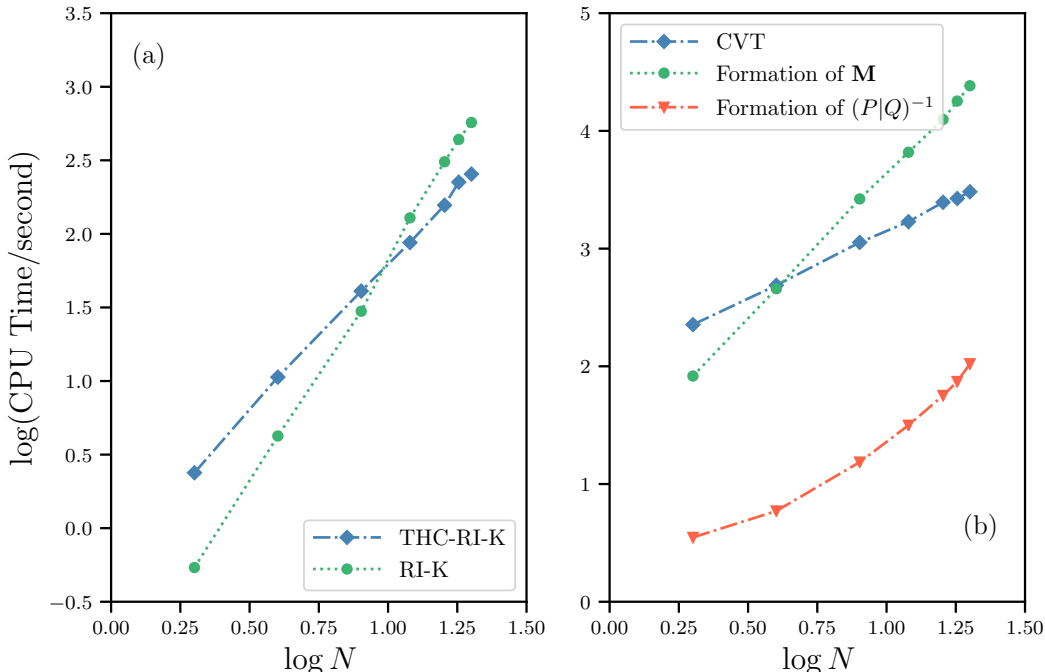


Figure 6: Measured CPU timing for RI-K, THC-RI-K and preparation steps for THC-RI-K in the case of water clusters ($N = 2, 4, 8, 12, 16, 18, 20$) with ω B97X-V/cc-pVTZ and $c_{\text{ISDF}} = 6.5$. (a) Comparison of RI-K and THC-RI-K and (b) CPU times for CVT, the formation of \mathbf{M} , and the formation of $(P|Q)^{-1}$.

As an example, we report CPU timing for running ω B97X-V on water clusters ($N = 2, 4, 8, 12, 16, 18, 20$) as shown in Fig. 6. We used 32 cores on an AMD Opteron 6376 Processor to obtain the timing results. Even with our pilot implementation (i.e., no sparsity beyond the shell-pair sparsity), the crossover between THC-RI-K and RI-K occurs at about 10 water molecules ($N_e = 80$). As shown in Fig. 6 (b), however, the preparation steps for obtaining the THC factorization takes most of the time for systems considered in this timing benchmark. Asymptotically, since the preparation steps are cubic scaling and occur only once at the beginning of the SCF, there will be eventually a crossover between THC-RI-K and RI-K including preparation steps for each. Given this promising result in Fig. 6, it will be interesting to further optimize computational kernels and seek earlier crossovers.

8 Conclusions and Outlook

In this work, we investigated the utility of tensor hypercontraction (THC) combined with the resolution-of-the-identity (RI) approximation and the interpolative separable density fitting (ISDF) grid point selection. We presented a modified ISDF point selection that can be straightforwardly combined with Becke-type atom-centered grids typically used in density functional theory calculations. Within this framework, we assessed the accuracy of THC-RI with ISDF points when applied to the evaluation of exact exchange for Hartree-Fock (HF) and a range-separated hybrid (RSH) functional ω B97X-V and the second and the third-order Møller-Plesset perturbation theory (MP2 and MP3) correlation energy. Since the framework developed in this work is general, we were able to investigate the accuracy of these methods with multiple basis sets (cc-pVDZ, cc-pVTZ, aug-cc-pVDZ and aug-cc-pVDZ) over 140 data points for total atomization energy included in the W4-11 set⁷⁶ and 24 data points for non-covalent interaction energy included in the A24 set⁷⁷. For each basis set and method, we made a recommendation for c_{ISDF} , which is a single parameter that is used in the ISDF point selection. The number of interpolation points N_{IP} is determined by $N_{\text{IP}} = c_{\text{ISDF}} \times N_{\text{X}}$ where N_{X} is the number of RI basis functions.

Exact exchange: A cubic-scaling algorithm for exact exchange (K) was formulated within the THC-RI framework. In THC-RI-K, fitting molecular orbital integrals (i.e., MO-THC) directly leads to a modified Fock matrix. Therefore, for simplicity, we investigated THC with fitting atomic orbital integrals (i.e., AO-THC). Since THC-RI-K does not guarantee variationality, subsequent SCF calculations with it often exhibits non-variational collapse. Regardless of the basis set size, we recommend $c_{\text{ISDF}} = 8.5$ and $c_{\text{ISDF}} = 6.5$ for THC-RI-K when used in HF and ω B97X-V, respectively. This was found to be sufficient in achieving chemical accuracy (1 kcal/mol) for every data point in the W4-11 benchmarks considered in this work (and much better for the A24 intermolecular interactions).

MP2: A quartic-scaling algorithm for evaluating MP2 correlation energy was formulated, which is identical to what was proposed by Martínez and co-workers. Using MO-THC, we

achieved more or less the same accuracy as RI-MP2 with $c_{\text{ISDF}} = 2.5$ for cc-pVDZ and $c_{\text{ISDF}} = 3.0$ for cc-pVTZ, aug-cc-pVDZ and aug-cc-pVTZ. Such accuracy was also achieved with hand-optimized grids developed for each basis set by Martínez and co-workers⁵⁸. We emphasize that the ISDF point selection can achieve both compactness and accuracy of the hand-optimized grids and is more general because it can be straightforwardly applied to any basis sets that a user specifies.

MP3: We developed a new quartic-scaling algorithm for MP3 that utilizes THC-factorized first-order wavefunction amplitudes. With factorized amplitudes, the THC-RI-MP3 algorithm no longer has the double Laplace transformation (LT) in it. Instead, the transformation is folded into the factorized amplitudes. As a result, this algorithm removes a prefactor of 49 (assuming 7 quadrature points for LT) in the previously developed THC-RI-MP3 algorithm⁵⁰. The assessment of this algorithm with MO-THC over W4-11 and A24 suggested that $c_{\text{ISDF}} = 2.5$ for cc-pVDZ, $c_{\text{ISDF}} = 4.0$ for aug-cc-pVDZ, and $c_{\text{ISDF}} > 5.0$ for cc-pVTZ and aug-cc-pVTZ basis sets. Since MP3 needs all MO blocks of two-electron integrals, it is more difficult for THC-RI to achieve high accuracy than it is for MP2. This was attributed to the metric matrix (particularly for the virtual-virtual block) becoming more singular with larger basis sets. Furthermore, the amplitude factorization through a simple least-squares fit was found to be insufficient for accurate results in large basis sets (aug-cc-pVTZ and cc-pVTZ).

Stress tests on larger systems: We applied THC-RI-K ω B97X-V and THC-RI-MP2 to larger systems such as $(\text{H}_2\text{O})_N$ and C_{20} . This was done by using cc-pVTZ combined with ISDF grid points generated from recommended c_{ISDF} values for this basis set. Increasing the size of water cluster, we concluded that the THC error scales linearly with system size. Based on the results from $(\text{H}_2\text{O})_N$ and C_{20} , we concluded that the THC error is about 20-60 μE_h . When combined with RI (whose error is about 50 μE_h per atom or molecule), this provides enough margin for error cancelation to achieve chemical accuracy.

Remaining challenges and outlook: Based on the ISDF point selection, it is clear that THC-RI-MP2 is easiest to make robust and accurate compared to other THC methods

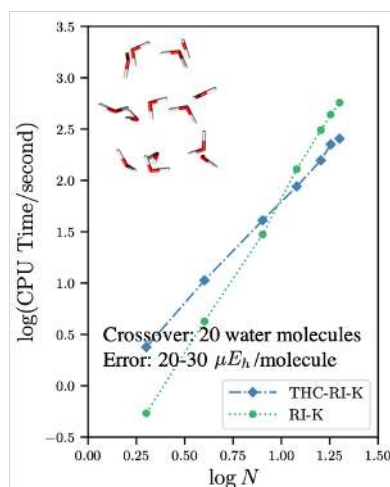
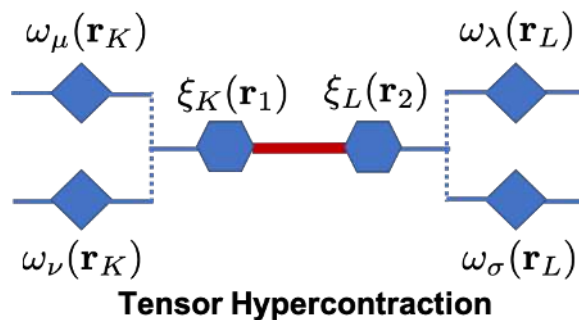
examined in this work. Although the energy evaluation presented in this work is quite straightforward to implement, its nuclear gradient can be challenging to implement within the MO-THC formalism. For exact exchange, more thorough investigation to understand frequent variational collapse is needed and further improvements are highly desirable. It may be interesting to investigate Dunlap’s variational ansatz²⁶ with the THC factorization to have a robust, variational fit. For MP3 (and other correlated methods such as coupled-cluster theory), there is a need to develop a more robust alternative fitting or representation for amplitude factorization. Furthermore, it was difficult to observe any speed-up from THC-RI-MP3 compared to RI-MP3 for the systems studied in this work (i.e., the crossover is not early). Therefore, it is important to implement THC-RI-MP3 with careful consideration of sparsity. Since a majority of THC-RI algorithms work in 3-space (where locality plays a crucial role), sparsity may greatly reduce the cost at the expense of more difficult implementation. Therefore, it is highly desirable to develop a sparsity-aware THC implementation as was done for scaled opposite spin MP2⁵⁹. It will be interesting to see whether a sparse implementation of the THC-RI-MP3 algorithm provides a substantial reduction in its cost over the conventional RI-MP3 algorithm even for medium-sized molecules.

The application of THC-RI to other methods is worthwhile such as mean-field excited state methods (e.g., configuration interaction with singles and time-dependent density functional theory), more sophisticated random phase approximations such as second-order screened exchange¹⁰⁷, coupled cluster methods⁵², orbital-optimized MP2⁸⁶, and etc. Furthermore, a theoretical question about whether short-range or long-range is easier to compress via THC will be interesting to investigate. We are currently exploring some of these issues in our group.

9 Acknowledgement

We thank Fionn Malone and Miguel Morales for pointing out the Coulomb singularity problem in THC for molecular systems in the beginning of this project. Furthermore, we thank Fionn Malone and Luke Bertels for useful discussions on the W4-11 and A24 benchmarks, Evgeny Epifanovsky for help on verifying the Becke quadrature implementation used in this work, and Srimukh Prasad for providing water cluster geometries used for timing benchmarks as well as discussions related to them. J. L. thanks Soojin Lee for consistent encouragement. This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Scientific Discovery through Advanced Computing (SciDAC) program.

10 Table of Contents Graphic



References

- (1) Pople, J. A.; Hehre, W. J. Computation of electron repulsion integrals involving contracted Gaussian basis functions. *J. Comput. Phys.* **1978**, *27*, 161–168.
- (2) Rys, J.; Dupuis, M.; King, H. F. Computation of electron repulsion integrals using the rys quadrature method. *J. Comput. Chem.* **1983**, *4*, 154–157.
- (3) Obara, S.; Saika, A. Efficient recursive computation of molecular integrals over Cartesian Gaussian functions. *J. Chem. Phys.* **1986**, *84*, 3963–3974.
- (4) Head-Gordon, M.; Pople, J. A. A method for two-electron Gaussian integral and integral derivative evaluation using recurrence relations. *J. Chem. Phys.* **1988**, *89*, 5777–5786.
- (5) Gill, P. M. W.; Head-Gordon, M.; Pople, J. A. An efficient algorithm for the generation of two-electron repulsion integrals over gaussian basis functions. *Int. J. Quantum Chem.* **1989**, *36*, 269–280.
- (6) Gill, P. M. W.; Pople, J. A. The prism algorithm for two-electron integrals. *Int. J. Quantum Chem.* **1991**, *40*, 753–772.
- (7) Gill, P. M. W.; Johnson, B. G.; Pople, J. A. Two-electron repulsion integrals over Gaussians functions. *Int. J. Quantum Chem.* **1991**, *40*, 745–752.
- (8) Almlöf, J.; Fægri Jr, K.; Korsell, K. Principles for a direct SCF approach to LICAOMOab-initio calculations. *J. Comput. Chem.* **1982**, *3*, 385–399.
- (9) Häser, M.; Ahlrichs, R. Improvements on the direct SCF method. *J. Comput. Chem.* **1989**, *10*, 104–111.
- (10) Häser, M. Møller-Plesset (MP2) perturbation theory for large molecules. *Theor. Chim. Acta.* **1993**, *87*, 147–173.

- (11) Koch, H.; Christiansen, O.; Kobayashi, R.; Jørgensen, P.; Helgaker, T. A direct atomic orbital driven implementation of the coupled cluster singles and doubles (CCSD) model. *Chem. Phys. Lett.* **1994**, *228*, 233–238.
- (12) White, C. A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M. The continuous fast multipole method. *Chem. Phys. Lett.* **1994**, *230*, 8–16.
- (13) White, C. A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M. Linear scaling density functional calculations via the continuous fast multipole method. *Chem. Phys. Lett.* **1996**, *253*, 268–278.
- (14) Ochsenfeld, C.; White, C. A.; Head-Gordon, M. Linear and sublinear scaling formation of Hartree-Fock-type exchange matrices. *J. Chem. Phys.* **1998**, *109*, 1663.
- (15) Ochsenfeld, C.; Kussmann, J.; Lambrecht, D. S. *Reviews in Computational Chemistry*; John Wiley & Sons, Ltd, 2007; Chapter 1, pp 1–82.
- (16) Doser, B.; Lambrecht, D. S.; Kussmann, J.; Ochsenfeld, C. Linear-scaling atomic orbital-based second-order Møller-Plesset perturbation theory by rigorous integral screening criteria. *J. Chem. Phys.* **2009**, *130*, 064107.
- (17) Baerends, E.; Ellis, D.; Ros, P. Self-consistent molecular Hartree-Fock-Slater calculations I. The computational procedure. *Chem. Phys.* **1973**, *2*, 41–51.
- (18) Whitten, J. L. Coulombic potential energy integrals and approximations. *J. Chem. Phys.* **1973**, *58*, 4496–4501.
- (19) Jafri, J. A.; Whitten, J. L. Electron repulsion integral approximations and error bounds: Molecular applications. *J. Chem. Phys.* **1974**, *61*, 2116–2121.
- (20) Vahtras, O.; Almlöf, J.; Feyereisen, M. W. Integral approximations for LCAO-SCF calculations. *Chem. Phys. Lett.* **1993**, *213*, 514–518.

- (21) Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. Auxiliary basis sets to approximate Coulomb potentials. *Chem. Phys. Lett.* **1995**, *240*, 283–290.
- (22) Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R. Auxiliary basis sets for main row atoms and transition metals and their use to approximate Coulomb potentials. *Theor. Chem. Acc.* **1997**, *97*, 119–124.
- (23) Dunlap, B. I.; Connolly, J. W.; Sabin, J. R. On some approximations in applications of $X\alpha$ theory. *J. Chem. Phys.* **1979**, *71*, 3396–3402.
- (24) Dunlap, B. I. Fitting the Coulomb potential variationally in $X\alpha$ molecular calculations. *J. Chem. Phys.* **1983**, *78*, 3140–3142.
- (25) Dunlap, B. I. Robust variational fitting: Gaspar’s variational exchange can accurately be treated analytically. *J. Mol. Struct. THEOCHEM* **2000**, *501-502*, 221–228.
- (26) Dunlap, B. Robust and variational fitting: Removing the four-center integrals from center stage in quantum chemistry. *J. Mol. Struct. THEOCHEM* **2000**, *529*, 37–40.
- (27) Werner, H.-J.; Manby, F. R.; Knowles, P. J. Fast linear scaling second-order Møller-Plesset perturbation theory (MP2) using local and density fitting approximations. *J. Chem. Phys.* **2003**, *118*, 8149–8160.
- (28) Polly, R.; Werner, H.-J.; Manby, F. R.; Knowles, P. J. Fast Hartree-Fock theory using local density fitting approximations. *Mol. Phys.* **2004**, *102*, 2311–2321.
- (29) Sodt, A.; Subotnik, J. E.; Head-Gordon, M. Linear scaling density fitting. *J. Chem. Phys.* **2006**, *125*, 194109.
- (30) Merlot, P.; Kjaergaard, T.; Helgaker, T.; Lindh, R.; Aquilante, F.; Reine, S.; Pedersen, T. B. Attractive electron-electron interactions within robust local fitting approximations. *J. Comput. Chem.* **2013**, *34*, 1486–1496.

- (31) Hollman, D. S.; Schaefer, H. F.; Valeev, E. F. Semi-exact concentric atomic density fitting: Reduced cost and increased accuracy compared to standard density fitting. *J. Chem. Phys.* **2014**, *140*, 064109.
- (32) Manzer, S. F.; Epifanovsky, E.; Head-Gordon, M. Efficient implementation of the pair atomic resolution of the identity approximation for exact exchange for hybrid and range-separated density functionals. *J. Chem. Theory Comput.* **2015**, *11*, 518–27.
- (33) Beebe, N. H.; Linderberg, J. Simplifications in the generation and transformation of two-electron integrals in molecular calculations. *Int. J. Quantum Chem.* **1977**, *12*, 683–705.
- (34) Früchtl, H. A.; Kendall, R. A.; Harrison, R. J.; Dyall, K. G. An implementation of RI-SCF on parallel computers. *Int. J. Quantum Chem.* **1997**, *64*, 63–69.
- (35) Manzer, S.; Horn, P. R.; Mardirossian, N.; Head-Gordon, M. Fast, accurate evaluation of exact exchange: The occ-RI-K algorithm. *J. Chem. Phys.* **2015**, *143*, 024113.
- (36) Feyereisen, M.; Fitzgerald, G.; Komornicki, A. Use of approximate integrals in ab initio theory. An application in MP2 energy calculations. *Chem. Phys. Lett.* **1993**, *208*, 359–363.
- (37) Bernholdt, D. E.; Harrison, R. J. Large-scale correlated electronic structure calculations: the RI-MP2 method on parallel computers. *Chem. Phys. Lett.* **1996**, *250*, 477–484.
- (38) Rendell, A. P.; Lee, T. J. Coupled-cluster theory employing approximate integrals: An approach to avoid the input/output and storage bottlenecks. *J. Chem. Phys.* **1994**, *101*, 400–408.
- (39) Epifanovsky, E.; Zuev, D.; Feng, X.; Khistyayev, K.; Shao, Y.; Krylov, A. I. General implementation of the resolution-of-the-identity and Cholesky representations of

- electron repulsion integrals within coupled-cluster and equation-of-motion methods: Theory and benchmarks. *J. Chem. Phys.* **2013**, *139*, 134105.
- (40) Friesner, R. A. Solution of self-consistent field electronic structure equations by a pseudospectral method. *Chem. Phys. Lett.* **1985**, *116*, 39–43.
- (41) Friesner, R. A. Solution of the Hartree-Fock equations by a pseudospectral method: Application to diatomic molecules. *J. Chem. Phys.* **1986**, *85*, 1462–1468.
- (42) Friesner, R. A. Solution of the Hartree-Fock equations for polyatomic molecules by a pseudospectral method. *J. Chem. Phys.* **1987**, *86*, 3522–3531.
- (43) Ringnalda, M. N.; Belhadj, M.; Friesner, R. A. Pseudospectral Hartree - Fock theory: Applications and algorithmic improvements. *J. Chem. Phys.* **1990**, *93*, 3397–3407.
- (44) Martínez, T. J.; Mehta, A.; Carter, E. A. Pseudospectral full configuration interaction. *J. Chem. Phys.* **1992**, *97*, 1876–1880.
- (45) Martínez, T. J.; Carter, E. A. Pseudospectral double excitation configuration interaction. *J. Chem. Phys.* **1993**, *98*, 7081–7085.
- (46) Martínez, T. J.; Carter, E. A. Pseudospectral Møller-Plesset perturbation theory through third order. *J. Chem. Phys.* **1994**, *100*, 3631–3638.
- (47) Won, Y.; Lee, J. G.; Ringnalda, M. N.; Friesner, R. A. Pseudospectral Hartree-Fock gradient calculations. *J. Chem. Phys.* **1991**, *94*, 8152–8157.
- (48) Greeley, B. H.; Russo, T. V.; Mainz, D. T.; Friesner, R. A.; Langlois, J. M.; Goddard, W. A.; Donnelly, R. E.; Ringnalda, M. N. New pseudospectral algorithms for electronic structure calculations: Length scale separation and analytical two-electron integral corrections. *J. Chem. Phys.* **1994**, *101*, 4028–4041.

- (49) Neese, F.; Wennmohs, F.; Hansen, A.; Becker, U. Efficient, approximate and parallel Hartree-Fock and hybrid DFT calculations. A 'chain-of-spheres' algorithm for the Hartree-Fock exchange. *Chem. Phys.* **2009**, *356*, 98–109.
- (50) Hohenstein, E. G.; Parrish, R. M.; Martínez, T. J. Tensor hypercontraction density fitting. I. Quartic scaling second- and third-order Møller-Plesset perturbation theory. *J. Chem. Phys.* **2012**, *137*, 1085.
- (51) Parrish, R. M.; Hohenstein, E. G.; Martínez, T. J.; Sherrill, C. D. Tensor hypercontraction. II. Least-squares renormalization. *J. Chem. Phys.* **2012**, *137*, 224106.
- (52) Hohenstein, E. G.; Parrish, R. M.; Sherrill, C. D.; Martínez, T. J. Communication: Tensor hypercontraction. III. Least-squares tensor hypercontraction for the determination of correlated wavefunctions. *J. Chem. Phys.* **2012**, *137*, 221101.
- (53) Parrish, R. M.; Hohenstein, E. G.; Martínez, T. J.; Sherrill, C. D. Discrete variable representation in electronic structure theory: Quadrature grids for least-squares tensor hypercontraction. *J. Chem. Phys.* **2013**, *138*, 194107.
- (54) Hohenstein, E. G.; Kokkila, S. I.; Parrish, R. M.; Martínez, T. J. Quartic scaling second-order approximate coupled cluster singles and doubles via tensor hypercontraction: THC-CC2. *J. Chem. Phys.* **2013**, *138*, 124111.
- (55) Hohenstein, E. G.; Kokkila, S. I.; Parrish, R. M.; Martínez, T. J. Tensor hypercontraction equation-of-motion second-order approximate coupled cluster: Electronic excitation energies in $O(N^4)$ time. *J. Phys. Chem. B* **2013**, *117*, 12972–12978.
- (56) Parrish, R. M.; Hohenstein, E. G.; Schunck, N. F.; Sherrill, C. D.; Martínez, T. J. Exact tensor hypercontraction: A universal technique for the resolution of matrix elements of local finite-range N -body potentials in many-body quantum problems. *Phys. Rev. Lett.* **2013**, *111*, 132505.

- (57) Parrish, R. M.; Sherrill, C. D.; Hohenstein, E. G.; Kokkila, S. I.; Martínez, T. J. Communication: Acceleration of coupled cluster singles and doubles via orbital-weighted least-squares tensor hypercontraction. *J. Chem. Phys.* **2014**, *140*, 181102.
- (58) Kokkila Schumacher, S. I.; Hohenstein, E. G.; Parrish, R. M.; Wang, L. P.; Martínez, T. J. Tensor Hypercontraction Second-Order Møller-Plesset Perturbation Theory: Grid Optimization and Reaction Energies. *J. Chem. Theory Comput.* **2015**, *11*, 3042–3052.
- (59) Song, C.; Martínez, T. J. Atomic orbital-based SOS-MP2 with tensor hypercontraction. I. GPU-based tensor construction and exploiting sparsity. *J. Chem. Phys.* **2016**, *144*, 174111.
- (60) Song, C.; Martínez, T. J. Atomic orbital-based SOS-MP2 with tensor hypercontraction. II. Local tensor hypercontraction. *J. Chem. Phys.* **2017**, *146*, 034104.
- (61) Song, C.; Martínez, T. J. Analytical gradients for tensor hyper-contracted MP2 and SOS-MP2 on graphical processing units. *J. Chem. Phys.* **2017**, *147*, 161723.
- (62) Benedikt, U.; Böhm, K.-H.; Auer, A. A. Tensor decomposition in post-Hartree–Fock methods. II. CCD implementation. *J. Chem. Phys.* **2013**, *139*, 224101.
- (63) Hummel, F.; Tsatsoulis, T.; Grüneis, A. Low rank factorization of the Coulomb integrals for periodic coupled cluster theory. *J. Chem. Phys.* **2017**, *146*, 124105.
- (64) Schutski, R.; Zhao, J.; Henderson, T. M.; Scuseria, G. E. Tensor-structured coupled cluster theory. *J. Chem. Phys.* **2017**, *147*, 184113.
- (65) Song, C.; Martínez, T. J. Reduced scaling CASPT2 using supporting subspaces and tensor hyper-contraction. *J. Chem. Phys.* **2018**, *149*, 044108.
- (66) Shenvi, N.; Van Aggelen, H.; Yang, Y.; Yang, W. Tensor hypercontracted ppRPA:

- Reducing the cost of the particle-particle random phase approximation from $O(r^6)$ to $O(r^4)$. *J. Chem. Phys.* **2014**, *141*, 024119.
- (67) Duchemin, I.; Blase, X. Separable resolution-of-the-identity with all-electron Gaussian bases: Application to cubic-scaling RPA. *J. Chem. Phys.* **2019**, *150*, 174120.
- (68) Lu, J.; Ying, L. Compression of the electron repulsion integral tensor in tensor hypercontraction format with cubic scaling cost. *J. Comput. Phys.* **2015**, *302*, 329–335.
- (69) Dong, K.; Hu, W.; Lin, L. Interpolative Separable Density Fitting through Centroidal Voronoi Tessellation with Applications to Hybrid Functional Electronic Structure Calculations. *J. Chem. Theory Comput.* **2018**, *14*, 1311–1320.
- (70) Hu, W.; Lin, L.; Yang, C. Interpolative Separable Density Fitting Decomposition for Accelerating Hybrid Density Functional Calculations with Applications to Defects in Silicon. *J. Chem. Theory Comput.* **2017**, *13*, 5420–5431.
- (71) Lu, J.; Thicke, K. Cubic scaling algorithms for RPA correlation using interpolative separable density fitting. *J. Comput. Phys.* **2017**, *351*, 187–202.
- (72) Hu, W.; Shao, M.; Cepellotti, A.; da Jornada, F. H.; Lin, L.; Thicke, K.; Yang, C.; Louie, S. G. Accelerating Optical Absorption Spectra and Exciton Energy Computation via Interpolative Separable Density Fitting. *Lecture Notes in Computer Science* (including subseries *Lecture Notes in Artificial Intelligence* and *Lecture Notes in Bioinformatics*). 2018; pp 604–617.
- (73) Lin, L.; Xu, Z.; Ying, L. Adaptively compressed polarizability operator for accelerating large scale ab initio phonon calculations. *Multiscale Model Sim.* **2017**, *15*, 29–55.
- (74) Malone, F. D.; Zhang, S.; Morales, M. A. Overcoming the Memory Bottleneck in Auxiliary Field Quantum Monte Carlo Simulations with Interpolative Separable Density Fitting. *J. Chem. Theory Comput.* **2018**, *15*, 256–264.

- (75) Becke, A. D. A multicenter numerical integration scheme for polyatomic molecules. *J. Chem. Phys.* **1988**, *88*, 2547.
- (76) Karton, A.; Daon, S.; Martin, J. M. W4-11: A high-confidence benchmark dataset for computational thermochemistry derived from first-principles W4 data. *Chem. Phys. Lett.* **2011**, *510*, 165–178.
- (77) Řezáč, J.; Hobza, P. Describing Noncovalent Interactions beyond the Common Approximations: How Accurate Is the Gold Standard, CCSD(T) at the Complete Basis Set Limit? *J. Chem. Theory Comput.* **2013**, *9*, 2151–2155.
- (78) Scuseria, G. E. Linear scaling density functional calculations with Gaussian orbitals. *J. Phys. Chem. A* **1999**, *103*, 4782–4790.
- (79) Van Voorhis, T.; Head-Gordon, M. A geometric approach to direct minimization. *Mol. Phys.* **2002**, *100*, 1713–1721.
- (80) Pulay, P. Convergence acceleration of iterative sequences. the case of scf iteration. *Chem. Phys. Lett.* **1980**, *73*, 393–398.
- (81) Pulay, P. Improved SCF convergence acceleration. *J. Comput. Chem* **1982**, *3*, 556–560.
- (82) Wirz, L. N.; Reine, S. S.; Pedersen, T. B. On Resolution-of-the-Identity Electron Repulsion Integral Approximations and Variational Stability. *J. Chem. Theory Comput.* **2017**, *13*, 4897–4906.
- (83) Häser, M.; Almlöf, J. Laplace transform techniques in Møller–Plesset perturbation theory. *J. Chem. Phys.* **1992**, *96*, 489–494.
- (84) Jung, Y.; Lochan, R. C.; Dutoi, A. D.; Head-Gordon, M. Scaled opposite-spin second order Møller–Plesset correlation energy: an economical electronic structure method. *J. Chem. Phys.* **2004**, *121*, 9793–9802.

- (85) Ayala, P. Y.; Scuseria, G. E. Linear scaling second-order Moller–Plesset theory in the atomic orbital basis for large molecular systems. *J. Chem. Phys.* **1999**, *110*, 3660–3671.
- (86) Lee, J.; Head-Gordon, M. Regularized Orbital-Optimized Second-Order Møller–Plesset Perturbation Theory: A Reliable Fifth-Order-Scaling Electron Correlation Model with Orbital Energy Dependent Regularizers. *J. Chem. Theory Comput.* **2018**, *14*, 5203–5219.
- (87) Lee, J.; Head-Gordon, M. Distinguishing artificial and essential symmetry breaking in a single determinant: approach and application to the C₆₀, C₃₆, and C₂₀ fullerenes. *Phys. Chem. Chem. Phys.* **2019**, *21*, 4763–4778.
- (88) Lee, J.; Head-Gordon, M. Two single-reference approaches to singlet biradicaloid problems: Complex, restricted orbitals and approximate spin-projection combined with regularized orbital-optimized Møller-Plesset perturbation theory. *J. Chem. Phys.* **2019**, *150*, 244106.
- (89) Bertels, L. W.; Lee, J.; Head-Gordon, M. Third-Order Møller-Plesset Perturbation Theory Made Useful? Choice of Orbitals and Scaling Greatly Improves Accuracy for Thermochemistry, Kinetics, and Intermolecular Interactions. *J. Phys. Chem. Lett.* **2019**, *10*, 4170–4176.
- (90) Boys, S. F.; Bernardi, F. d. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **1970**, *19*, 553–566.
- (91) Mardirossian, N.; Head-Gordon, M. ω B97X-V: A 10-parameter, range-separated hybrid, generalized gradient approximation density functional with nonlocal correlation, designed by a survival-of-the-fittest strategy. *Phys. Chem. Chem. Phys.* **2014**, *16*, 9904–9924.

- (92) Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
- (93) Woon, D. E.; Dunning Jr, T. H. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *J. Chem. Phys.* **1993**, *98*, 1358–1371.
- (94) Kendall, R. A.; Dunning Jr, T. H.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96*, 6796–6806.
- (95) Weigend, F.; Köhn, A.; Hättig, C. Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations. *J. Chem. Phys.* **2002**, *116*, 3175–3183.
- (96) Seeger, R.; Pople, J. A. Self-consistent molecular orbital methods. XVIII. Constraints and stability in Hartree-Fock theory. *J. Chem. Phys.* **1977**, *66*, 3045–3050.
- (97) Gill, P. M. W.; Johnson, B. G.; Pople, J. A. A standard grid for density functional calculations. *Chem. Phys. Lett.* **1993**, *209*, 506–512.
- (98) Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X.; Ghosh, D.; Goldey, M.; Horn, P. R.; Jacobson, L. D.; Kaliman, I.; Khaliullin, R. Z.; Kuš, T.; Landau, A.; Liu, J.; Proynov, E. I.; Rhee, Y. M.; Richard, R. M.; Rohrdanz, M. A.; Steele, R. P.; Sundstrom, E. J.; Woodcock, H. L.; Zimmerman, P. M.; Zuev, D.; Albrecht, B.; Alguire, E.; Austin, B.; Beran, G. J.; Bernard, Y. A.; Berquist, E.; Brandhorst, K.; Bravaya, K. B.; Brown, S. T.; Casanova, D.; Chang, C. M.; Chen, Y.; Chien, S. H.; Closser, K. D.; Crittenden, D. L.; Diedenhofen, M.; Distasio, R. A.; Do, H.; Dutoi, A. D.; Edgar, R. G.; Fatehi, S.; Fusti-Molnar, L.; Ghysels, A.; Golubeva-Zadorozhnaya, A.; Gomes, J.; Hanson-Heine, M. W.; Harbach, P. H.; Hauser, A. W.; Hohenstein, E. G.; Holden, Z. C.;

Jagau, T. C.; Ji, H.; Kaduk, B.; Khistyayev, K.; Kim, J.; Kim, J.; King, R. A.; Klunzinger, P.; Kosenkov, D.; Kowalczyk, T.; Krauter, C. M.; Lao, K. U.; Laurent, A. D.; Lawler, K. V.; Levchenko, S. V.; Lin, C. Y.; Liu, F.; Livshits, E.; Lochan, R. C.; Luenser, A.; Manohar, P.; Manzer, S. F.; Mao, S. P.; Mardirossian, N.; Marenich, A. V.; Maurer, S. A.; Mayhall, N. J.; Neuscamman, E.; Oana, C. M.; Olivares-Amaya, R.; Oneill, D. P.; Parkhill, J. A.; Perrine, T. M.; Peverati, R.; Prociuk, A.; Rehn, D. R.; Rosta, E.; Russ, N. J.; Sharada, S. M.; Sharma, S.; Small, D. W.; Sodt, A.; Stein, T.; Stück, D.; Su, Y. C.; Thom, A. J.; Tsuchimochi, T.; Vanovschi, V.; Vogt, L.; Vydrov, O.; Wang, T.; Watson, M. A.; Wenzel, J.; White, A.; Williams, C. F.; Yang, J.; Yeganeh, S.; Yost, S. R.; You, Z. Q.; Zhang, I. Y.; Zhang, X.; Zhao, Y.; Brooks, B. R.; Chan, G. K.; Chipman, D. M.; Cramer, C. J.; Goddard, W. A.; Gordon, M. S.; Hehre, W. J.; Klamt, A.; Schaefer, H. F.; Schmidt, M. W.; Sherrill, C. D.; Truhlar, D. G.; Warshel, A.; Xu, X.; Aspuru-Guzik, A.; Baer, R.; Bell, A. T.; Besley, N. A.; Chai, J. D.; Dreuw, A.; Dunietz, B. D.; Furlani, T. R.; Gwaltney, S. R.; Hsu, C. P.; Jung, Y.; Kong, J.; Lambrecht, D. S.; Liang, W.; Ochsenfeld, C.; Rassolov, V. A.; Slipchenko, L. V.; Subotnik, J. E.; Van Voorhis, T.; Herbert, J. M.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Mol. Phys.* **2015**, *113*, 184–215.

- (99) Lee, J.; Small, D. W.; Epifanovsky, E.; Head-Gordon, M. Coupled-cluster valence-bond singles and doubles for strongly correlated systems: Block-tensor based implementation and application to oligoacenes. *J. Chem. Theory Comput.* **2017**, *13*, 602–615.
- (100) Lee, J.; Small, D. W.; Head-Gordon, M. Open-shell coupled-cluster valence-bond theory augmented with an independent amplitude approximation for three-pair correlations: Application to a model oxygen-evolving complex and single molecular magnet. *J. Chem. Phys.* **2018**, *149*, 244121.
- (101) Lee, J.; Huggins, W. J.; Head-Gordon, M.; Whaley, K. B. Generalized Unitary Coupled

- Cluster Wave functions for Quantum Computation. *J. Chem. Theory Comput.* **2019**, *15*, 311–324.
- (102) Lee, J.; Bertels, L. W.; Small, D. W.; Head-Gordon, M. Kohn-sham density functional theory with complex, spin-restricted orbitals: Accessing a new class of densities without the symmetry dilemma. *Phys. Rev. Lett.* **2019**, *123*, 113001.
- (103) Lee, J.; Malone, F. D.; Morales, M. A. An auxiliary-Field quantum Monte Carlo perspective on the ground state of the dense uniform electron gas: An investigation with Hartree-Fock trial wavefunctions. *J. Chem. Phys.* **2019**, *151*, 064122.
- (104) ErgoSCF. <http://www.ergoscf.org/xyz/h2o.php>, Accessed: 2019-09-01.
- (105) Cambridge Cluster Databse. <http://www-wales.ch.cam.ac.uk/CCD.html>, Accessed: 2019-09-01.
- (106) Manna, D.; Martin, J. M. L. What Are the Ground State Structures of C₂₀ and C₂₄? An Explicitly Correlated Ab Initio Approach. *J. Phys. Chem. A* **2016**, *120*, 153–160.
- (107) Grüneis, A.; Marsman, M.; Harl, J.; Schimka, L.; Kresse, G. Making the random phase approximation to electronic correlation accurate. *J. Chem. Phys.* **2009**, *131*, 154115.