The Journal of Chemical Physics

Low rank factorization of the Coulomb integrals for periodic coupled cluster theory

Cite as: J. Chem. Phys. **146**, 124105 (2017); https://doi.org/10.1063/1.4977994
Submitted: 10 November 2016 • Accepted: 21 February 2017 • Published Online: 22 March 2017

Felix Hummel, Theodoros Tsatsoulis and Andreas Grüneis

Low rank factorization of the Coulomb integrals for periodic coupled cluster theory

Felix Hummel,[1,a)] Theodoros Tsatsoulis,[1] and Andreas Grüneis[1,2,b)]
[1]*Max Planck Institute for Solid State Research, Heisenbergstraße 1, D-70569 Stuttgart, Germany*
[2]*Department Chemie, Technische Universität München (TUM), Lichtenbergstraße 4, D-85747 Garching, Germany*

(Received 10 November 2016; accepted 21 February 2017; published online 22 March 2017)

We study a tensor hypercontraction decomposition of the Coulomb integrals of periodic systems where the integrals are factorized into a contraction of six matrices of which only two are distinct. We find that the Coulomb integrals can be well approximated in this form already with small matrices compared to the number of real space grid points. The cost of computing the matrices scales as $\mathcal{O}(N^4)$ using a regularized form of the alternating least squares algorithm. The studied factorization of the Coulomb integrals can be exploited to reduce the scaling of the computational cost of expensive tensor contractions appearing in the amplitude equations of coupled cluster methods with respect to system size. We apply the developed methodologies to calculate the adsorption energy of a single water molecule on a hexagonal boron nitride monolayer in a plane wave basis set and periodic boundary conditions.
[http://dx.doi.org/10.1063/1.4977994]

I. INTRODUCTION

The high dimensionality of the many electron wave function is one of the most limiting factors in applying highly accurate electronic structure theories to the solution of the many electron Schrödinger equation for real materials on an *ab initio* level.[1] Many of the most widely used wave function based theories have a good balance between accuracy, computational cost, and the number of parameters used to approximate the exact wave function. The efficiency is strongly affected by the computational complexity required to evaluate resultant expectation values.

Tensor rank decompositions (TRDs) and low rank tensor approximations are ubiquitous in the field of electronic structure theory calculations. These techniques are essential to reduce the computational cost and memory footprints to calculate and store the approximate many electron wave function. Already the simplest level of approximation, Hartree–Fock (HF) theory can be regarded as a low rank approximation, employing an antisymmetrized outer product of one electron orbitals to approximate the full wave function. This low rank tensor approximation is identical to a single Slater determinant. However, HF theory neglects electronic correlation effects. Electronic correlation effects can be captured by extending the wave function basis with additional determinants. For this purpose, excited HF determinants can be employed. They are constructed by replacing occupied orbitals with unoccupied orbitals, forming a complete and orthogonal basis. Computationally, the basis of (excited) Slater determinants is very convenient. It introduces a large degree of sparsity to the full many electron Hamiltonian and simplifies the solution of the many electron problem. Most entries in the sparse many electron Hamiltonian can be calculated directly from electron repulsion integrals. The memory footprint for the storage of these integrals in a canonical basis is very large and grows rapidly with respect to the number of orbitals. Therefore it is often necessary to calculate these integrals in a computationally efficient on-the-fly manner. The most widely used schemes for the calculation of electron repulsion integrals include the following:[2–6] (i) prior calculation of the integrals in the employed atomic orbital basis and its subsequent transformation into a molecular orbital basis and (ii) employing the resolution of identity approach. Computationally the resolution of identity approach is more efficient because it requires the calculation and storage of intermediate quantities with at most three indices. In passing we note that the expression of the integrals in terms of these intermediate quantities allows for rearranging nested summations in ring coupled cluster theories such that the scaling of the computational cost with respect to the system size can be reduced.[7]

Coupled cluster theory can also be viewed as a low rank tensor approximation to the exact configuration interaction wave function coefficients in the Slater determinant basis. The exponential ansatz used in coupled cluster theories effectively approximates the coefficients of highly excited determinants by outer products of cluster amplitudes with a lower rank. However, increasingly accurate levels of coupled cluster theories lead to increasingly steep polynomial scalings of the computational cost and memory with respect to the studied system sizes. In this work, we seek to reduce the computational cost of coupled cluster theories without introducing additional approximations on the level of the employed wave function. This can be achieved by employing low rank tensor

a)Electronic mail: f.hummel@fkf.mpg.de
b)Electronic mail: a.grueneis@fkf.mpg.de

approximation techniques for the decomposition of the two electron integrals and the corresponding intermediate quantities obtained from the resolution of identity approach. Using the low rank decomposition, the nested summations in the amplitude equations of distinguishable cluster doubles (DCD) and distinguishable cluster singles and doubles (DCSD) theories,[8–10] as well as of linearized coupled cluster singles and doubles (CCSD) theory can be rearranged such that the scaling of the computational cost is reduced from $\mathcal{O}(N^6)$ to $\mathcal{O}(N^5)$ without any further approximations. Additionally, we highlight that the low rank factorization of the Coulomb integrals also allows for reducing the scaling of the computationally most expensive terms in coupled cluster doubles (CCD) and coupled cluster singles and doubles (CCSD) theories using a plane wave basis set.

We note that the methods outlined in this work share many similarities with other approaches that aim at the reduction of the computational cost in correlated wave function based theories. In particular we want to point out that the tensor hypercontraction (THC) technique introduced by Hohenstein *et al.* in Refs. 11–13 also performs a low rank tensor decomposition of the Coulomb integrals. Furthermore, a similar approach to the tensor rank decomposition method introduced in this work was discussed by Shenvi *et al.* in Ref. 14. In the work of Benedikt *et al.*, it was shown that tensor rank decomposition techniques can even be applied to the decomposition of the coupled cluster amplitudes directly.[15] The previous work on THC in Ref. 11 forms the basis for the methodological developments outlined in the present manuscript. We seek to improve upon the efficiency and convergence of the employed algorithm for obtaining the low rank tensor approximation. Furthermore we aim at applying these methods to periodic systems. We note in passing that the more recent work on THC for molecular systems achieves highly efficient implementations of coupled cluster and other perturbation theories by employing a real space quadrature grid to determine one of the two factor matrices.

A. Structure of this work

The factorization of the Coulomb integrals tensor is obtained in two steps. In Section II we first discuss how the Coulomb integrals can be decomposed into a contraction of two third order tensors: $V^{pq}_{sr} \approx \Gamma^{*pF}_{s}\Gamma^{q}_{rF}$, where we refer to Γ^{q}_{rF} as the optimized Coulomb vertex. Subsequently, we perform a Tensor Rank Decomposition (TRD) of the optimized Coulomb vertex into a contraction of three matrices: $\Gamma^{q}_{rF} \approx \Lambda^{R}_{F}\Pi^{*q}_{R}\Pi^{R}_{r}$. Section III describes the employed algorithms to compute this factorization.

Section IV outlines how this factorization can be employed by quantum chemistry methods and in Section V we study the application of the discussed approximations to different systems. Subsection V A focuses on the convergence of the TRD for total energies of the LiH solid, while we compute coupled cluster adsorption energies of water on the surface of a single BN sheet in Subsections V B and V C.

B. Notation

We imply a sum over all free indices occurring at least twice within a product but nowhere else. We will use the letters i, j, k, l to label occupied spin orbitals, a, b, c, d to label virtual spin orbitals, and p, q, r, s to label general spin orbitals. The letters R, S, T, U are used to denote elements of the rank decomposition. The conjugate transpose of a tensor such as A^{q}_{r} is denoted by A^{*r}_{q} where lower and upper indices are swapped. Sequence numbers in iterations are given in superscript within parentheses, as in $A^{(n)}$. The Frobenius norm of a tensor A is denoted by $\|A\|$. Examples are as follows:

$$V^{ab}_{ij}V^{*ij}_{ab} = \sum_{a,b\in\text{virt.},i,j\in\text{occ.}} V^{ab}_{ij}V^{*ij}_{ab} \tag{1}$$

$$T_{ijk} - A_{iR}B_{jR}C_{kR} = T_{ijk} - \sum_{R=1}^{N_R} A_{iR}B_{jR}C_{kR}, \tag{2}$$

$$\|\Gamma^{q}_{rF}\|^2 = \sum_{q,r,F} \Gamma^{*rF}_{q}\Gamma^{q}_{rF}, \tag{3}$$

II. OPTIMIZED AUXILIARY FIELD APPROXIMATION

In this section, we discuss how to approximate the Coulomb integrals, a tensor of fourth order, by a contraction of two considerably smaller tensors of third order: $V^{pq}_{sr} \approx \Gamma^{*pF}_{s}\Gamma^{q}_{rF}$, without actually calculating the entire tensor V^{pq}_{sr}.

Given the spin orbitals $\psi_q(\mathbf{x})$ from a Hartree–Fock (HF) or density functional theory (DFT) calculation, the (nonantisymmetrized) Coulomb integrals are defined by

$$V^{pq}_{sr} = \int\int \mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{x}'\,\psi^{*p}(\mathbf{x})\psi^{*q}(\mathbf{x}')\frac{1}{|\mathbf{r}-\mathbf{r}'|}\psi_r(\mathbf{x}')\psi_s(\mathbf{x}), \tag{4}$$

with $\mathbf{x} = (\sigma, \mathbf{r})$ and $\int \mathrm{d}\mathbf{x} = \sum_\sigma \int \mathrm{d}\mathbf{r}$. Owing to the translational invariance of the Coulomb kernel, we can separate the Coulomb integrals as follows:

$$V^{pq}_{sr} = \int \frac{\mathrm{d}\mathbf{G}}{(2\pi)^3}\,\Gamma^{*p}_{s}(\mathbf{G})\Gamma^{q}_{r}(\mathbf{G}), \tag{5}$$

where the *Coulomb vertex* $\Gamma^{q}_{r}(\mathbf{G})$ is given by

$$\Gamma^{q}_{r}(\mathbf{G}) := \sqrt{\frac{4\pi}{\mathbf{G}^2}}\int \mathrm{d}\mathbf{x}\,\mathrm{e}^{-\mathrm{i}\mathbf{G}\cdot\mathbf{r}}\,\psi^{*q}(\mathbf{x})\psi_r(\mathbf{x}). \tag{6}$$

The above equation can be derived directly from Eq. (4) by using Fourier transforms and provides the physical justification for the factorization of the Coulomb vertex that we will investigate in Section III. This factorization is also referred to as tensor hypercontraction (THC) and employed in a wide range of related studies that form the foundation for the present work.[11–13,16–18] We let the discretization of Eq. (6) be $\tilde{\Gamma}^{q}_{rG} = \sqrt{w_G}\,\Gamma^{q}_{r}(\mathbf{G}_G)$ with the momentum grid points $\mathbf{G}_G$ and the numerical integration weights w_G such that

$$\int \frac{\mathrm{d}\mathbf{G}}{(2\pi)^3}\,\Gamma^{*p}_{s}(\mathbf{G})\Gamma^{q}_{r}(\mathbf{G}) \approx \sum_{G=1}^{N_G}\tilde{\Gamma}^{*pG}_{s}\tilde{\Gamma}^{q}_{rG}\,. \tag{7}$$

We note in passing that the above approximation is of the same algebraic form as obtained by Cholesky decomposition[19–21] or density fitting[22–24] methods for the calculation of the Coulomb integrals employing atom centered basis sets. However, in the case of plane wave basis sets, Eq. (7) can be

derived directly from Eq. (4) and a series of Fourier transforms. The Coulomb vertex can be computed from the spin orbitals in $\mathcal{O}(N_p^2 N_G \log N_r)$ time employing a Fast Fourier Transform (FFT), where N_p, N_G, and N_r denote the number of spin orbitals, momentum grid points, and real space grid points, respectively. We note that the orbital overlap charge density $\psi^{*q}(\mathbf{x})\psi_r(\mathbf{x})$ is approximated in the projector augmented wave method using Eq. (2.87) of Ref. 25 as implemented in the Vienna *ab initio* simulation package (VASP).[26–28]

In general, any two-body operator can be split into a product of two single body operators coupled by an auxiliary field.[29] In the case of the Coulomb interaction, the auxiliary field has only one variable due to translational invariance, which is the momentum $\mathbf{G}$ mediated by the interaction. Although $\tilde{\Gamma}^q_{rG}$ is in practice, already a third order tensor, the number of momentum grid points N_G of the HF or DFT calculation is usually too large to continue with the tensor rank decomposition of the Coulomb vertex directly. A large set of momenta will have a small but nonnegligible contribution to the correlation energy and we seek a more compact set of auxiliary field variables with fewer relevant elements.

Let

$$\tilde{\Gamma}^I_G = U^F_G \Sigma^J_F W^{*I}_J, \quad \begin{array}{l} G \\ \boxed{\tilde{\Gamma}} \\ I \end{array} = \begin{array}{l} G \\ \boxed{U} \\ \boxed{\Sigma} \\ \boxed{W^*} \\ I \end{array} \tag{8}$$

be a singular value decomposition of the Coulomb vertex $\tilde{\Gamma}^I_G$, written as a matrix with the compound index $I = (q, r)$, where the singular values in Σ are sorted in descending order. Eq. (8) is also shown in the form of a tensor wiring diagram of the involved tensor contractions on the right. Taking only the largest $N_F < N_G$ singular values of the unapproximated Coulomb vertex $\tilde{\Gamma}$ into account, we can define the *optimized auxiliary field (OAF) Coulomb vertex*

$$\Gamma^I_F := \Sigma^J_F W^{*I}_J = U^{*G}_F \tilde{\Gamma}^I_G. \quad \begin{array}{l} F \\ \boxed{\Gamma} \\ I \end{array} := \begin{array}{l} F \\ \boxed{U^*} \\ \boxed{\tilde{\Gamma}} \\ I \end{array} \tag{9}$$

Note that we write Γ without a tilde for the approximated Coulomb vertex, in contrast to usual convention, simply because we will not use the unapproximated vertex in any subsequent step.

We are only interested in the left singular vectors U^F_G associated to the largest singular values so we contract Eq. (8) from the right with $\tilde{\Gamma}^{*G'}_I$,

$$\tilde{\Gamma}^I_G \tilde{\Gamma}^{*G'}_I = U^F_G \Sigma^{2F}_F U^{*G'}_F =: E^{G'}_G, \tag{10}$$

transforming a singular value problem of a large $N_G \times N_p^2$ matrix into an eigenvalue problem of a comparatively small $N_G \times N_G$ hermitian matrix. The eigenvalues of E are the squares of the singular values of $\tilde{\Gamma}^I_G$, and the left eigenvectors of E associated to the largest eigenvalues are also the left singular vectors of $\tilde{\Gamma}$ we need in order to transform the Coulomb vertex $\tilde{\Gamma}^I_G$ into the optimized auxiliary field Coulomb vertex Γ^I_F according to Eq. (9). Note that this approach becomes numerically problematic for very small singular values since one only has access to their squares. However, we find that all N_F largest singular values needed for an accurate approximation of the Coulomb vertex are sufficiently large.

Inserting the singular value decomposition of the Coulomb vertex with sorted singular values from Eq. (8) into the discretized definition of the Coulomb integrals given in Eq. (7) yields a singular value decomposition of the Coulomb integrals, also with sorted singular values

$$V^I_J = W^K_J \Sigma^{2K}_K W^{*I}_K, \quad \begin{array}{l} J \\ \boxed{V} \\ I \end{array} = \begin{array}{l} J \\ \boxed{W} \\ \boxed{\Sigma^2} \\ \boxed{W^*} \\ I \end{array} \tag{11}$$

where the Coulomb integrals V^{pq}_{sr} are now written in the matrix form V^I_J with $I = (q, r)$ and $J = (s, p)$. Thus, using the optimized auxiliary field Coulomb vertex Γ^I_F instead of the full Coulomb vertex $\tilde{\Gamma}^I_G$ best approximates the Coulomb integrals with respect to the Frobenius norm of the difference

$$V^{pq}_{sr} \approx \Gamma^{*pF}_s \Gamma^q_{rF}. \quad \begin{array}{l} p \quad q \\ \diagup\!\sim\!\diagdown \\ s \quad r \end{array} \approx \begin{array}{l} p \quad q \\ \boxed{\Gamma^*} \quad \boxed{\Gamma} \\ s \quad r \end{array} \quad F \tag{12}$$

We stress that the method outlined above for obtaining the optimized auxiliary field is expected to be less accurate than the more widely used Cholesky decomposition method that employs an upper bound of error to any integral. However, we find that the achieved level of accuracy is sufficient for the present purpose and the level of accuracy can be systematically improved by including a larger number of singular values.

III. DECOMPOSITION OF THE COULOMB VERTEX

The form of the Coulomb vertex in real space on the right hand side in Eq. (6) suggests that the optimized Coulomb vertex Γ^q_{rF} can be decomposed in an analogous manner into a product of three tensors of second order, denoted and depicted as follows:

$$\Gamma^q_{rF} \approx \Lambda^R_F \Pi^{*q}_R \Pi^R_r. \quad \begin{array}{l} q \\ \boxed{\Gamma} \\ r \end{array} \quad F \approx \begin{array}{l} q \\ \boxed{\Pi^*} \\ R \\ \boxed{\Pi} \\ r \end{array} \quad \boxed{\Lambda} \quad F \tag{13}$$

We let N_R denote the number of *vertex indices* R and refer to it as the *rank* of the Coulomb vertex for a given quality of the approximation. We call the matrices Π^R_r and Λ^R_F *factor orbitals* and *Coulomb factors* of the Coulomb vertex, respectively.

The decomposition is invariant under scaling of the Coulomb factors Λ^R_F with any real scalar $a_R > 0$ while scaling the factor orbitals Π^R_r with a complex scalar c_R with $|c_R| = 1/\sqrt{a_R}$ for each value of R. One can also choose an alternative ansatz to Eq. (13) for approximating the Coulomb vertex Γ^q_{rF} which does not involve the conjugation of the factor orbitals on the outgoing index q. This ansatz reads

$$\Gamma^q_{rF} \approx \Lambda^R_F \Pi^R_q \Pi^R_r. \quad \begin{array}{l} q \\ \boxed{\Gamma} \\ r \end{array} \quad F \approx \begin{array}{l} q \\ \boxed{\Pi} \\ R \\ \boxed{\Pi} \\ r \end{array} \quad \boxed{\Lambda} \quad F \tag{14}$$

The above decomposition is invariant under scaling of the Coulomb factors Λ^R_F with any complex scalar $c_R \neq 0$ while

scaling the factor orbitals Π_r^R with $\pm 1/\sqrt{c_R}$ for each value of R. This is in contrast to the symmetries of the ansatz according to Eq. (13). In the Alternating Least Squares (ALS) approximation scheme, which we employ for fitting the factor orbitals Π and the Coulomb factors Λ, it is preferable to have the symmetries of Eq. (14) since fixing one of the two factors removes all continuous symmetries from the other. This accelerates the convergence and allows a smaller rank N_R in practice. A downside of this ansatz is that one loses the simple notion of particle and hole propagators as given in Eqs. (34) and (35), respectively. The propagators can, for instance, be employed to calculate second order Møller–Plesset theory (MP2) correlation energies in $\mathcal{O}(N^4)$, as outlined in Subsection IV A. Unless otherwise stated we use the ansatz of Eq. (14) for the applications presented in Section V but we will continue to discuss the more widely applicable ansatz of Eq. (13).

If one chooses $N_R = N_r$, where N_r is the number of real space grid points, the validity of the ansatz follows directly from Eq. (6). We want to investigate how low N_R can be chosen compared to N_r for a sufficiently faithful decomposition having an error below 1% in the energies calculated from the factor matrices. For reference we use the error in the MP2 energy assuming that other terms, occurring for instance in the coupled cluster amplitude equations, exhibit a similar behavior. This accuracy is assumed sufficient since in practice only those terms will be calculated from the factor matrices that pose either computational or memory bottlenecks. Furthermore, we want to show that the ratio N_R/N_r for a sufficiently faithful decomposition is independent of the system size if the system is not too small.

A. Canonical polyadic decomposition algorithms

A factorization of a tensor according to the ansatz of Eq. (13) or (14) is referred to as canonical polyadic decomposition[30] (CPD). For a given rank N_R, the factor orbitals Π_r^R and the Coulomb factors Λ_F^R can be fit by minimizing the square of the Frobenius norm of the difference

$$(\Lambda, \Pi) = \underset{\Lambda,\,\Pi}{\operatorname{argmin}} \, \|\Lambda_F^R \Pi_R^{*q} \Pi_r^R - \Gamma_{rF}^q\|^2. \tag{15}$$

The above optimization problem is high dimensional and nonquadratic. Conjugate gradient algorithms or other local algorithms may require thousands of steps until sufficiently converged. Global optimization algorithms try to tackle the problem by keeping a subset of the variables fixed and optimizing only the remaining variables. In the case of the ALS[31] algorithm, the optimization is done in turn over each matrix, while in the case of the cyclic coordinate descent[32] (CCD) algorithm, the optimization is done in turn over each value of the index R. We have studied the performance of a regularized version of the ALS here.

B. Alternating least squares

In the case of three distinct factors $T_{ijk} \approx A_{iR} B_{jR} C_{kR}$, two of them can be regarded fixed leaving a least squares problem for finding the optimal third factor. Each matrix is optimized in alternating order leading to the *alternating least squares* (ALS) algorithm

$$A^{(n+1)} := \underset{A}{\operatorname{argmin}} \left\| A_{iR} B_{jR}^{(n)} C_{kR}^{(n)} - T_{ijk} \right\|^2, \tag{16}$$

$$B^{(n+1)} := \underset{B}{\operatorname{argmin}} \left\| A_{iR}^{(n+1)} B_{jR} C_{kR}^{(n)} - T_{ijk} \right\|^2, \tag{17}$$

$$C^{(n+1)} := \underset{C}{\operatorname{argmin}} \left\| A_{iR}^{(n+1)} B_{jR}^{(n+1)} C_{kR} - T_{ijk} \right\|^2, \tag{18}$$

which has to be solved iteratively until it is sufficiently converged, starting with random matrices $A^{(0)}$, $B^{(0)}$, and $C^{(0)}$.

Each least squares problem has a unique solution, which can be written explicitly. For Eq. (16) it is for instance given by

$$A_{iR}^{(n+1)} = T_{ijk} B^{*jS} C^{*kS} G_{SR}^{+}, \tag{19}$$

omitting the iteration specification on B and C for brevity. G^+ denotes the Moore–Penrose pseudoinverse[33,34] of the Gramian matrix G. For Eq. (16) the Gramian matrix is given by

$$G_{RS} = B^{*jS} C^{*kS} B_{jR} C_{kR}\,. \tag{20}$$

The expressions for the other matrices can be written in an analogous manner. When applying the ALS algorithm to decompose the Coulomb vertex, the computationally most demanding steps are the calculation of the pseudoinverse G^+ scaling as $\mathcal{O}(N_R^3)$, as well as the contraction of T_{ijk} with either factor B^{*jS} or C^{*kS} in Eq. (19), depending on which is larger, scaling as $\mathcal{O}(N_p^2 N_F N_R)$.

We stress that the ALS algorithm and the employed CPD of the Coulomb integrals has already been studied in Ref. 11 under the guise of the tensor hypercontraction (THC) method. To overcome the slow convergence of the ALS algorithm, the authors of the THC method have switched in Refs. 12, 13, and 18 to determining the factor orbitals Π using a real space quadrature grid, which was shown to have physical justification. Having fixed the factor orbitals, the Coulomb factors Λ can then be efficiently calculated by a single least squares fit.

C. Regularized alternating least squares

In this work, we apply a modification to the ALS algorithm that can substantially improve the convergence which does not require the choice of a real space quadrature grid. Although the ALS algorithm guarantees an improvement of the fit quality in each iteration, the convergence can be very slow, especially when there are multiple local minima for a factor A in different regions having all similar minimal values. In that case, the best choice for A may vary strongly from iteration to iteration since updating the other factors B and C can change the order of the minima. This behavior is referred to as swamping[35] and it takes many iterations before the ALS algorithm converges to one region for each factor that globally minimizes the fit quality. Introducing a penalty on the distance to the previous iteration limits swamping and leads to the regularized ALS[36] (RALS) algorithm

$$A^{(n+1)} := \underset{A}{\operatorname{argmin}} \left(\left\| A_{iR} B_{jR}^{(n)} C_{kR}^{(n)} - T_{ijk} \right\|^2 + \lambda_A^{(n)} \left\| A_{iR} - A_{iR}^{(n)} \right\|^2 \right), \tag{21}$$

$$B^{(n+1)} := \underset{B}{\operatorname{argmin}}\left(\left|\left|A_{iR}^{(n+1)}B_{jR}C_{kR}^{(n)} - T_{ijk}\right|\right|^2 + \lambda_B^{(n)}\left|\left|B_{jR} - B_{jR}^{(n)}\right|\right|^2\right), \tag{22}$$

$$C^{(n+1)} := \underset{C}{\operatorname{argmin}}\left(\left|\left|A_{iR}^{(n+1)}B_{jR}^{(n+1)}C_{kR} - T_{ijk}\right|\right|^2 + \lambda_C^{(n)}\left|\left|C_{kR} - C_{kR}^{(n)}\right|\right|^2\right). \tag{23}$$

The solution of each regularized least squares problem can again be given explicitly, here, for instance, for Eq. (21), and again omitting the iteration specification on B and C,

$$A_{iR}^{(n+1)} = \left(T_{ijk}B^{*jS}C^{*kS} + \lambda_A^{(n)}A_{iS}^{(n)}\right)G_{SR}^{+}. \tag{24}$$

In the regularized case, the Gramian G depends on the regularization parameter $\lambda_A^{(n)}$,

$$G_{RS} = B^{*jS}C^{*kS}B_{jR}C_{kR} + \lambda_A^{(n)}\delta_{RS} \tag{25}$$

where δ_{RS} denotes the Kronecker delta.

The regularization parameter $\lambda_A^{(n)}$ for finding $A^{(n+1)}$ in the nth iteration still remains to be determined. Too low values allow swamping to occur while too large values unnecessarily slow down the convergence. To estimate an efficient regularization parameter we assume that the fit quality $||AB^{(n)}C^{(n)} - T||^2$ in the term to minimize in Eq. (21) varies little from one iteration to the next. We also assume this for the local change of the fit quality with respect to each value in A. This allows us to relate the minimized term in the previous step $n-1$ to the minimized term in the step n for which we want to determine the regularization parameter

$$\lambda^{(n-1)}||A^{(n)} - A^{(n-1)}||^2 \approx \lambda^{(n)}||A^{(n+1)} - A^{(n)}||^2. \tag{26}$$

If $A^{(n)}$ and $A^{(n+1)}$ have similar norm, we can also relate their relative step sizes $s_A^{(n)}$ and $s_A^{(n+1)}$, by $\lambda^{(n-1)}s_A^{2\,(n)} \approx \lambda^{(n)}s_A^{2\,(n+1)}$, where the relative step size in the nth iteration is given by

$$s_A^{(n)} := ||A^{(n)} - A^{(n-1)}||/||A^{(n)}||. \tag{27}$$

We want the relative iteration step size $s_A^{(n+1)}$ of the next iteration to be approximately as large as a chosen maximum value s_0, which we refer to as *swamping threshold*. From that we define the estimated regularization parameter for the nth iteration

$$\hat{\lambda}_A^{(n)} := \lambda_A^{(n-1)}s_A^{2\,(n)}/s_0^2. \tag{28}$$

Directly using the above estimate results in a regularization which we find alternately to be too strong and too weak. To ameliorate this, we introduce a mixing of the estimated regularization parameter $\hat{\lambda}_A^{(n)}$ for the nth iteration, as above, with the regularization parameter $\lambda_A^{(n-1)}$ of the previous iteration to obtain the regularization parameter $\lambda_A^{(n)}$ employed for the nth iteration in the RALS

$$\lambda_A^{(0)} := 1, \tag{29}$$

$$\lambda_A^{(n)} := \alpha\hat{\lambda}_A^{(n)} + (1-\alpha)\lambda_A^{(n-1)}. \tag{30}$$

Regarding the choice of the swamping threshold s_0 and the mixing factor α, we find that $s_0 = 1.0$ and $\alpha = 0.8$ offer a good compromise allowing quick convergence while still preventing swamping for the systems studied so far.

D. Quadratically occurring factors

In the case of the Coulomb vertex, the factor orbitals Π_r^R occur quadratically. For finding the next estimate $\Pi^{(n+1)}$ in the alternating least squares algorithm, we use an iterative algorithm similar to the Babylonian square root algorithm. Each subiteration is given by

$$\Pi^{(n+1,m+1)} := (1-\beta)\,\Pi^{(n+1,m)} + \beta\,\underset{\Pi}{\operatorname{argmin}}\left(\left|\left|\Lambda_F^{R\,(n+1)}\Pi^{*q\,(n+1,m)}_{R}\Pi_r^R - \Gamma_{rF}^q\right|\right|^2 + \lambda_\Pi^{(n+1,m)}\left|\left|\Pi_r^R - \Pi_r^{R(n+1,m)}\right|\right|^2\right), \tag{31}$$

with $\Pi^{(n+1,0)} := \Pi^{(n)}$ and the mixing factor $0 < \beta < 1$. Note that Π^* is a fixed parameter rather than a fitted one and that the regularization parameter $\lambda_\Pi^{(n+1,m)}$ needs to be determined for the mth subiteration similar to Eq. (30), however with $\lambda_\Pi^{(n+1,0)} = \lambda_\Pi^{(n)}$. The above iteration converges towards a solution of the quadratic problem. We use $\Pi^{(n+1,M)}$ and $\lambda_\Pi^{(n+1,M)}$ for the next estimate of $\Pi^{(n+1)}$ and $\lambda_\Pi^{(n+1)}$ in the RALS algorithm, respectively. The number of subiterations M needs to be sufficiently large, such that $\Pi^{(n+1)}$ is at least an improved solution of the entire fit problem compared to $\Pi^{(n)}$. A large number M of subiterations gives an estimate $\Pi^{(n+1)}$ that is close to the optimal choice of Π for a given $\Lambda^{(n+1)}$. However, the cost of each subiteration is similar to the cost of the fit of Λ in the RALS algorithm and as a good choice for minimizing the overall computational cost, we find $\beta = 0.8$ and $M \geq 2$, but only as large such that the solution is an improvement. We point out that there are alternative methods for solving the quadratically occurring factors.[37]

IV. APPLICATION OF THE LOW RANK FACTORIZATION

The algorithms described so far yield an approximate factorization of the Coulomb integrals of the form

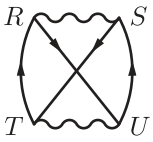
$$V_{sr}^{pq} \approx \Pi^{*p}_{R}\Pi^{*q}_{S}\Lambda^{*F}_{R}\Lambda^{S}_{F}\Pi^{S}_{r}\Pi^{R}_{s}, \tag{32}$$

where the factors Π and Λ are $N_p \times N_R$ and $N_F \times N_R$ matrices, respectively. We find that the rank of the decomposition N_R is about an order of magnitude lower than the number of real space grid points of the original factors of the Coulomb integrals, being the orbitals $\psi_q(\mathbf{x})$. In Section V we study the convergence of the approximation in detail. In this section, we discuss how this factorization can be applied to lower the scaling of the computational cost of wave function based methods such as second order Møller–Plesset (MP2) theory or coupled cluster theory.

A. MP2 from imaginary time propagators

The factorization of the Coulomb integrals permits evaluating the terms in the perturbation expansion by summing over all vertex indices $R, S, T, \ldots$ occurring in the term's diagram, contracting propagator matrices for each particle, hole, and Coulomb line. For instance, the exchange term of second order Møller–Plesset (MP2) theory can be evaluated as

follows:



$$= -\frac{1}{2}\int_0^\infty \mathrm{d}\tau \sum_{RSTU} G_T^R(\tau)G_R^U(-\tau)V_R^S \\ G_U^S(\tau)G_S^T(-\tau)V_T^U, \tag{33}$$

where the imaginary time dependent propagator matrices are given in terms of the decomposed factor orbitals Π_r^R and the Coulomb factors Λ_F^R,

$$G_S^R(\tau > 0) := +\sum_a \Pi^{*a}_S \Pi_a^R \, \mathrm{e}^{-(\varepsilon_a-\mu)\tau}, \tag{34}$$

$$G_S^R(\tau \leq 0) := -\sum_i \Pi^{*i}_S \Pi_i^R \, \mathrm{e}^{-(\varepsilon_i-\mu)\tau}, \tag{35}$$

$$V_R^S := \Lambda^{*F}_R \Lambda_F^S, \tag{36}$$

where μ is the Fermi level energy at zero temperature. The imaginary time integration can be done numerically on a small grid, for instance, the minimax grid,[38] which allows an evaluation of the MP2 correlation energy in $\mathcal{O}(N^4)$ in principle. In practice, however, this approach outperforms canonical MP2 calculations only for very large systems. The propagator matrices $G_S^R(\pm\tau)$ are analogous to the one-body particle/hole Green's functions in real space and imaginary time $G_0(\mathbf{x}, \mathbf{x}'; \pm\tau)$. The time independent matrix propagator V_R^S corresponds to the Coulomb kernel. The factorization of the one-body energies $\varepsilon_{a,i}$ in imaginary time is equivalent to the Laplace transformed MP2 ansatz of Almlöf.[39] Reducing the scaling of MP2 theory to $\mathcal{O}(N^4)$ has also been extensively discussed in Refs. 11, 16, and 40.

Note that the imaginary time dependent matrices defined above are actually not propagators in the sense that

$$G_T^S(\tau_1)G_S^R(\tau_2) = G_T^R(\tau_1 + \tau_2), \tag{37}$$

for all R, T and $\tau_1, \tau_2 > 0$ for particles as well as $\tau_1, \tau_2 \leq 0$ for holes. They can only be used to directly connect vertices of the Coulomb interaction. If propagators in the above sense are required, one needs to employ a stricter ansatz for the factorization of the Coulomb vertex, namely,

$$\Gamma^q_{rF} \approx \Lambda_F^R \Pi^{+q}_R \Pi_r^R, \tag{38}$$

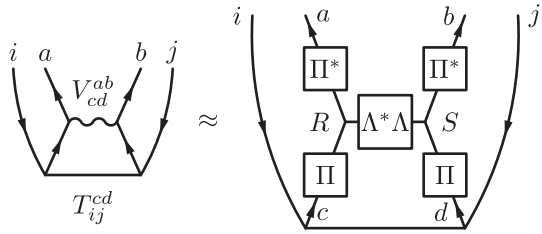
where Π^+ denotes the Moore–Penrose pseudoinverse of Π. The convergence behavior of this ansatz remains, however, to be studied.

One can also evaluate the perturbation terms stochastically, directly using the real space Green's functions $G_0(\mathbf{x}, \mathbf{x}'; \pm\tau)$ rather than using the low rank propagators $G_S^R(\pm\tau)$. This has been done for one dimensional solids[41,42] and for three dimensional solids.[43]

B. Reduced scaling coupled cluster theory

This section outlines how the obtained factorization can be used to reduce the computational cost of coupled cluster theory. We closely follow the work of Parrish *et al.* in Ref. 17. A related acceleration scheme for coupled cluster theory is the chain of spheres exchange (COSX) method as described in Ref. 44.

The most demanding step in the canonical CCD (DCD) method using a plane wave basis set is the calculation of the particle/particle ladder contribution $T_{ij}^{cd}V_{cd}^{ab}$ in the amplitude equation, scaling as $\mathcal{O}(N_\mathrm{o}^2N_\mathrm{v}^4)$ in time. Here, the factorized form of the Coulomb integrals V_{cd}^{ab} can be exploited to break down the simultaneous contraction over the indices c and d into a sequence of contractions involving only at most one index, as can be seen from the tensor wiring diagram of the involved tensors



(39)

The most expensive term in this sequence of contractions leads to a scaling of $\mathcal{O}(N_\mathrm{o}^2N_\mathrm{v}^2N_R)$ in time, without exceeding the memory complexity of the coupled cluster amplitudes. As will be demonstrated in Section V, we find N_R to be proportional to the system size N, resulting in an $\mathcal{O}(N^5)$ scaling behavior in time of the particle/particle ladder contribution. Furthermore, the DCD amplitude equations can be solely reformulated in an $\mathcal{O}(N^5)$ implementation with the use of the Coulomb vertex and its decomposed approximation, due to the absence of exchange terms between different clusters. Likewise, the most expensive term in CCSD (DCSD) amplitude equations includes the singles contribution to the ladder diagrams ($T_{ij}^{cd}T_k^aV_{cd}^{kb}$, $T_{ij}^{cd}T_l^bV_{cd}^{al}$, $T_{ij}^{cd}T_k^aT_l^bV_{cd}^{kl}$). Similarly, these terms can be evaluated via the factor orbitals and the Coulomb factors in an $\mathcal{O}(N^5)$ scaling in time and in the DCSD approximation no term exceeds this scaling behavior.

V. RESULTS EMPLOYING THE LOW RANK FACTORIZATION

A. Total energies of the LiH solid

We first seek to discuss the convergence of the low rank factorization with respect to the number of iterations, N_R and the system size. To this end we study different supercell sizes constructed from two atomic LiH crystal unit cells including $2 \times 2 \times 2$, $3 \times 2 \times 2$, $3 \times 3 \times 2$, $4 \times 3 \times 2$, and $3 \times 3 \times 3$, corresponding to 16, 24, 36, 48, and 54 atoms, respectively. In this subsection, we only employ MP2 theory to investigate the behavior of the correlation energy calculated from the factorized Coulomb integrals. This study focuses on the decomposition of the Coulomb vertex, neither employing the optimized auxiliary field nor the pseudized Gaussian type virtual orbitals[45] technique. The resulting Coulomb vertices that need to be fit are large, such that only the particle/hole part Γ_{iF}^a is used for this study. The kinetic energy cutoff defining N_F was set to 200 eV. The Li $2s^1$ and H $1s^1$ states have been treated as valence states.

Fig. 1 shows the relative error of the MP2 correlation energy retrieved as a function of the number of iterations. The

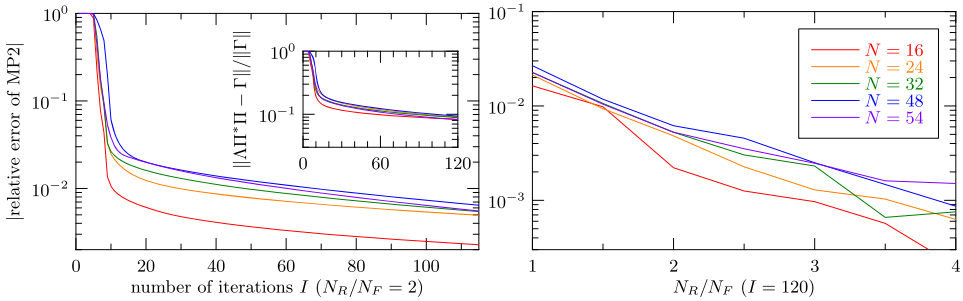


FIG. 1. On the left, the convergence of the relative error of the MP2 energy for $N_R = 2N_F$ with respect to fit iterations for different system sizes is shown. The inset plots the relative error of the Coulomb vertex approximation in comparison. The convergence with respect to the rank N_R using 120 iterations is given on the right. Both plots use the same colors denoting different system sizes.

relative error is computed from MP2 energies that employ integrals that have been calculated with and without the low rank tensor factorization. Fig. 1 reveals that the rate of convergence for the relative error of the MP2 energy is very similar in all different system sizes. We note that the smallest supercell containing 16 atoms only exhibits a slightly faster convergence. From these results, we conclude that the required number of iterations in the tensor factorization algorithm is system size independent for intensive properties. Furthermore we find that 100 iterations are sufficient to achieve a relative accuracy of 1% with a rank that corresponds to $2N_F$. Although the accuracy of the fit is still improving when going beyond 100 iterations, we find it more economic to increase the rank N_R rather than increasing the number of iterations if a higher accuracy is required.

The right side of Fig. 1 explores the convergence of the relative error in the MP2 correlation energy error retrieved as a function of N_R/N_F. Note that N_F corresponds to the number of plane wave vectors and scales linearly with respect to the system size. This plot shows that the required rank N_R needed to achieve a certain relative level of accuracy also scales linearly with respect to the system size. Furthermore we find that systematically improvable exponential convergence can be achieved for this system and property by increasing N_R/N_F.

The computational cost for obtaining the TRD of the Coulomb vertex with $N_\mathrm{o} = 27$, $N_\mathrm{v} = 8469$, $N_F = N_G = 1830$, and $N_R = 1830$ is roughly 1000 CPU hours. Therefore the computational cost of the TRD exceeds the computational cost of a full MP2 calculation which is roughly 10 CPU hours in the present case despite the fact that the TRD formally scales more favorably with system size. However, we note that the present TRD algorithm is sufficiently efficient to reduce the total computational cost of coupled cluster theory calculations as discussed in the following.

B. Molecular adsorption of water on hexagonal boron nitride

We now turn to the application of the newly developed methodologies to some more challenging problems. We calculate the interaction between a water molecule and a hexagonal boron nitride (*h*BN) monolayer. We employ periodic CCD theory and examine to what extent the TRD and the optimized auxiliary field approximations are accurate and efficient. We used the structures obtained by Al-Hamdani *et al.*,[46] whereby the molecule is oriented on top of an N site and the geometry has been optimized using the optB86b-vdW functional. The water–N distance was set to 3.2 Å. The *h*BN monolayer is modeled by 32 atoms in the periodic cell and the distance between two BN sheets was set to 16 Å. After checking convergence, we employed a 500 eV kinetic energy cutoff for the one particle orbitals along with Γ point sampling of the Brillouin zone. The B $2s^2 2p^1$, N $2s^2 2p^3$, O $2s^2 2p^4$, and H $1s^1$ states have been treated as valence states. Occupied HF states were converged within the full plane wave basis. The vast number of virtual orbitals in a plane wave basis is circumvented by mapping a virtual orbital manifold, expanded in plane waves, onto atom centered basis functions. Dunning's contracted aug-cc-pVDZ (AVDZ) and aug-cc-pVTZ (AVTZ)[47,48] pseudized Gaussians were chosen to represent the atom centered functions. By orthogonalizing these virtual states to the occupied space, we mimic an AVDZ and AVTZ basis set via orbitals which are expanded in plane waves. The underlying procedure is detailed in Ref. 45. This allows the usage of well tested basis sets with reliable extrapolation schemes. Counterpoise corrections to the basis set superposition error (BSSE) were included in all correlated calculations. The adsorption energy is defined as the difference in energy between the noninteracting fragments and the interacting system

$$E_\mathrm{ads} = E_{\mathrm{H_2O}} + E_\mathrm{BN} - E_{\mathrm{H_2O+BN}}. \tag{40}$$

We note that in Ref. 46, the adsorption energy has been calculated as the difference between the total energy of water and *h*BN at the largest possible oxygen–surface distance of 8 Å and the total energy of water and *h*BN at the adsorption oxygen–surface distance.

Initially, we investigate the convergence of the adsorption energy with respect to the number of momentum grid points N_G, employed to evaluate the Coulomb vertex $\tilde{\Gamma}^q_{rG}$ according to Eq. (7). The selection of the plane waves vectors $\mathbf{G}$ is determined by a kinetic energy cutoff E_χ such that

$$\frac{\hbar^2 \mathbf{G}^2}{2m_e} < E_\chi. \tag{41}$$

For this purpose, we utilize the pseudized AVDZ basis set for the virtual orbitals. The current system consists of $N_\mathrm{o} = 68$ and $N_\mathrm{v} = 780$ occupied and virtual orbitals, respectively. Kinetic energy cutoff values from 100 to 300 eV were employed for the calculation of the adsorption energy. The results are shown in the inset of Fig. 2. The adsorption energy behaves as $CE_\chi^{-3/2}$ [49,50] up to 200 eV, however, at higher cutoffs one

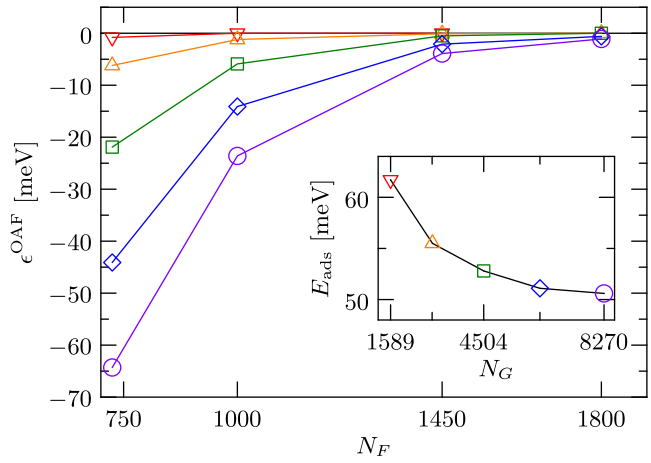


FIG. 2. Optimized auxiliary field (OAF) approximation error $\epsilon^{\mathrm{OAF}}(N_F) = E_{\mathrm{ads}}(N_F) - E_{\mathrm{ads}}(N_G)$ of the CCD adsorption energy as a function of the number of field variables N_F used to approximate the Coulomb vertex. The number of G vectors N_G of the unapproximated Coulomb vertex $\tilde{\Gamma}^q_{rG}$ is indicated by the shape of the markers. The inset shows which marker corresponds to which N_G and plots the convergence of the adsorption energy with respect to N_G, corresponding to kinetic energy cutoff values of 100, 150, 200, 250, and 300 eV, respectively.

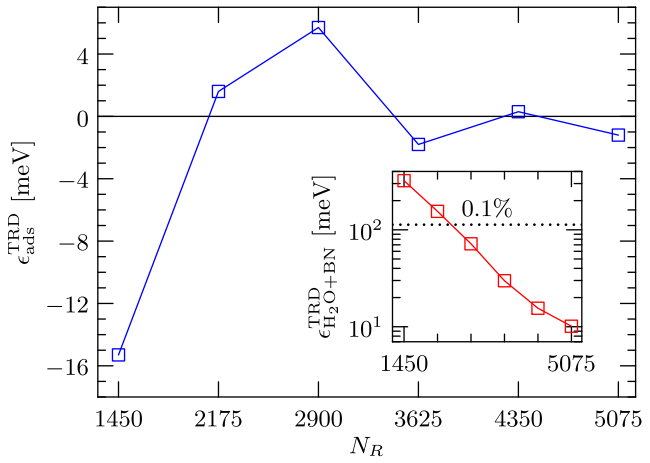


FIG. 3. Low rank approximation error $\epsilon^{\mathrm{TRD}}_{\mathrm{ads}}(N_R) = E_{\mathrm{ads}}(N_F, N_R) - E_{\mathrm{ads}}(N_F)$ of the CCD adsorption energy as a function of the rank N_R using the OAF approximated Coulomb vertex with $N_F = 1450$ field variables. The inset shows the respective approximation error for surface + molecule fragment $\epsilon^{\mathrm{TRD}}_{\mathrm{H_2O+BN}} = E_{\mathrm{H_2O+BN}}(N_F, N_R) - E_{\mathrm{H_2O+BN}}(N_F)$ only, revealing an error cancellation of one to two orders of magnitude in the adsorption energy.

observes a plateau in the curve as a result of the truncation of the virtual orbital space via the pseudized Gaussian basis functions. We conclude that a cutoff energy of 200 eV is sufficient to converge the adsorption energy to within 2–3 meV. We then investigate how accurately can the optimized auxiliary field approximate the Coulomb vertex $\tilde{\Gamma}^q_{rG}$ for the current system at the level of CCD theory. First we obtain the optimized Coulomb vertex $\Gamma^q_{rF} = U^{*G}_F \tilde{\Gamma}^q_{rG}$, where U^F_G consists of the left singular values of $\tilde{\Gamma}^q_{rG}$ associated to the N_F largest singular values, according to Eq. (8). The optimized Coulomb vertex is expected to be efficient since most of the space in the simulation cell is vacant, and the plane wave auxiliary basis contains redundant information. Different numbers of field variables N_F were employed to approximate the plane wave vectors N_G for the various cutoff energies. The behavior of the adsorption energy with respect to the number of field variables is shown in Fig. 2. The rapid convergence of the energy with increasing number of field variables owes to the locality of the molecular orbitals in the supercell. The adsorption energy obtained with a cutoff of 200 eV can be calculated within 0.5 meV accuracy using $N_F = 1450$ field variables to approximate $N_G = 4504$ plane wave vectors.

Since we can conclude that the adsorption energy can be computed within approximately 3 meV using a cutoff energy of 200 eV for the auxiliary plane wave basis and $N_F = 1450$ field variables to construct the optimized Coulomb vertex, we chose these settings to assess the accuracy of the low rank factorization of the Coulomb integrals. We used different numbers of vertex indices R to compute the factor orbitals and Coulomb factors of the Coulomb vertex, following Eq. (14). We approximate the particle/particle ladder contribution $T^{cd}_{ij} V^{ab}_{cd}$ in the amplitude equation of CCD via the factor orbitals and the Coulomb factors as shown in Eq. (39). The adsorption energy versus the number of the vertex indices N_R is shown in Fig. 3. The energy does not converge monotonically as in the case of the optimized auxiliary field approximation. This is due to the nonlinear nature of the canonical polyadic decomposition of the Coulomb vertex and the random initial choice for its factors. Nevertheless we observe a converged behavior with increasing N_R. Furthermore, a value of $N_R = 4350 = 3N_F$ is sufficient to yield an adsorption energy within 1 meV accuracy. This suggests that the TRD of the Coulomb vertex is a controllable approximation that can yield increasingly accurate results with increasing decomposition rank N_R. In order to further validate the accuracy of the TRD method, we show the convergence of the absolute energy of the interacting system with respect to the decomposition rank N_R in the inset of Fig. 3. We observe an exponential convergence of the total energy. An accuracy better than 0.1% is achieved already with $N_R = 2N_F$. Nevertheless we stress that the corresponding accuracy in the adsorption energy is a result of an error cancellation of one to two orders of magnitude.

Having assessed the accuracy of the TRD, we now calculate the adsorption energy of the water molecule on *h*BN using the AVTZ pseudized Gaussian basis set. The evaluation involves the decomposition of a Coulomb vertex with $N_o = 68$, $N_v = 1564$, $N_F = 0.33 N_G = 1450$, and $N_R = 3N_F = 4350$. The computational cost to obtain the decomposed matrices is roughly 3000 CPU hours with 256 iterations. The results of the adsorption energy are shown in Table I. In order to grasp a physical insight of the system, we compare the CCD results with RPA + SOSEX and MP2 calculations.[51] MP2 theory usually overestimates dispersion driven

TABLE I. Adsorption energies of water on *h*BN obtained using the pseudized Gaussian basis sets at different levels of theory. RPA + SOSEX calculations were performed using DFT PBE orbitals as reference, whereas MP2 and CCD using HF ones.

Basis set	RPA + SOSEX	MP2	CCD
AVDZ	62	83	54
AVTZ	72	92	62
AV(D,T)Z	76	95	65

interactions although in the description of BN bilayer interaction it is fortuitously accurate.[52] Consequently, one expects MP2 theory to slightly overestimate the adsorption energy, whereas RPA + SOSEX is likely to yield a very accurate estimate. It is not surprising that CCD underbinds the water molecule, since there exist findings that indicate the inability of CCD for an accurate description. Higher levels of theories, such as inclusion of the single excitations and the perturbative triples, are required for a more appropriate treatment. Nevertheless, the purpose of the current work is to examine the accuracy and efficiency of the newly developed methodologies rather than the accuracy of the method itself. The CPU hours required for the CCD calculations obtained with and without the TRD technique are summarized in Table II. The time for the evaluation of the particle/particle ladder term per iteration is as much as 43 times faster using a decomposition with $N_R = 2N_F$ and 22 times faster with $N_R = 3N_F$. This constitutes a significant gain in the computational effort of coupled cluster methods with only slight compromise in accuracy.

C. Potential energy surface smoothness

Finally, we examine the potential energy surface (PES) obtained from the TRD method. In order to assess the smoothness of the potential energy landscape, we examine the adsorption energy curve of a water molecule on a periodic *h*BN monolayer. A supercell consisting of an 8 atom *h*BN substrate was employed for the calculation of the CCD energy curve. The orientation of the water molecule on top of the surface was kept the same as in Sec. V B, while the atomic positions of the monolayer were fixed to their pristine structure. The energy curve was obtained by moving the molecule vertically on top of the surface without altering its orientation. We follow a similar computational procedure as in Sec. V B. HF oribtals were obtained using a 600 eV kinetic energy cutoff along with Γ point sampling of the first Brillouin zone. The virtual states were expanded in plane waves and were chosen in a way to resemble closely an AVDZ basis set orthogonalized to the HF occupied orbitals, in the same manner as in Sec. V B. Since we are solely interested in the smoothness of the PES and not the absolute adsorption energies, this small basis set is sufficient. The selection of the plane wave vectors $\mathbf{G}$ for the evaluation of the Coulomb vertex $\tilde{\Gamma}^q_{rG}$ according to Eq. (7) was determined by a kinetic energy cutoff $E_\chi = 300$ eV. The resulting number of momentum grid points is $N_G = 2066$. The optimized Coulomb vertex $\Gamma^q_{rF} = U^{*G}_F \tilde{\Gamma}^q_{rG}$ was obtained using the $N_F = 826$ largest singular values of $\tilde{\Gamma}^q_{rG}$ according to Eq. (8). The adsorption energy agrees to within 1 μeV using the optimized Coulomb vertex.

TABLE II. CPU hours per iteration comparing CCD calculation with and without the factorized Coulomb integrals. In parenthesis we denote the part for evaluating the particle/particle ladder term.

Basis set	$N_R = 2N_F$	$N_R = 3N_F$	No TRD
AVDZ	39 (13)	49 (24)	100 (75)
AVTZ	259 (28)	258 (55)	1443[a] (1212)

[a]Estimation based on the AVDZ basis set.

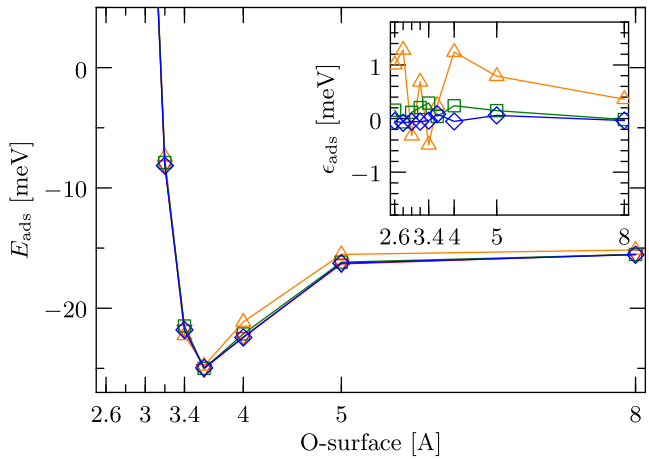


FIG. 4. Smoothness of the potential energy landscape employing the tensor rank decomposition. Triangles (orange) represent the energy curve obtained with $N_R = N_F$ vertex indices, squares (green) the energy curve with $N_R = 2N_F$, and diamonds (blue) the energy curve with $N_R = 3N_F$. The potential energy landscape of the optimized Coulomb vertex is represented by the mostly occluded red triangles pointing downwards. The inset shows the nonparallelity errors.

We approximate the particle/particle ladder contribution $T^{cd}_{ij} V^{ab}_{cd}$ in the amplitude equation of CCD via the factor orbitals and the Coulomb factors as shown in Eq. (39). The adsorption energy curve obtained using different numbers of decomposed vertex indices N_R and 128 iterations is shown in Fig. 4. On the scale of the absolute adsorption energy, only the $N_R = N_F$ curve deviates from the optimized Coulomb vertex result. The PES obtained using $N_R = 2N_F$ and $N_R = 3N_F$ does not exhibit a deviation visible on the scale of the absolute adsorption energy. In order to examine further the accuracy of the PES landscape of the TRD method we show on the inset of Fig. 4 the relative error with respect to the energy curve of the optimized Coulomb vertex. The error of the $N_R = N_F$ energy curve is of the order of 1 meV and oscillates rather randomly. The $N_R = 2N_F$ curve exhibits a much smaller error (of the order of 0.1 meV) and the $N_R = 3N_F$ an even smaller one. This demonstrates that sub meV accuracy and very smooth potential energy surfaces can be achieved in a systematic manner for increasing N_R.

VI. CONCLUSIONS

In this work, we have outlined an algorithm to obtain a low rank tensor approximation of the Coulomb integrals having the same algebraic structure as its definition from the molecular orbitals

$$V^{pq}_{sr} = \iint_{\mathrm{dx}\,\mathrm{dx}'} \psi^{*p}(\mathbf{x})\, \psi^{*q}(\mathbf{x}')\, \frac{1}{|\mathbf{r}-\mathbf{r}'|}\, \psi_r(\mathbf{x}')\, \psi_s(\mathbf{x}) \approx \sum_{RS} \Pi^{*p}_R \Pi^{*q}_S \Lambda^{*F}_R \Lambda^S_F \Pi^S_r \Pi^R_s . \tag{42}$$

The developed methods for periodic systems are based on previous work referred to as tensor hypercontraction (THC) outlined in Refs. 11–13. The factorization is obtained by fitting $\Lambda^S_F \Pi^{*q}_S \Pi^S_r$ to auxiliary three index quantities referred to as Coulomb vertices that are calculated from a resolution of identity approach using a plane wave basis set. In this

manner, the scaling of the computational cost for obtaining the low rank tensor approximation with respect to system size does not exceed $\mathcal{O}(N^4)$. To reduce the prefactor of the computational cost further we have outlined an approach to further compactify the representation of the Coulomb vertices. We linearly transform the momentum index of the Coulomb vertices into a (truncated) basis referred to as an optimized auxiliary field. The accuracy of this truncation is systematically improvable using a single parameter that is used for the truncation of a singular value decomposition. The tensor factorization of the transformed Coulomb vertices is achieved using a regularized alternating least squares algorithm that converges rapidly using about 10^2 iterations only. In contrast, the non-regularized alternating least squares algorithm would require 10^5–10^6 iterations. We stress that we employ no prior assumptions for the real space grids used for expanding the low order tensors.

Once obtained, the tensor factorization of the Coulomb integrals can be employed to reduce the scaling of the computational cost of distinguishable coupled cluster theory to $\mathcal{O}(N^5)$ without further approximations. We demonstrate that the factorization can also be used to reduce the computational cost for evaluating the computationally most expensive term (particle/particle ladder diagram) in the CCD amplitude equations for the case of water adsorption on the *h*BN monolayer system. For system sizes containing 136 electrons in 1632 orbitals, we achieve substantial reductions in the computational cost that are on the order of a factor 10–20 without compromising the accuracy and introducing any further approximation.

Future work will focus on combining the outlined techniques with explicitly correlated methods and finite size corrections in order to significantly expand the scope of periodic coupled cluster theory calculations using plane wave basis sets for solid state systems.[53,54]

ACKNOWLEDGMENTS

We want to thank Edgar Solomonik and Alexander Auer for fruitful discussions. We also want to acknowledge the Cyclops Tensor Framework (CTF)[55] for providing automatic and efficient parallelization of tensor operations.

[1] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory* (Dover Publications, Mineola, NY, 1996).
[2] S. Reine, T. Helgaker, and R. Lindh, "Multi-electron integrals," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **2**, 290 (2012).
[3] A. Hellweg, C. Hättig, S. Höfener, and W. Klopper, "Optimized accurate auxiliary basis sets for RI-MP2 and RI-CC2 calculations for the atoms Rb to Rn," Theor. Chem. Acc. **117**(4), 587–597 (2007).
[4] R. Ahlrichs, "Efficient evaluation of three-center two-electron integrals over Gaussian functions," Phys. Chem. Chem. Phys. **6**(22), 5119–5121 (2004).
[5] F. Weigend and R. Ahlrichs, "Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy," Phys. Chem. Chem. Phys. **7**(18), 3297–3305 (2005).
[6] F. R. Manby and P. J. Knowles, "Poisson equation in the Kohn-Sham Coulomb problem," Phys. Rev. Lett. **87**, 163001 (2001).
[7] A. Grüneis, M. Marsman, J. Harl, L. Schimka, and G. Kresse, "Making the random phase approximation to electronic correlation accurate," J. Chem. Phys. **131**(15), 154115 (2009).
[8] D. Kats, "The distinguishable cluster approach from a screened Coulomb formalism," J. Chem. Phys. **144**, 044102 (2016).
[9] D. Kats and F. R. Manby, "Communication: The distinguishable cluster approximation," J. Chem. Phys. **139**, 021102 (2013).
[10] D. Kats, D. Kreplin, H.-J. Werner, and F. R. Manby, "Accurate thermochemistry from explicitly correlated distinguishable cluster approximation," J. Chem. Phys. **142**, 064111 (2015).
[11] E. G. Hohenstein, R. M. Parrish, and T. J. Martínez, "Tensor hypercontraction density fitting. I. Quartic scaling second- and third-order Møller-Plesset perturbation theory," J. Chem. Phys. **137**, 044103 (2012).
[12] R. M. Parrish, E. G. Hohenstein, T. J. Martínez, and C. D. Sherrill, "Tensor hypercontraction. II. Least-squares renormalization," J. Chem. Phys. **137**, 224106 (2012).
[13] E. G. Hohenstein, R. M. Parrish, C. D. Sherrill, and T. J. Martínez, "Communication: Tensor hypercontraction. III. Least-squares tensor hypercontraction for the determination of correlated wavefunctions," J. Chem. Phys. **137**, 221101 (2012).
[14] N. Shenvi, H. V. Aggelen, Y. Yang, W. Yang, C. Schwerdtfeger, and D. Mazziotti, "The tensor hypercontracted parametric reduced density matrix algorithm: Coupled-cluster accuracy with O(r^4) scaling," J. Chem. Phys. **139**, 054110 (2013).
[15] U. Benedikt, K.-H. Böhm, and A. A. Auer, "Tensor decomposition in post-Hartree–Fock methods. II. CCD implementation," J. Chem. Phys. **139**, 224101 (2013).
[16] S. I. L. Kokkila Schumacher, E. G. Hohenstein, R. M. Parrish, L.-P. Wang, and T. J. Martínez, "Tensor hypercontraction second-order Møller–Plesset perturbation theory: Grid optimization and reaction energies," J. Chem. Theory Comput. **11**, 3042–3052 (2015).
[17] R. M. Parrish, C. D. Sherrill, E. G. Hohenstein, S. I. L. Kokkila, and T. J. Martínez, "Communication: Acceleration of coupled cluster singles and doubles via orbital-weighted least-squares tensor hypercontraction," J. Chem. Phys. **140**(18), 181102 (2014).
[18] R. M. Parrish, E. G. Hohenstein, T. J. Martínez, and C. D. Sherrill, "Discrete variable representation in electronic structure theory: Quadrature grids for least-squares tensor hypercontraction," J. Chem. Phys. **138**, 194107 (2013).
[19] N. H. F. Beebe and J. Linderberg, "Simplifications in the generation and transformation of two-electron integrals in molecular calculations," Int. J. Quantum Chem. **12**, 683–705 (1977).
[20] I. Røeggen and E. Wisløff-Nilssen, "On the Beebe-Linderberg two-electron integral approximation," Chem. Phys. Lett. **132**, 154–160 (1986).
[21] H. Koch, A. Sánchez de Merás, and T. B. Pedersen, "Reduced scaling in electronic structure calculations using Cholesky decompositions," J. Chem. Phys. **118**, 9481–9484 (2003).
[22] B. I. Dunlap, J. W. D. Connolly, and J. R. Sabin, "On first-row diatomic molecules and local density models," J. Chem. Phys. **71**(12), 4993 (1979).
[23] M. Feyereisen, G. Fitzgerald, and A. Komornicki, "Use of approximate integrals in *ab initio* theory. An application in MP2 energy calculations," Chem. Phys. Lett. **208**, 359–363 (1993).
[24] O. Vahtras, J. Almlöf, and M. W. Feyereisen, "Integral approximations for LCAO-SCF calculations," Chem. Phys. Lett. **213**, 514–518 (1993).
[25] J. Harl, "The linear response function in density functional theory," Ph.D. thesis, University of Vienna, Vienna, 2008.
[26] P. E. Blöchl, "Projector augmented-wave method," Phys. Rev. B **50**, 17953–17979 (1994).
[27] G. Kresse and J. Hafner, "Norm-conserving and ultrasoft pseudopotentials for first-row and transition elements," J. Phys.: Condens. Matter **6**(40), 8245 (1994).
[28] G. Kresse and J. Furthmüller, "Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set," Phys. Rev. B **54**, 11169–11186 (1996).
[29] R. Blankenbecler, D. J. Scalapino, and R. L. Sugar, "Monte Carlo calculations of coupled boson-fermion systems, I," Phys. Rev. D **24**, 2278–2286 (1981).
[30] F. L. Hitchcock, "The expression of a tensor or a polyadic as a sum of products," J. Math. Phys. **6**(1), 164–189 (1927).
[31] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," SIAM Rev. **51**, 455–500 (2009).
[32] L. Karlsson, D. Kressner, and A. Uschmajew, "Parallel algorithms for tensor completion in the CP format," Parallel Comput. **57**, 222–234 (2015).
[33] E. H. Moore, "On the reciprocal of the general algebraic matrix," Bull. Am. Math. Soc. **26**(9), 394–395 (1920).
[34] R. Penrose and J. A. Todd, "A generalized inverse for matrices," Math. Proc. Cambridge Philos. Soc. **51**, 406 (1955).

[35]S. Kindermann and C. Navasca, "New Algorithms for tensor decomposition based on a reduced functional," Num. Linear Algebra Appl. **21**, 340–374 (2014).
[36]A. Cichocki and R. Zdunek, "Regularized alternating least squares algorithms for non-negative matrix/tensor factorization," *Advances in Neural Networks–ISNN* (Springer, 2007), pp. 793–802.
[37]N. Li, S. Kindermann, and C. Navasca, "Some convergence results on the regularized alternating least-squares method for tensor decomposition," Linear Algebra Appl. **438**(2), 796–812 (2013).
[38]M. Kaltak, J. Klimeš, and G. Kresse, "Low scaling algorithms for the random phase approximation: Imaginary time and Laplace transformations," J. Chem. Theory Comput. **10**, 2498–2507 (2014).
[39]J. Almlöf, "Elimination of energy denominators in Møller–Plesset perturbation theory by a Laplace transform approach," Chem. Phys. Lett. **181**, 319–320 (1991).
[40]T. Schäfer, B. Ramberger, and G. Kresse, "Quartic scaling MP2 for solids: A highly parallelized algorithm in the plane-wave basis," e-print arXiv:1611.06797 (2016).
[41]S. Y. Willow, K. S. Kim, and S. Hirata, "Stochastic evaluation of second-order many-body perturbation energies," J. Chem. Phys. **137**, 204122 (2012).
[42]D. Neuhauser, R. Baer, and D. Zgid, "Stochastic self-consistent Green's function second-order perturbation theory (sGF2)," e-print arXiv:1603.04141 (2016).
[43]T. Schäfer and G. Kresse, private communication (2016).
[44]A. K. Dutta, F. Neese, and R. Izsák, "Speeding up equation of motion coupled cluster theory with the chain of spheres approximation," J. Chem. Phys. **144**(3), 034102 (2016).
[45]G. H. Booth, T. Tsatsoulis, G. K.-L. Chan, and A. Grüneis, "From plane waves to local Gaussians for the simulation of correlated periodic systems," J. Chem. Phys. **145**, 084111 (2016).
[46]Y. S. Al-Hamdani, M. Ma, D. Alfè, O. A. V. Lilienfeld, and A. Michaelides, "Communication: Water on hexagonal boron nitride from diffusion Monte Carlo," J. Chem. Phys. **142**, 181101 (2015).
[47]T. H. Dunning, Jr., "Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen," J. Chem. Phys. **90**, 1007–1023, (1989).
[48]D. Feller, "The role of databases in support of computational chemistry calculations," J. Comput. Chem. **17**, 1571–1586 (1996).
[49]J. Harl and G. Kresse, "Cohesive energy curves for noble gas solids calculated by adiabatic connection fluctuation-dissipation theory," Phys. Rev. B **77**, 045136 (2008).
[50]M. Marsman, A. Grüneis, J. Paier, and G. Kresse, "Second-order Møller–Plesset perturbation theory applied to extended systems. I. Within the projector-augmented-wave formalism using a plane wave basis set," J. Chem. Phys. **130**, 184103 (2009).
[51]J. Paier, X. Ren, P. Rinke, G. E. Scuseria, A. Grüneis, G. Kresse, and M. Scheffler, "Assessment of correlation energies based on the random-phase approximation," New J. Phys. **14**(4), 043002 (2012).
[52]F. Hummel, T. Gruber, and A. Grüneis, "A many-electron perturbation theory study of the hexagonal boron nitride bilayer system*," Eur. Phys. J. B **89**(11), 235 (2016).
[53]A. Grüneis, "Efficient explicitly correlated many-electron perturbation theory for solids: Application to the Schottky defect in MgO," Phys. Rev. Lett. **115**, 66402 (2015).
[54]K. Liao and A. Grüneis, "Communication: Finite size correction in periodic coupled cluster theory calculations of solids," J. Chem. Phys. **145**, 141102 (2016).
[55]E. Solomonik, D. Matthews, J. R. Hammond, J. F. Stanton, and J. Demmel, "A massively parallel tensor contraction framework for coupled-cluster computations," J. Parallel Distrib. Comput. **74**(12), 3176–3190 (2014).