

Externally Corrected CCSD with Renormalized Perturbative Triples (R-ecCCSD(T)) and the Density Matrix Renormalization Group and Selected Configuration Interaction External Sources

Seunghoon Lee,* Huanchen Zhai, Sandeep Sharma,* C. J. Umrigar,* and Garnet Kin-Lic Chan*



Cite This: *J. Chem. Theory Comput.* 2021, 17, 3414–3425



Read Online

ACCESS |



Metrics & More

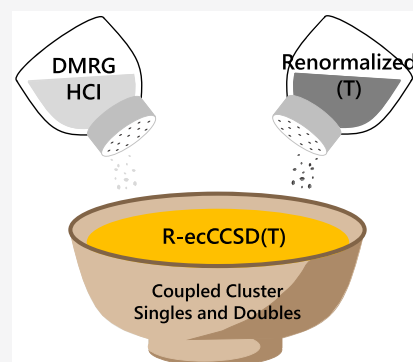


Article Recommendations



Supporting Information

ABSTRACT: We investigate the renormalized perturbative triples correction together with the externally corrected coupled-cluster singles and doubles (ecCCSD) method. We use the density matrix renormalization group (DMRG) and heat-bath CI (HCI) as external sources for the ecCCSD equations. The accuracy is assessed for the potential energy surfaces of H_2O , N_2 , and F_2 . We find that the triples correction significantly improves upon ecCCSD, and we do not see any instability of the renormalized triples with respect to dissociation. We explore how to balance the cost of computing the external source amplitudes against the accuracy of the subsequent CC calculation. In this context, we find that very approximate wave functions (and their large amplitudes) serve as an efficient and accurate external source. Finally, we characterize the domain of correlation treatable using the ecCCSD and renormalized triples combination studied in this work via a well-known wave function diagnostic.



1. INTRODUCTION

Electronic structure methods that can efficiently handle both static and dynamic correlation remain an important area of investigation. Because there is a wide spectrum of strongly correlated problems, ranging from mildly “quasi”-degenerate scenarios (e.g., in the electronic structure of diradicals^{1,2}) to extensively near-degenerate problems (e.g., in the electronic structure of multicenter transition metal clusters^{3–5}), a variety of theoretical strategies have been proposed.

For problems with more than a few nearly degenerate orbitals, it is common to employ methods that combine an explicit multireference (MR) state with dynamic correlation methods. The multireference state can range from an exact complete active space (CAS) representation^{6–8} (for small numbers of nearly degenerate orbitals), to a density matrix renormalization group (DMRG)^{9–23} which can be used even when there are many degenerate orbitals, and to selected configuration interaction^{24–44} and Monte Carlo approximations^{45–50} for intermediate cases. On top of these, various flavors of perturbation theory,^{51–63} configuration interaction,^{64–67} and exponential approximations^{59,68–70} have been explored for including dynamic correlation. However, the combination of dynamic correlation with multireference representations is not straightforward and usually leads to added conceptual, implementation, and computational complexity.

For problems with a few degenerate orbitals (say <6), an alternative strategy has been investigated, which adds some static correlation on top of an existing single-reference (SR) method. This has been particularly popular in conjunction with

SR coupled cluster methods.^{71–76} Some examples include variants of tailored coupled cluster^{77–82} and externally corrected coupled cluster methods.^{83–86} The simple formulation of SR coupled cluster methods compared to MR coupled cluster methods makes such methods attractive. However, their accuracy for problems with more than a few degenerate orbitals remains to be examined.

In this work, we will investigate the externally corrected coupled cluster method.^{83–86} This extracts static correlation from a MR method by using the MR wave function as an “external” source of higher order coupled cluster amplitudes. For example, in the externally corrected coupled-cluster singles and doubles (ecCCSD) approximation, the T_3 and T_4 amplitudes are extracted from the external source, and a new set of T_1 and T_2 amplitudes is computed in their presence. Should the T_3 and T_4 amplitudes be exact, then the T_1 and T_2 amplitudes and the energy will be exact. A different but similar approximation is tailored CCSD,^{77–79} which has been of renewed interest of late.^{80–82} Here, instead of higher order cluster amplitudes, the large (active space) T_1 and T_2 amplitudes are fixed from an external source.

Received: March 1, 2021

Published: May 21, 2021



The ecCCSD method has a long history, and external sources, ranging from unrestricted Hartree–Fock^{87–90} to CASSCF and CASCI^{91–95} and, most recently, full configuration interaction quantum Monte Carlo⁹⁶ and selected configuration interaction,⁹⁷ have been used. One of the early and more successful applications of ecCCSD is the reduced multireference (RMR) CCSD method,^{98,99} which uses a multireference configuration interaction method including single and double excitations (MR-CISD) wave function as the external source. RMR-CCSD(T), which incorporates some of the residual dynamic correlation through perturbative triples, has also been studied.¹⁰⁰ Despite the promising performance of RMR-CCSD(T) in several studies,^{101–109} it suffers from two practical limitations. First, conventional MR-CISD can only be applied for modest sizes of active spaces (typically, up to about 16 orbitals as limited by the exact CAS treatment). Second, the (T) correction, although not divergent like its single-reference counterpart, still overcorrects the dynamical correlation in the bond-stretched region.¹¹⁰

In this work, we make two modifications to ecCCSD to overcome or ameliorate the above limitations. First, we utilize variational DMRG and heat-bath configuration interaction (HCI) wave functions as external sources for ecCCSD. (The HCI method is a selected CI method. A detailed description of the method and its semistochastic perturbatively corrected extension, SHCI, can be found in ref 44.) This allows for two qualitative improvements in the types of external source to be investigated: large active space wave functions (i.e., with a large number of near-degenerate orbitals) and external sources which include a significant fraction of the dynamic correlation effects already. With respect to the former, in polymetallic cofactors, it has been possible to compute DMRG wave functions in an active space with 40 transition metal *d* orbitals and approximately 40 bridging ligand orbitals.⁵ With respect to the latter, in calculations on molecules drawn from the G2 set, almost 500 nondegenerate orbitals have been used in SHCI.¹¹¹ The second modification we make is to explore the renormalized perturbative triples correction. This has been shown in the single reference setting to ameliorate the overcorrection of standard perturbative triples,¹¹² without affecting the computational scaling.

The plan for the rest of the paper is as follows. We describe the use of DMRG and HCI as external sources for ecCCSD in Section 2.1. It is possible to create a near-exact method by using a near-exact external source. Since near-exact external sources can be computed by DMRG and HCI for the small systems we employ as test cases in this paper, the critical question is not simply the accuracy of the method, but whether accurate energies can be obtained at a reduced computational cost using intentionally far from exact sources. To this end, we explore a variety of approximate external DMRG and HCI sources as discussed in Section 2.2. The various triples approximations for ecCCSD are discussed in Section 2.3. Computational details are provided in Section 3. The accuracy of the renormalized perturbative triples correction is assessed for three potential energy surfaces (PESs) in Sections 4.1–4.4. We characterize the range of quasi-degenerate correlations captured in this work in Section 4.6. Finally we discuss the limitations of this method in Section 4.7 and summarize our findings in Section 5.

2. THEORY

In this work, we use a reference determinant $|0\rangle$ with the same occupancy as the Hartree–Fock (HF) determinant. We will denote excited determinants by $|m_k\rangle$ where k is an integer labeling the excitation level, e.g., $|m_1\rangle$ denotes the m -th singly excited determinant. It is also useful to define a projection operator onto the space of k -tuply excited determinants relative to the reference; we denote this $Q_k = \sum_m |m_k\rangle\langle m_k|$. The external source is used to provide an important subset of the triples and quadruples determinants; the projector onto this subset is denoted Q_k^{ec} , and its complement is Q_k^{c} . Thus

$$Q_k = Q_k^{\text{ec}} + Q_k^{\text{c}}, \quad k = 3, 4 \quad (1)$$

2.1. Externally Corrected CCSD with DMRG and HCI Wave Functions. Configuration interaction wave functions have the form

$$|\Psi\rangle = (1 + C)|0\rangle \quad (2)$$

whereas CC wave functions have the form

$$|\Psi\rangle = e^T|0\rangle \quad (3)$$

where $C = \sum_{n=1}^M C_n$, $T = \sum_{n=1}^M T_n$, and C_n and T_n are operators that generate n -body excitations from the reference determinant. In ecCCSD, the coupled cluster operator T is given by

$$T = T_1 + T_2 + T_3^{\text{ec}} + T_4^{\text{ec}} \quad (4)$$

where T_3^{ec} and T_4^{ec} respectively excite the ground-state determinant into the space of triples and quadruples determinants extracted from the external source (and which define the projectors Q_3^{ec} and Q_4^{ec}). In this work, we extract T_k^{ec} , $k = 3, 4$ from the DMRG or the HCI variational wave function. For the DMRG wave function, the triply and quadruply excited determinants m_k that define Q_k^{ec} , $k = 3, 4$ are chosen to be those where the magnitude of the CI coefficient c_{m_k} is above a threshold, i.e.

$$|c_{m_k}| > s\sqrt{\omega} \quad (5)$$

where s is an arbitrary scaling factor, and ω is the largest discarded weight of the density matrix at the maximum bond dimension in two-dot DMRG sweeps (carried out without noise). An efficient algorithm to convert a matrix-product state to CI coefficients above a given threshold is described in the Appendix. For the HCI variational wave function, m_k is included in the projectors Q_k^{ec} , $k = 3, 4$ using the HCI algorithm with a threshold ϵ_1 ,⁴² i.e., it is included if

$$|\langle m_k | H | m' \rangle c_{m'}| > \epsilon_1 \quad (6)$$

for at least one determinant m' which is already in the variational space. The CI coefficients are converted into cluster amplitudes using the relations

$$T_3 = C_3 - C_1 C_2 + \frac{C_1^3}{3} \quad (7)$$

$$T_4 = C_4 - C_1 C_3 - \frac{C_2^2}{2} + C_1^2 C_2 - \frac{C_1^4}{4} \quad (8)$$

which follow from eqs 2 and 3, except that if a CI coefficient is zero, we set the corresponding cluster amplitude to zero. We arrived at this prescription empirically, motivated by the idea that in situations where CC works, the magnitude of the CI coefficient in the exact wave function will be correlated with

the magnitude of the corresponding T amplitude. (In a recent paper,⁹⁷ the same choice is made by arguing that there are classes of CI wave functions for which the ecCC approach provides no improvement to the CI energy. This can be seen as follows: assume all C_1 and C_2 amplitudes are included in the CI wave function. Then, $\langle m_k | H(1 + C) | 0 \rangle = E_{\text{CI}} \langle m_k | (1 + C) | 0 \rangle$, for $k = 0, 1, 2$, is satisfied by the CI solution for any choice of C_3 and C_4 in the CI wave function. If all T_3 and T_4 amplitudes generated by eq 8 are used in the external source, then $e^T | 0 \rangle = (1 + C) | 0 \rangle$, up to quadruply excited determinants, and $\langle m_k | H e^T | 0 \rangle = E_{\text{CI}} \langle m_k | e^T | 0 \rangle$, for $k = 0, 1, 2$. Then, we can show that such a wave function satisfies the CCSD solution condition, $\langle m_k | e^{-T} H e^T | 0 \rangle = 0$ for $k = 1, 2$. This is because for $k = 1, 2$

$$\begin{aligned} \langle m_k | e^{-T} H e^T | 0 \rangle &= \sum_{m'} \langle m_k | e^{-T} | m' \rangle \langle m' | H e^T | 0 \rangle \\ &= \sum_{m'; k' \in \{0,1,2\}} \langle m_k | e^{-T} | m' \rangle \langle m' | H e^T | 0 \rangle \\ &= \sum_{m'} \langle m_k | e^{-T} | m' \rangle \langle m' | E_{\text{CI}} e^T | 0 \rangle = E_{\text{CI}} \langle m_k | 0 \rangle = 0 \end{aligned} \quad (9)$$

where the middle line follows because $\langle m_k | e^{-T} | m' \rangle$ is nonzero only if m' is a determinant of lower or equal excitation to m_k . Thus, ecCC in this case offers no change in the energy. However, while true, this does not itself justify omitting T_3 and T_4 amplitudes where the corresponding C_3 and C_4 coefficients are zero, as can be seen by constructing an example where the C_3 and C_4 coefficients are zero in the exact solution.)

The T_1 and T_2 amplitudes are obtained by solving the ecCCSD equations using fixed T_3^{ec} and T_4^{ec}

$$0 = \langle m_k | e^{-T} H_N e^T | 0 \rangle; \quad k = 1, 2 \quad (10)$$

where H_N is the Hamiltonian in normal-ordered form, and T is defined in eq 4. With the relaxed T_1 and T_2 , the ecCCSD correlation energy of the ground state is obtained as

$$\Delta E_0^{\text{ecCCSD}} = \langle 0 | e^{-T} H_N e^T | 0 \rangle \quad (11)$$

2.2. Approximations in the External Source. While DMRG and HCI can, in small molecules, be a source of nearly exact T_3 and T_4 amplitudes even in the full orbital space, this provides no computational advantage as such calculations are more expensive than the subsequent coupled cluster calculation. Consequently, it is important to find approximate inexpensive sources that lead to acceptable errors in the final coupled cluster calculation.

In this work, we consider six different types of approximate external sources with different sizes of active spaces and different values of parameters summarized in Table 1. Type-I uses CASSCF-like external sources in minimal active spaces. Since the minimal active space of F_2 , i.e., two electrons in two orbitals (2e,2o), does not contain T_3 and T_4 , we perform minimal active space ecCC calculations for only H_2O and N_2 (with (4e,4o) and (6e,6o), respectively). These provide amplitudes that are very close to the exact CASSCF amplitudes. However, these amplitudes lack the relaxation that comes from allowing excitations within a larger space of orbitals, and of course, the amplitudes outside the minimal active space are completely absent. Type-III uses CASCI-like external sources (the orbitals are not optimized to save computer time) in larger active spaces ((8e,18o), (10e,16o), and (14e,16o) for H_2O , N_2 , and F_2 , respectively).

In either case, one can potentially introduce bad external amplitudes if the effect of relaxation on the amplitude upon going to a larger space is large relative to the size of the

Table 1. Six Types of Approximate External Sources^d

type	method	active space	cut-off parameters	
			wave function	cluster amplitude
I	DMRG	minimal ^a	$M = 2000$	100% with $s = 0.1$
	HCI		$\epsilon_1 = 10^{-7}, 10^{-15}$	100%
II	DMRG	minimal ^a	$M = 2000$	80% with $s = 0.1$
	HCI		$\epsilon_1 = 10^{-7}, 10^{-15}$	80%
III	DMRG	larger ^b	$M = 2000$	100% with $s = 0.1$
IV	DMRG	larger ^b	$M = 2000$	80% with $s = 0.1$
V	DMRG	larger ^b	$M = 25, 50, 100$	80% for H_2O and N_2
				100% for F_2 with $s = 0.01$
VI	HCI	full ^c	$\epsilon_1 = 0.01, 0.003$	100%

^a(4e, 4o) for H_2O , (6e, 6o) for N_2 . ^b(8e, 18o) for H_2O , (10e, 16o) for N_2 , (14e, 16o) for F_2 . ^c(10e, 58o) for H_2O (14e, 60o) for N_2 , (14e, 58o) for F_2 . ^dA detailed explanation is to be found in the main text.

amplitude (e.g., changes its sign). Thus, we study also Type-II and Type-IV external sources which are similar to Type-I and Type-III external sources, respectively, except that they employ an additional threshold to screen out all except the largest T_3 and T_4 amplitudes. The absolute values of the T_3 and T_4 elements at the most stretched geometry of each molecule are sorted in a single large vector. The norm of the vector is computed. Only the largest elements of the vector are retained such that the resulting norm is more than 80% of the norm of the full vector. Along the PESs of each molecule, we used the same elements of T_3 and T_4 (but with the appropriate values for each geometry) as the external sources, to maintain the smoothness of the PESs.

As discussed in Section 4.4, the Type-III and Type-IV sources improve upon the PESs obtained from the Type-I and Type-II sources, but the DMRG calculations to obtain the sources incur a higher computational cost than the subsequent CC calculations. To reduce the cost, we have also tried Type-V sources, which employ loosely converged DMRG wave functions with small bond dimensions of $M = 25, 50$, and 100. Finally, Type-VI sources employ large thresholds $\epsilon_1 = 0.01$ and 0.003 to obtain loosely converged HCI wave functions in the full orbital spaces. This combination has the advantage that it can be considered a black box method wherein a single parameter ϵ_1 controls the trade-off between accuracy and cost.

2.3. Perturbative Triples Corrections. Expressions for the standard, renormalized, and completely renormalized perturbative triples corrections can be written down in analogy with their single reference definitions.¹¹² We first define the completely renormalized (CR)-ecCCSD(T) correction. We use the state $|\Psi^{\text{ecCCSD}(T)}\rangle$ defined as

$$|\Psi^{\text{ecCCSD}(T)}\rangle = (1 + T_1 + T_2 + T_3^{\text{ec}} + Q_3^c T_3^{[2]} + Q_3^c Z_3) | 0 \rangle \quad (12)$$

$$T_3^{[2]} | 0 \rangle = R_0^{(3)} (V_N T_2)_C | 0 \rangle \quad (13)$$

$$Z_3 | 0 \rangle = R_0^{(3)} V_N T_1 | 0 \rangle \quad (14)$$

where V_N is the two-body part of the Hamiltonian in normal-ordered form, and $R_0^{(3)}$ denotes the three-body component of the reduced resolvent operator in many-body perturbation theory, given by differences of orbital energies in the denominator.⁷⁶ In addition, $(V_N T_2)_C$ denotes the connected

part of $V_N T_2$. The resulting formula for the CR-ecCCSD(T) energy correction is

$$\delta_0^{\text{CR-ecCCSD(T)}} = \frac{\langle \Psi^{\text{ecCCSD(T)}} | Q_3^c e^{-T} H_N e^T | 0 \rangle}{\langle \Psi^{\text{ecCCSD(T)}} | e^T | 0 \rangle} \quad (15)$$

The energy corrections for renormalized R-ecCCSD(T) and ecCCSD(T) can be obtained by taking the lowest-order estimates of the correction and by assuming the denominator to be one, i.e.

$$\delta_0^{\text{R-ecCCSD(T)}} = \frac{\langle \Psi^{\text{ecCCSD(T)}} | Q_3^c (V_N T_2)_C | 0 \rangle}{\langle \Psi^{\text{ecCCSD(T)}} | e^T | 0 \rangle} \quad (16)$$

$$\delta_0^{\text{ecCCSD(T)}} = \langle \Psi^{\text{ecCCSD(T)}} | Q_3^c (V_N T_2)_C | 0 \rangle \quad (17)$$

Unlike perturbative triples without external correction (as in CCSD(T)), the approximate perturbative expression $T_3^{[2]}$ is only evaluated for determinants omitted in the external source. Thus, it is not expected to diverge as long as the external source includes all degeneracies. Nonetheless, it can still overestimate the triples correlation. The role of the denominator in the “renormalized” triples approximations is to rescale this correction, which is expected to reduce the overestimation.

3. COMPUTATIONAL DETAILS

All CC calculations were performed using cc-pVTZ basis sets.¹¹³ All excitations are permitted for H₂O and N₂, whereas excitations from the HF 1s core are not considered for F₂, resulting in 58, 60, and 58 orbitals for the three systems, respectively. We used spin-restricted orbitals in all the calculations. Type-I and -II source calculations were done using CASSCF natural orbitals, and Type-III and -IV source calculations were done with HF orbitals.

The errors in the ecCC PESs were evaluated by comparing against highly accurate PESs obtained from the semistochastic HCI (SHCI) method with a sufficiently small value of ϵ_1 (4×10^{-5}) to ensure good convergence of the total energy. The SHCI total energy is obtained using a semistochastic method to evaluate a second-order perturbation theory correction to the variational HCI energy.

All CC calculations were carried out using a local version of PySCF¹¹⁴ interfaced with StackBLOCK^{115–118} for DMRG and Arrow^{44,119,120} for HCI. We used Dice^{119–121} to get accurate SHCI PESs.

4. NUMERICAL RESULTS

4.1. PESs with Type-I External Sources. The dissociation PESs of H₂O and N₂, shown in Figure 1, were obtained by the ecCC methods using the Type-I external sources in Table 1 (i.e., near exact wave functions in the minimal active spaces and all amplitudes of T_3 and T_4). We show the PESs of ecCC using DMRG external sources (colored solid lines) and the PESs of CC (colored dotted lines) in Figure 1. These are compared against accurate PESs represented as black lines obtained by SHCI in the full space. All the ecCC energies studied in this work, the accurate energies from SHCI, the source energies, and an example of timing for ecCC calculations are given in the Supporting Information. The mean absolute errors (MAEs) and the nonparallelity errors (NPEs) of the PESs are listed in Table 2.

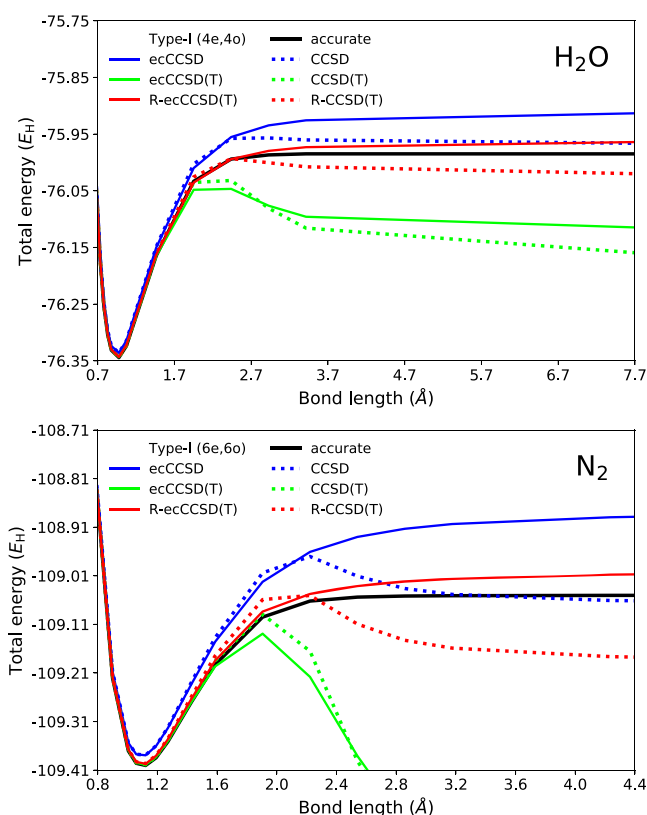


Figure 1. PESs of H₂O and N₂ in the top and bottom panels, respectively, obtained with the CC methods and the ecCC methods using the Type-I external sources. The Type-I sources correspond to near exact wave functions from the minimal active space and use all the amplitudes of T_3 and T_4 . The blue, green, and red solid lines are the ecCCSD, ecCCSD(T), and R-ecCCSD(T) PESs, respectively. These are to be compared with the PESs of CC represented as dotted lines. The black lines are accurate PESs obtained by SHCI with a sufficiently small value of ϵ_1 .

The CCSD curves (blue dotted lines) have an unphysical dip due to an inadequate treatment of static correlation. This problem is completely eliminated within ecCCSD (blue solid line). However, there are significant errors with respect to the accurate (black) curve at large distances, giving an (MAE,NPE) of (23.9, 65.8) mE_H for H₂O and (63.1, 146.8) mE_H for N₂.

The (T) correction captures much of the missing dynamic correlation, such that the green ecCCSD(T) curves reach an accuracy of 1 mE_H around the equilibrium geometry. However, when the bond is stretched, the (T) correction overestimates the dynamic correlation, leading to another unphysical dip in the PESs. Although this overestimation can easily be reduced by increasing the size of the active space for the small systems treated here, this would be very expensive for large systems. Alternatively, we can use the renormalized triples formula to damp the (T) correction in eq 16. R-ecCCSD(T) (the solid red curves) completely removes the unphysical dips in the PESs. These attain an (MAE,NPE) = (3.8, 23.3) and (11.1, 47.1) mE_H for H₂O and N₂, respectively.

As expected, the ecCC PESs with tightly converged HCI external sources are almost identical to those with the DMRG sources and are therefore not shown in Figure 1 but can be found in the Supporting Information. However, for systems where the molecule has a low-spin ground state, but the

Table 2. MAE and NPE (mE_H) of H_2O , N_2 , and F_2 PESs in a Range of Geometries $R \in [0.68, 7.80]$ Å, $R \in [0.79, 4.23]$ Å, and $R \in [1.14, 5.00]$ Å, Respectively^b

method	type	H_2O			N_2			F_2		
		active	MAE	NPE	active	MAE	NPE	active	MAE	NPE
DMRG/HCI	I, II	4e,4o	215.5	59.8	6e,6o	274.3	32.1	2e,2o	<i>a</i>	<i>a</i>
DMRG	III, IV	8e,18o	160.9	52.5	10e,16o	224.7	59.2	14e,16o	334.9	27.7
CCSD			15.9	30.4		32.5	102.5		34.8	58.0
ecCCSD	I	4e,4o	23.9	65.8	6e,6o	63.1	146.8	2e,2o	<i>a</i>	<i>a</i>
	II		24.0	65.8		57.6	114.4			
	III	8e,18o	17.6	49.9	10e,16o	48.4	112.9	14e,16o	25.3	41.7
	IV		15.6	39.4		41.5	74.0		27.0	40.8
R-CCSD(T)			6.8	43.0		32.1	162.3		6.2	10.6
R-ecCCSD(T)	I	4e,4o	3.8	23.3	6e,6o	11.1	47.1	2e,2o	<i>a</i>	<i>a</i>
	II		3.9	23.1		5.5	13.8			
	III	8e,18o	2.9	18.0	10e,16o	12.2	44.2	14e,16o	0.9	1.6
	IV		1.1	8.1		4.7	8.7		1.7	1.5

^aNo T_3 and T_4 in the minimal active space of F_2 . ^bThe second column shows the types of the external sources defined in Table 1.

dissociated fragments have high-spin ground states, the HCI wave functions at stretched geometries can be spin-contaminated when ϵ_1 is not enough to be small, even when time-reversal symmetry is employed to reduce the spin contamination, and this can cause the energy to dip down. Our DMRG calculations employ full spin symmetry and therefore avoid this problem. Full spin symmetry could also be employed in HCI, but that is not the case in this paper. From Table S2 of the Supporting Information, we see that the R-ecCCSD(T) energies with HCI sources are the same as those with DMRG sources to better than 10^{-4} – 10^{-3} Hartree when the expectation value of the square of spin operator ($\langle S^2 \rangle$) with the HCI wave function is less than 10^{-7} – 10^{-6} (the exact ground states are singlets). Using ϵ_1 of 10^{-7} , $\langle S^2 \rangle < 10^{-7}$ is achieved up to $3R_e$ (2.9 Å) for H_2O and up to 3.3 Bohr (1.7 Å) for N_2 , and using ϵ_1 of 10^{-15} , $\langle S^2 \rangle < 10^{-7}$ is achieved up to 5.8 Bohr (3.1 Å) for N_2 .

4.2. PESs with Type-II External Sources. The difference between the Type-I and Type-II external sources is that the former use all the amplitudes of T_3 and T_4 , while the latter use only a small number of the largest amplitudes, which contribute about 80% of the total weight of T_3 and T_4 (see Table 1 and Section 2.2). For the minimal active space of H_2O , i.e. (4e,4o), the external source contains only one nonzero T_4 amplitude, and all elements of T_3 are zero to within numerical noise. Thus, the PESs of ecCC using 80% of the external amplitudes (Type-II) are almost identical to those using 100% of the external amplitudes (Type-I). On the other hand, for the minimal active space of N_2 , i.e. (6e,6o), the external source contains several large T_3 and T_4 amplitudes. At the stretched geometry corresponding to a bond length of 4.23 Å, 80% of the total T_3 and T_4 amplitude weight is recovered by the nine largest elements of T_4 . Figure 2 shows that the Type-II external sources, although using fewer amplitudes, improve the PESs of ecCCSD and R-ecCCSD(T) relative to the Type-I sources. The PES of R-ecCCSD(T) displays an (MAE,NPE) = (5.5, 13.8) mE_H .

One concern with partial use of the amplitudes in the minimal active space is that those amplitudes that are not taken from the external source are calculated perturbatively and may diverge because of zeros in the energy denominator. Using the HF occupation of the CASSCF natural orbitals, the recomputed orbital energies (diagonal parts of the Fock matrix) retain a sizable gap between the occupied and

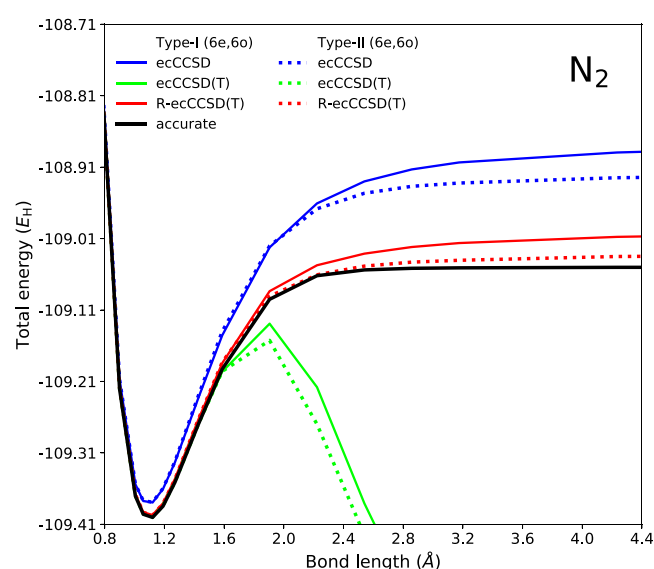


Figure 2. PESs of N_2 obtained by the ecCC methods using all the external amplitudes (Type-I) and some of the largest external amplitudes (Type-II) from the minimal active space (solid and dotted lines, respectively). Other descriptions are the same as in Figure 1.

unoccupied orbitals. For example, for N_2 at the largest bond length studied (4.23 Å), the gap between the three occupied orbitals and the three unoccupied orbitals, which become degenerate in the infinite distance limit, is about 0.1 hartree. Hence, we see from Figure 2 that Type-II sources overestimate the ecCCSD(T) correlation only a little more than Type-I sources do. Note that the R-ecCCSD(T) curve using Type-II sources is considerably more accurate than the one using Type-I sources.

Similar to what is seen with the Type-I source, the ecCC PESs with Type-II HCI sources are almost identical to those from DMRG sources when $\langle S^2 \rangle < 10^{-7}$ in the HCI source. Using a truncation threshold of $\epsilon_1 = 10^{-7}$ for H_2O and 10^{-15} for N_2 is sufficient to create an accurate Type-II HCI source over the same range of geometries as discussed for the Type-I source. (See Table S3 of the Supporting Information.)

4.3. PESs with Type-III and Type-IV External Sources. For DMRG sources, we investigated PESs of R-ecCCSD(T) using larger active spaces with all amplitudes (Type-III) and

80% of the amplitudes (Type-IV) to further reduce the errors. We used active spaces of (8e,18o), (10e,16o), and (14e,16o) for H_2O , N_2 , and F_2 , respectively, and obtained near exact wave functions in these spaces. The PESs of ecCC using the Type-III and Type-IV external sources are shown as colored solid and dotted lines, respectively, in Figure 3. The MAE and NPE of the PESs are given in Table 2.

For H_2O and N_2 , the resulting PESs of R-ecCCSD(T) with all amplitudes (red solid lines in the top and middle panels) achieve an $(\text{MAE}, \text{NPE}) = (2.9, 18.0)$ and $(12.2, 44.2)$ mE_H , respectively. These are minor improvements compared to the results using the minimal active space external sources.

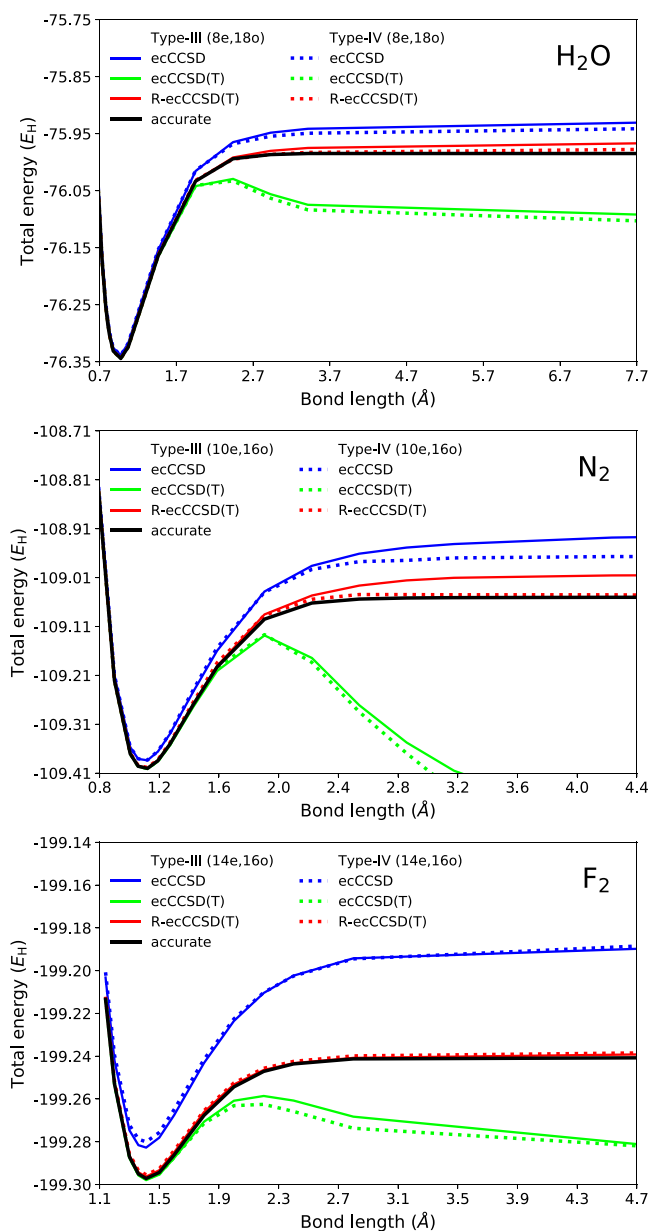


Figure 3. PESs of H_2O , N_2 , and F_2 in the top, middle, and bottom panels, respectively, for the ecCC methods using the larger than minimal active spaces and the near exact external sources. The solid and dotted lines correspond to PESs obtained using all external amplitudes of T_3 and T_4 (Type-III) and only the largest amplitudes of T_4 (Type-IV), respectively. The descriptions are otherwise the same as those in Figure 2.

However, the PESs obtained using only 80% of the external amplitudes (red dotted lines) are much closer to the black accurate PES and reach an $(\text{MAE}, \text{NPE}) = (1.1, 8.1)$ and $(4.7, 8.7)$ mE_H . Similar to the minimal active space, 80% of the amplitude weight in the larger active space corresponds to only one element of T_4 for H_2O and the nine largest elements of T_4 for N_2 . When we have a few large elements, the partial use of the amplitudes has advantages in both accuracy and efficiency. Although the same elements of T_4 are used in Type-II and Type-IV sources, the extra relaxation of their values coming from the larger CAS space of Type-IV sources contributes to their accuracy.

Unlike in H_2O and N_2 , there are no particularly large elements in the T_3 and T_4 amplitudes of F_2 . At the bond length of 5.0 Å, one requires approximately the 400 largest elements of T_3 and T_4 for their total weights to be 80% of the total T_3 and T_4 weights. The red solid and dotted lines in the bottom panel of Figure 3 show the PESs of R-ecCCSD(T) using all and 80% of the amplitudes, respectively. These two PESs are very close to the accurate black curve and reach an $(\text{MAE}, \text{NPE}) = (0.9, 1.6)$ and $(1.7, 1.5)$ mE_H for 100% and 80% of the amplitudes, respectively. Although both are accurate, the partial use of the external source in this case slightly degrades the accuracy.

4.4. PESs with Type-V External Sources. In the previous section, we showed that the use of larger active spaces significantly improves the ecCC PESs. However, obtaining tightly converged DMRG wave functions in large active spaces requires more CPU time than the subsequent CC calculation. (See Table S9 of the Supporting Information.) For example, optimizing an external wave function with 16 orbitals requires around a few minutes of CPU time for one DMRG sweep with $M = 2000$. Although a few minutes is not prohibitively large, it is large compared to the subsequent ecCC calculation which only takes tens of seconds for the small molecules considered here. In addition, the fact that in some cases using only the large amplitudes led to better results in the last section suggests that it is not a good use of computational time to tightly converge the external source.

Hence, we consider loosely converged DMRG sources (Type-V) in the larger active spaces. We obtained R-ecCCSD(T) PESs using DMRG sources with bond dimensions 25, 50, and 100. We used the truncation to 80% amplitude weight for H_2O and N_2 , while we used all the external amplitudes for F_2 . Table 3 shows the corresponding MAE and NPE in the same range of geometries in Table 2.

For all cases, when we reduced the bond dimension to $M = 100$, the MAE increased by 0.3–1.2 mE_H , and the NPE increased by 0.0–0.8 mE_H , compared to using $M = 2000$. However, for $M = 100$, one DMRG sweep with 16 orbitals took only a few seconds of CPU time at the 4.23 Å bond length of N_2 , giving a better computational balance between the DMRG calculation and the subsequent ecCC calculations. When we further reduced the bond dimension to $M = 25$, the MAE increased by 2.5 and 0.9 mE_H , and the NPE increased by 0.6 and 2.4 mE_H for H_2O and N_2 . In the case of F_2 , which does not have a small number of large T_3 and T_4 elements, the MAE and NPE increased more, by 3.2 and 5.2 mE_H .

We have so far shown that the errors of R-ecCCSD(T) can be reduced by increasing the size of active space and replacing the small magnitude cluster amplitudes by those computed from perturbation theory. Further, we find only a small degradation in accuracy upon going from a DMRG bond

Table 3. MAE and NPE (mE_H) of H_2O , N_2 , and F_2 PESs in a Range of Geometries $R \in [0.68, 7.80]$ Å, $R \in [0.79, 4.23]$ Å, and $R \in [1.14, 5.00]$ Å, Respectively, for Type-V Sources Defined in Table 1

M	H_2O			N_2			F_2		
	active	MAE	NPE	active	MAE	NPE	active	MAE	NPE
25	(8e,18o)	3.6	8.7	(10e,16o)	5.6	11.1	(14e,16o)	4.1	6.8
50		2.8	8.2		5.2	9.6		3.5	3.4
100		2.3	8.1		5.0	9.5		1.3	2.3
2000		1.1	8.1		4.7	8.7		0.9	1.6

dimension of 2000 to 100. However, achieving both accuracy and efficiency requires tuning several parameters—the size of the active space, the screening parameter for the amplitudes, and the bond dimension in DMRG. We next consider a source that can be used in a more black box manner.

4.5. PESs with Type-VI External Sources. Type-VI sources use the full orbital space, avoiding the complications of choosing an active space. To make such sources black box, we choose to specify only a single parameter to control the accuracy of the source. For example, for a DMRG wave function, this could be just the bond dimension. Here, we use HCI for the Type-VI source and control the accuracy with the single HCI parameter, ϵ_1 , to select the large T_3 and T_4 amplitudes. It is normal practice to improve upon the variational HCI energy using a semistochastic implementation of second-order perturbation theory.^{44,119} We refer to the resulting energies as SHCI total energies. Hence, an important question is whether the R-ecCCSD(T) energies are more accurate than the SHCI total energies. In Figure 4, we make this comparison for two rather large values of ϵ_1 , namely, 0.01 and 0.003. Also shown in the figure are the essentially exact energies obtained from SHCI using a small value of ϵ_1 (4×10^{-5}) (black lines) and the energies of the sources (red lines) so that the improvement coming from R-ecCCSD(T) and the perturbative correction of SHCI can be seen. It is apparent that both R-ecCCSD(T) and the perturbative correction of SHCI capture much of the missing energy in the source, particularly at short and medium bond lengths. At long bond lengths, the R-ecCCSD(T) dips down because the source wave functions are spin contaminated, which is why the energies in Figure 4 are plotted over a smaller range of geometries than in the other figures. Since N_2 is a singlet that dissociates into atoms that are quartets, it has particularly large errors. The spin contamination and the dip in the energy of N_2 are considerably reduced upon going from $\epsilon_1 = 0.01$ to $\epsilon_1 = 0.003$. A better solution would be to impose S^2 symmetry on the source wave function. For H_2O , the R-ecCCSD(T) PESs, for both values of ϵ_1 , are more accurate than the corresponding SHCI energies. For F_2 , the R-ecCCSD(T) PES for $\epsilon_1 = 0.01$ is more accurate than the SHCI PES, but for $\epsilon_1 = 0.003$, they are both equally good. From the limited number of calculations presented in this paper, it appears that R-ecCCSD(T) is better than SHCI for large ϵ_1 but not for small ϵ_1 . Hence, it may be a useful method for large systems where SHCI calculations with small ϵ_1 require large computer resources (mostly memory rather than time).

Recently a similar study used a different selected CI method (CIPSI) as the source and studied the performance of the method on the H_2O molecule using a smaller basis than the one we used (cc-pVDZ instead of cc-pVTZ). Their ecCC method differs from our R-ecCCSD(T) in detail but has in common that it does not use T_3 or T_4 amplitudes from the source if the corresponding C_3 or C_4 coefficients are zero and

that it adds in a perturbative triples correction for amplitudes that are not set by the source. They too found that, for small numbers of determinants in the source, their ec-CC-II₃ method outperforms the selected CI plus perturbation theory but that, for large numbers of determinants, it underperforms, despite the fact that they impose S^2 symmetry on their CIPSI source wave functions.

4.6. Error Analysis. In this section, we present the errors of R-ecCCSD(T) for the systems in this work and analyze them using the well-known CC error diagnostic D_2 , defined by the matrix 2-norm of the T_2 amplitudes.¹²² The magnitude of D_2 can be used to distinguish between the SR and MR character of the different geometries on the PESs. Organic molecules are sometimes considered to have MR character when D_2 is larger than 0.18.¹²² Figure 5 shows the absolute errors on a log scale versus the D_2 diagnostic. Each symbol represents a geometry on the PESs of one of the molecules, H_2O , N_2 , and F_2 . Square symbols denote R-CCSD(T), and circle symbols denote the various externally corrected theories, using the Type-V and the Type-VI external sources. For D_2 ranging from 0 to 1.7, the absolute errors of R-ecCCSD(T) (red circles) are less than 0.015 E_H , and most of the errors are smaller than those of R-CCSD(T) (black squares) and ecCCSD (blue circles). The absolute errors of R-ecCCSD(T) are less than those of ecCCSD(T) (green circles) in the MR region where $D_2 > 0.4$, while they are mostly greater than those of ecCCSD(T) in the range $0.0 < D_2 < 0.4$. Overall, it is clear to see that R-ecCCSD(T) offers the most balanced treatment of errors across a wide range of SR and MR character. However, for very weakly correlated systems, the original (T) correction is slightly more accurate than the renormalized (T) correction.

4.7. Limitations. Although the work above shows that it is possible to obtain quantitative accuracy across the full potential energy surface, at a reasonable computational cost, using the combination of external sources and the renormalized (T) correction, SR ecCC approaches have a fundamental limitation when the CI coefficient of the reference configuration (the HF configuration in this work) in the external source is small. The smallest reference coefficient value we encountered in this work was 0.2 at the stretched (4.23 Å) geometry of N_2 , and it is difficult to converge the ecCCSD energy, so we followed the following procedure. We extracted initial amplitudes of T_1 and T_2 from the external source and then iteratively converged the ecCCSD energy using a damping parameter of 0.05 to update the amplitudes. (We did not use the direct inversion in the iterative subspace (DIIS) algorithm.) At this geometry, the energy could be converged monotonically up to a threshold of 10^{-5} Hartrees, although the norm of the amplitudes could not be converged, and we simply used amplitudes from the last iteration in the set of monotonically decreasing energies.

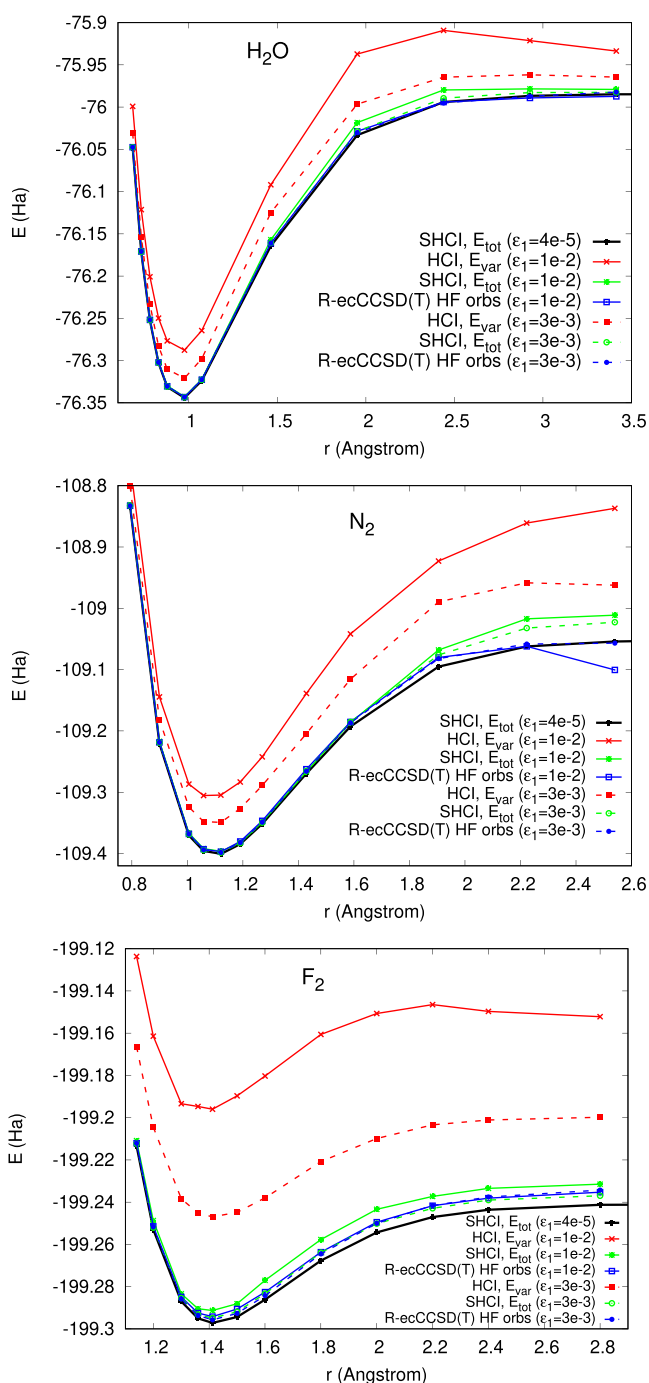


Figure 4. PESs of H_2O , N_2 , and F_2 in the top, middle, and bottom panels, respectively, for the R-ecCCSD(T) method (blue) using the full orbital space and large values (0.01 (solid) and 0.003 (dashed)) of ϵ_1 in HCI (Type-VI sources). Also shown are the source HCI variational energies (red), the SHCI total energies (green), and the essentially exact SHCI energy using $\epsilon_1 = 4 \times 10^{-5}$ (black).

5. CONCLUSION

In this work, we explored the externally corrected coupled cluster with a renormalized triples correction method (R-ecCCSD(T)), using DMRG and HCI and external sources. The critical question is how to best balance the accuracy and cost of computing the external source with the cost of the overall method. To this end, we considered multiple types of external sources: “exact” external sources, where the DMRG

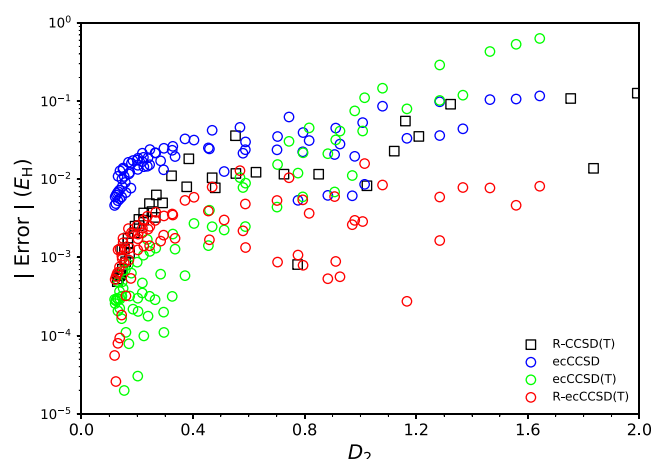


Figure 5. Absolute errors of R-CCSD(T) and ecCC methods on a log scale plotted against the D_2 diagnostic. A detailed explanation can be found in the main text.

and HCI wave functions were tightly converged within small active spaces, and “approximate” external sources, where they were loosely converged within larger active spaces. We also considered both “full” usage of the T_3 and T_4 amplitudes and “partial” usage, wherein we retained only the largest elements.

For all systems considered here, we found that R-ecCCSD(T) can significantly improve the potential energy surface obtained from either the external source alone or CC alone. For example, the unphysical dips in the PES obtained from CC methods in the bond-stretched region are eliminated. The use of approximate external sources, possibly with truncation to only the large T_3 and T_4 amplitudes, appears to be a practical way to balance the cost of the external calculation and the coupled cluster calculation in small molecules. Using the D_2 diagnostic to characterize the different points on the potential energy surfaces, we find that R-ecCCSD(T) gives absolute errors of less than 15 mE_H in the range of D_2 from 0.0 to 1.7. In fact, the errors of R-ecCCSD(T) are less than those of ecCCSD and R-CCSD(T) in almost all cases, except when D_2 is very small, where the renormalized (T) correction appears to be slightly less accurate than the simple (T) correction.

There are several interesting questions remaining which lie beyond what we have considered in this work. For example, while R-ecCCSD(T) appears quite stable up to large values of the D_2 diagnostic, what is the largest amount of multireference character which can be handled? Here, the difficulty in solving the CC equations, and the divergence of the amplitudes reflecting the problems of intermediate normalization, cannot be ignored. In addition, in the realm of quasidegenerate problems, we can ask whether other noniterative corrections such as the “completely renormalized” triples and quadruples corrections,¹¹² corresponding to CR-ecCCSD(T) and CR-ecCCSD(T,Q), would further improve on the present R-ecCCSD(T) method. The spin contamination of external sources, for example seen here in the HCI wave functions for stretched geometries, increases errors of ecCC methods, but this can be fixed by imposing S^2 symmetry on the source, as was done in the recent paper by Magoulas et al.⁹⁷ Finally, the current approximation, with its modest computational requirements on the external source, is applicable to the same scale of systems that can be handled by single reference coupled cluster methods. Thus, understanding the performance of this method

in larger correlated systems, particularly with approximate and low-cost Type-VI sources, is of interest.

■ APPENDIX: A SWEEP ALGORITHM CONVERTING MPS TO CI COEFFICIENTS WITH A THRESHOLD

In DMRG, the electronic wave function is represented by a matrix-product state (MPS)

$$|\Psi\rangle = \sum_{n_1, \dots, n_K} \sum_{\alpha_1, \dots, \alpha_{K-1}} A_{\alpha_1}^{n_1} A_{\alpha_1 \alpha_2}^{n_2} \cdots A_{\alpha_{K-1}}^{n_K} |n_1 n_2 \cdots n_K\rangle \quad (\text{A.1})$$

$$\{n\} = \{\text{vac}, \uparrow, \downarrow, \uparrow\downarrow\} \quad (\text{A.2})$$

where n_i is the occupation of orbital i , $|n_1 n_2 \cdots n_K\rangle$ is the occupation-number representation of a determinant, and α_i is an auxiliary index. Here, $\sum_{\alpha_i} A_{\alpha_i}^{n_i} A_{\alpha_i \alpha_{i+1}}^{n_{i+1}}$ denotes a matrix product, and it is assumed that the dimensions of all the auxiliary indices are the same (the bond dimension M). The wave function of the ground state can be optimized by the efficient DMRG sweep algorithm.

In order to get $Q_3^{\text{sc}}T_3$ and $Q_4^{\text{sc}}T_4$ in eq 4, we extract quadruple- and lower-order CI excitation amplitudes from the MPS by a sweep algorithm. To avoid repeated or unnecessary computation, we here describe how to obtain CI coefficients whose values are larger than a threshold in eq 5, with a concomitant reduction in computational cost and memory usage from a naive approach.

We first start with the MPS in left canonical form. During a sweep to compute the amplitude, at any given point (e.g., at some site p) one has a set of partial coefficients $c_{\alpha_p}(n_1 n_2 \cdots n_p) = \sum_{\alpha_1, \dots, \alpha_{p-1}} A_{\alpha_1}^{n_1} \cdots A_{\alpha_{p-1} \alpha_p}^{n_p}$. If $\sum_{\alpha_p} |c_{\alpha_p}(n_1 n_2 \cdots n_p)|^2 < \text{thresh}$, then this partial coefficient is dropped, as are all determinants involving the occupancy string $|n_1 n_2 \cdots n_p\rangle$. This is because if the MPS is in left canonical form, the above condition on the partial coefficient guarantees that the coefficient of any determinant generated by the MPS which contains $|n_1 n_2 \cdots n_p\rangle$ as a substring is also less than the threshold in magnitude. In addition, since the orbitals are associated with definite hole or particle character, we also drop any coefficient associated with more than four holes or four particles. Finally, in this process, we can take advantage of the conserved quantum numbers to only generate symmetry unique partial coefficients (e.g., if $S_z = 0$, then the values of $c_{\alpha_{p+1}}$ come in time-reversal pairs, and only one needs to be considered). Thus, using the above algorithm we can completely avoid generating any determinants with coefficients below the threshold.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.1c00205>.

Accurate SHCI energies of H_2O and N_2 (Table S1); source (DMRG and HCI) and ecCC energies (Tables S2–S8); CC energies (Table S10); and timing of ecCC calculations (Table S9) (XLSX)

■ AUTHOR INFORMATION

Corresponding Authors

Seunghoon Lee – Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, United States; orcid.org/0000-0003-3665-587X; Email: slee89@caltech.edu

Sandeep Sharma – Department of Chemistry, The University of Colorado at Boulder, Boulder, Colorado 80302, United States; orcid.org/0000-0002-6598-8887;

Email: sanshar@gmail.com

C. J. Umrigar – Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853, United States; Email: CyrusUmrigar@cornell.edu

Garnet Kin-Lic Chan – Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, United States; Email: gkc1000@gmail.com

Author

Huanchen Zhai – Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, United States; orcid.org/0000-0003-0086-0388

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.jctc.1c00205>

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Work by S.L., H.Z., and G.K.-L.C. was supported by the US National Science Foundation via Award CHE-1655333. G.K.-L.C. is a Simons Investigator. S.S. was supported by the US National Science Foundation grant CHE-1800584 and the Sloan research fellowship. C.J.U. was supported in part by the AFOSR under grant FA9550-18-1-0095.

■ REFERENCES

- (1) Abe, M. Diradicals. *Chem. Rev.* **2013**, *113*, 7011–7088.
- (2) Nakano, M. Electronic Structure of Open-Shell Singlet Molecules: Diradical Character Viewpoint. *Topics in Current Chemistry*; 2017; Vol. 375, DOI: [10.1007/978-3-319-93302-3_1](https://doi.org/10.1007/978-3-319-93302-3_1).
- (3) Kurashige, Y.; Chan, G. K.-L.; Yanai, T. Entangled quantum electronic wavefunctions of the Mn_4CaO_5 cluster in photosystem II. *Nat. Chem.* **2013**, *5*, 660–666.
- (4) Sharma, S.; Sivalingam, K.; Neese, F.; Chan, G. K.-L. Low-energy spectrum of iron-sulfur clusters directly from many-particle quantum mechanics. *Nat. Chem.* **2014**, *6*, 927–933.
- (5) Li, Z.; Guo, S.; Sun, Q.; Chan, G. K.-L. Electronic landscape of the P-cluster of nitrogenase as revealed through many-electron quantum wavefunction simulations. *Nat. Chem.* **2019**, *11*, 1026–1033.
- (6) Das, G.; Wahl, A. C. New Techniques for the Computation of Multiconfiguration Self-Consistent Field (MCSCF) Wavefunctions. *J. Chem. Phys.* **1972**, *56*, 1769–1775.
- (7) Werner, H.-J.; Meyer, W. A quadratically convergent multi-configuration-self-consistent field method with simultaneous optimization of orbitals and CI coefficients. *J. Chem. Phys.* **1980**, *73*, 2342–2356.
- (8) Werner, H.-J.; Knowles, P. J. A second order multiconfiguration SCF procedure with optimum convergence. *J. Chem. Phys.* **1985**, *82*, 5053–5063.
- (9) White, S. R.; Noack, R. M. Real-space quantum renormalization groups. *Phys. Rev. Lett.* **1992**, *68*, 3487.
- (10) White, S. R. Density matrix formulation for quantum renormalization groups. *Phys. Rev. Lett.* **1992**, *69*, 2863.
- (11) White, S. R. Density-matrix algorithms for quantum renormalization groups. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1993**, *48*, 10345.
- (12) White, S. R.; Martin, R. L. Ab initio quantum chemistry using the density matrix renormalization group. *J. Chem. Phys.* **1999**, *110*, 4127–4130.

- (13) Chan, G. K.-L.; Head-Gordon, M. Highly correlated calculations with a polynomial cost algorithm: A study of the density matrix renormalization group. *J. Chem. Phys.* **2002**, *116*, 4462–4476.
- (14) Legeza, Ö.; Röder, J.; Hess, B. Controlling the accuracy of the density-matrix renormalization-group method: The dynamical block state selection approach. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2003**, *67*, 125114.
- (15) Chan, G. K.-L. An algorithm for large scale density matrix renormalization group calculations. *J. Chem. Phys.* **2004**, *120*, 3172–3178.
- (16) Legeza, Ö.; Noack, R.; Sólyom, J.; Tincani, L. *Computational Many-Particle Physics*; Springer: 2008; Vol. 739, DOI: 10.1007/978-3-540-74686-7.
- (17) Chan, G.; Zgid, D. The Density Matrix Renormalization Group in Quantum Chemistry. *Annu. Rep. Comput. Chem.* **2009**, *5*, 149–162.
- (18) Marti, K. H.; Reiher, M. The density matrix renormalization group algorithm in quantum chemistry. *Z. Phys. Chem.* **2010**, *224*, 583–599.
- (19) Chan, G. K.-L.; Sharma, S. The density matrix renormalization group in quantum chemistry. *Annu. Rev. Phys. Chem.* **2011**, *62*, 465–481.
- (20) Sharma, S.; Chan, G. K.-L. Spin-adapted density matrix renormalization group algorithms for quantum chemistry. *J. Chem. Phys.* **2012**, *136*, 124121.
- (21) Wouters, S.; van Neck, D. The density matrix renormalization group for ab initio quantum chemistry. *Eur. Phys. J. D* **2014**, *68*, 272.
- (22) Szalay, S.; Pfeiffer, M.; Murg, V.; Barcza, G.; Verstraete, F.; Schneider, R.; Legeza, Ö. Tensor product methods and entanglement optimization for ab initio quantum chemistry. *Int. J. Quantum Chem.* **2015**, *115*, 1342.
- (23) Yanai, T.; Kurashige, Y.; Mizukami, W.; Chalupský, J.; Lan, T. N.; Saitow, M. Density matrix renormalization group for ab initio Calculations and associated dynamic correlation methods: A review of theory and applications. *Int. J. Quantum Chem.* **2015**, *115*, 283–299.
- (24) Ivanic, J.; Ruedenberg, K. Identification of deadwood in configuration spaces through general direct configuration interaction. *Theor. Chem. Acc.* **2001**, *106*, 339–351.
- (25) Huron, B.; Malrieu, J.; Rancurel, P. Iterative perturbation calculations of ground and excited state energies from multiconfigurational zeroth-order wavefunctions. *J. Chem. Phys.* **1973**, *58*, 5745–5759.
- (26) Garniron, Y.; Scemama, A.; Giner, E.; Caffarel, M.; Loos, P.-F. Selected configuration interaction dressed by perturbation. *J. Chem. Phys.* **2018**, *149*, 064103.
- (27) Loos, P.-F.; Scemama, A.; Blondel, A.; Garniron, Y.; Caffarel, M.; Jacquemin, D. A Mountaineering Strategy to Excited States: Highly Accurate Reference Energies and Benchmarks. *J. Chem. Theory Comput.* **2018**, *14*, 4360–4379.
- (28) Garniron, Y.; Scemama, A.; Loos, P.-F.; Caffarel, M. Hybrid stochastic-deterministic calculation of the second-order perturbative contribution of multireference perturbation theory. *J. Chem. Phys.* **2017**, *147*, 034101.
- (29) Buenker, R. J.; Peyerimhoff, S. D. Individualized configuration selection in CI calculations with subsequent energy extrapolation. *Theor. Chim. Acta* **1974**, *35*, 33–58.
- (30) Evangelista, S.; Daudey, J.-P.; Malrieu, J.-P. Convergence of an improved CIPSI algorithm. *J. Chem. Phys.* **1983**, *75*, 91–102.
- (31) Harrison, R. J. Approximating full configuration interaction with selected configuration interaction and perturbation theory. *J. Chem. Phys.* **1991**, *94*, 5021–5031.
- (32) Steiner, M.; Wenzel, W.; Wilson, K.; Wilkins, J. The efficient treatment of higher excitations in CI calculations: A comparison of exact and approximate results. *Chem. Phys. Lett.* **1994**, *231*, 263–268.
- (33) Neese, F. A spectroscopy oriented configuration interaction procedure. *J. Chem. Phys.* **2003**, *119*, 9428–9443.
- (34) Abrams, M. L.; Sherrill, C. D. Important configurations in configuration interaction and coupled-cluster wave functions. *Chem. Phys. Lett.* **2005**, *412*, 121–124.
- (35) Bytautas, L.; Ruedenberg, K. A priori identification of configurational deadwood. *Chem. Phys.* **2009**, *356*, 64–75. Moving Frontiers in Quantum Chemistry.
- (36) Evangelista, F. A. A driven similarity renormalization group approach to quantum many-body problems. *J. Chem. Phys.* **2014**, *141*, 054109.
- (37) Knowles, P. J. Compressive sampling in configuration interaction wavefunctions. *Mol. Phys.* **2015**, *113*, 1655–1660.
- (38) Schriber, J. B.; Evangelista, F. A. Communication: An adaptive configuration interaction approach for strongly correlated electrons with tunable accuracy. *J. Chem. Phys.* **2016**, *144*, 161106.
- (39) Tubman, N. M.; Lee, J.; Takeshita, T. Y.; Head-Gordon, M.; Whaley, K. B. A deterministic alternative to the full configuration interaction quantum Monte Carlo method. *J. Chem. Phys.* **2016**, *145*, 044112.
- (40) Liu, W.; Hoffmann, M. R. iCI: Iterative CI toward full CI. *J. Chem. Theory Comput.* **2016**, *12*, 1169–1178.
- (41) Caffarel, M.; Applencourt, T.; Giner, E.; Scemama, A. *Recent Progress in Quantum Monte Carlo*; ACS Publications: 2016; pp 15–46, DOI: 10.1021/bk-2016-1234.ch002.
- (42) Holmes, A. A.; Tubman, N. M.; Umrigar, C. J. Heat-Bath Configuration Interaction: An Efficient Selected Configuration Interaction Algorithm Inspired by Heat-Bath Sampling. *J. Chem. Theory Comput.* **2016**, *12*, 3674–3680.
- (43) Sharma, S.; Holmes, A. A.; Jeanmairet, G.; Alavi, A.; Umrigar, C. J. Semistochastic Heat-Bath Configuration Interaction Method: Selected Configuration Interaction with Semistochastic Perturbation Theory. *J. Chem. Theory Comput.* **2017**, *13*, 1595–1604.
- (44) Li, J.; Otten, M.; Holmes, A. A.; Sharma, S.; Umrigar, C. J. Fast Semistochastic Heat-Bath Configuration Interaction. *J. Chem. Phys.* **2018**, *149*, 214110.
- (45) Booth, G. H.; Thom, A. J.; Alavi, A. Fermion Monte Carlo without fixed nodes: A game of life, death, and annihilation in Slater determinant space. *J. Chem. Phys.* **2009**, *131*, 054106.
- (46) Cleland, D.; Booth, G. H.; Alavi, A. Communications: Survival of the fittest: Accelerating convergence in full configuration-interaction quantum Monte Carlo. *J. Chem. Phys.* **2010**, *132*, 041103.
- (47) Petruzielo, F. R.; Holmes, A. A.; Changlani, H. J.; Nightingale, M. P.; Umrigar, C. J. Semistochastic Projector Monte Carlo Method. *Phys. Rev. Lett.* **2012**, *109*, 230201.
- (48) Ghanem, K.; Lozovoi, A. Y.; Alavi, A. Unbiasing the initiator approximation in full configuration interaction quantum Monte Carlo. *J. Chem. Phys.* **2019**, *151*, 224108.
- (49) Sabzevari, I.; Sharma, S. Improved speed and scaling in orbital space variational Monte Carlo. *J. Chem. Theory Comput.* **2018**, *14*, 6276–6286.
- (50) Neuscamman, E. Subtractive manufacturing with geminal powers: making good use of a bad wave function. *Mol. Phys.* **2016**, *114*, 577–583.
- (51) Andersson, K.; Malmqvist, P. A.; Roos, B. O.; Sadlej, A. J.; Wolinski, K. Second-order perturbation theory with a CASSCF reference function. *J. Phys. Chem.* **1990**, *94*, 5483–5488.
- (52) Angeli, C.; Cimiraglia, R.; Evangelista, S.; Leininger, T.; Malrieu, J.-P. Introduction of n-electron valence states for multi-reference perturbation theory. *J. Chem. Phys.* **2001**, *114*, 10252–10264.
- (53) Angeli, C.; Cimiraglia, R.; Malrieu, J.-P. N-electron valence state perturbation theory: a fast implementation of the strongly contracted variant. *Chem. Phys. Lett.* **2001**, *350*, 297–305.
- (54) Angeli, C.; Cimiraglia, R.; Malrieu, J.-P. n-electron valence state perturbation theory: A spinless formulation and an efficient implementation of the strongly contracted and of the partially contracted variants. *J. Chem. Phys.* **2002**, *117*, 9138–9153.
- (55) Neuscamman, E.; Yanai, T.; Chan, G. K.-L. A review of canonical transformation theory. *Int. Rev. Phys. Chem.* **2010**, *29*, 231–271.
- (56) Yanai, T.; Kurashige, Y.; Neuscamman, E.; Chan, G. K.-L. Multireference quantum chemistry through a joint density matrix

renormalization group and canonical transformation theory. *J. Chem. Phys.* **2010**, *132*, 024105.

(57) Kurashige, Y.; Yanai, T. Second-order perturbation theory with a density matrix renormalization group self-consistent field reference function: Theory and application to the study of chromium dimer. *J. Chem. Phys.* **2011**, *135*, 094104.

(58) Sharma, S.; Chan, G. K.-L. Communication: A flexible multi-reference perturbation theory by minimizing the Hylleraas functional with matrix product states. *J. Chem. Phys.* **2014**, *141*, 111101.

(59) Sharma, S.; Alavi, A. Multireference linearized coupled cluster theory for strongly correlated systems using matrix product states. *J. Chem. Phys.* **2015**, *143*, 102815.

(60) Guo, S.; Watson, M. A.; Hu, W.; Sun, Q.; Chan, G. K.-L. N-electron valence state perturbation theory based on a density matrix renormalization group reference function, with applications to the chromium dimer and a trimer model of poly (p-phenylenevinylene). *J. Chem. Theory Comput.* **2016**, *12*, 1583–1591.

(61) Sokolov, A. Y.; Guo, S.; Ronca, E.; Chan, G. K.-L. Time-dependent N-electron valence perturbation theory with matrix product state reference wavefunctions for large active spaces and basis sets: Applications to the chromium dimer and all-trans polyenes. *J. Chem. Phys.* **2017**, *146*, 244102.

(62) Sharma, S.; Holmes, A. A.; Jeanmairet, G.; Alavi, A.; Umrigar, C. J. Semistochastic heat-bath configuration interaction method: Selected configuration interaction with semistochastic perturbation theory. *J. Chem. Theory Comput.* **2017**, *13*, 1595–1604.

(63) Guo, S.; Li, Z.; Chan, G. K.-L. Communication: An efficient stochastic algorithm for the perturbative density matrix renormalization group in large active spaces. *J. Chem. Phys.* **2018**, *148*, 221104.

(64) Lischka, H.; Shepard, R.; Brown, F. B.; Shavitt, I. New implementation of the graphical unitary group approach for multireference direct configuration interaction calculations. *Int. J. Quantum Chem.* **1981**, *20*, 91–100.

(65) Szalay, P. G.; Muller, T.; Gidofalvi, G.; Lischka, H.; Shepard, R. Multiconfiguration self-consistent field and multireference configuration interaction methods and applications. *Chem. Rev.* **2012**, *112*, 108–181.

(66) Lischka, H.; Shepard, R.; Pitzer, R. M.; Shavitt, I.; Dallos, M.; Müller, T.; Szalay, P. G.; Seth, M.; Kedziora, G. S.; Yabushita, S.; et al. High-level multireference methods in the quantum-chemistry program system COLUMBUS: Analytic MR-CISD and MR-AQCC gradients and MR-AQCC-LRT for excited states, GUGA spin-orbit CI and parallel CI density. *Phys. Chem. Chem. Phys.* **2001**, *3*, 664–673.

(67) Saitow, M.; Kurashige, Y.; Yanai, T. Fully internally contracted multireference configuration interaction theory using density matrix renormalization group: A reduced-scaling implementation derived by computer-aided tensor factorization. *J. Chem. Theory Comput.* **2015**, *11*, 5120–5131.

(68) Paldus, J.; Li, X. A critical assessment of coupled cluster method in quantum chemistry. *Adv. Chem. Phys.* **2007**, *110*, 1–175.

(69) Kong, L.; Shamasundar, K.; Demel, O.; Nooijen, M. State specific equation of motion coupled cluster method in general active space. *J. Chem. Phys.* **2009**, *130*, 114101.

(70) Jeziorski, B. Multireference coupled-cluster Ansatz. *Mol. Phys.* **2010**, *108*, 3043–3054.

(71) Čížek, J. On the correlation problem in atomic and molecular systems. Calculation of wavefunction components in Ursell-type expansion using quantum-field theoretical methods. *J. Chem. Phys.* **1966**, *45*, 4256–4266.

(72) Čížek, J. On the use of the cluster expansion and the technique of diagrams in calculations of correlation effects in atoms and molecules. *Adv. Chem. Phys.* **2007**, *35*, 89.

(73) Čížek, J.; Paldus, J. Correlation problems in atomic and molecular systems III. Rederivation of the coupled-pair many-electron theory using the traditional quantum chemical method. *Int. J. Quantum Chem.* **1971**, *5*, 359–379.

(74) Paldus, J.; Čížek, J.; Shavitt, I. Correlation Problems in Atomic and Molecular Systems. IV. Extended Coupled-Pair Many-Electron

Theory and Its Application to the BH₃ Molecule. *Phys. Rev. A: At, Mol., Opt. Phys.* **1972**, *5*, 50.

(75) Bartlett, R. J. In *Modern Electronic Structure Theory*; Yarkony, D. R., Ed.; World Scientific: 1995; Vol. 1, pp 1047–1131, DOI: 10.1142/9789812832115_0005.

(76) Shavitt, I.; Bartlett, R. J. *Many-body methods in chemistry and physics: MBPT and coupled-cluster theory*; Cambridge University Press: 2009; DOI: 10.1017/CBO9780511596834.

(77) Kinoshita, T.; Hino, O.; Bartlett, R. J. Coupled-cluster method tailored by configuration interaction. *J. Chem. Phys.* **2005**, *123*, 074106.

(78) Hino, O.; Kinoshita, T.; Chan, G. K.-L.; Bartlett, R. J. Tailored coupled cluster singles and doubles method applied to calculations on molecular structure and harmonic vibrational frequencies of ozone. *J. Chem. Phys.* **2006**, *124*, 114311.

(79) Lyakh, D. I.; Lotrich, V. F.; Bartlett, R. J. The ‘tailored’CCSD (T) description of the automerization of cyclobutadiene. *Chem. Phys. Lett.* **2011**, *501*, 166–171.

(80) Veis, L.; Antalík, A.; Brabec, J.; Neese, F.; Legeza, O.; Pittner, J. Coupled cluster method with single and double excitations tailored by matrix product state wave functions. *J. Phys. Chem. Lett.* **2016**, *7*, 4072–4078.

(81) Faulstich, F. M.; Máté, M.; Laestadius, A.; Csirik, M. A.; Veis, L.; Antalík, A.; Brabec, J.; Schneider, R.; Pittner, J.; Kvaal, S.; et al. Numerical and theoretical aspects of the DMRG-TCC method exemplified by the nitrogen dimer. *J. Chem. Theory Comput.* **2019**, *15*, 2206–2220.

(82) Vitale, E.; Alavi, A.; Kats, D. FCIQMC-tailored distinguishable cluster approach. *J. Chem. Theory Comput.* **2020**, *16*, 5621–5634.

(83) Paldus, J.; Planelles, J. Valence bond corrected single reference coupled cluster approach. *Theor. Chim. Acta* **1994**, *89*, 13–31.

(84) Planelles, J.; Paldus, J.; Li, X. Valence bond corrected single reference coupled cluster approach. *Theor. Chim. Acta* **1994**, *89*, 33.

(85) Planelles, J.; Paldus, J.; Li, X. Valence bond corrected single reference coupled cluster approach. *Theor. Chim. Acta* **1994**, *89*, 59.

(86) Paldus, J. Externally and internally corrected coupled cluster approaches: an overview. *J. Math. Chem.* **2017**, *55*, 477–502.

(87) Paldus, J.; Boyle, M. Cluster analysis of the full configuration interaction wave functions of cyclic polyene models. *Int. J. Quantum Chem.* **1982**, *22*, 1281–1305.

(88) Paldus, J.; Chin, E.; Grey, M. Bond length alternation in cyclic polyenes. II. Unrestricted hartree-fock method. *Int. J. Quantum Chem.* **1983**, *24*, 395–409.

(89) Paldus, J.; Čížek, J.; Takahashi, M. Approximate account of the connected quadruply excited clusters in the coupled-pair many-electron theory. *Phys. Rev. A: At, Mol., Opt. Phys.* **1984**, *30*, 2193.

(90) Piecuch, P.; Tobola, R.; Paldus, J. Approximate account of connected quadruply excited clusters in single-reference coupled-cluster theory via cluster analysis of the projected unrestricted Hartree-Fock wave function. *Phys. Rev. A: At, Mol., Opt. Phys.* **1996**, *54*, 1210.

(91) Peris, G.; Planelles, J.; Paldus, J. Single-reference CCSD approach employing three- and four-body CAS SCF corrections: A preliminary study of a simple model. *Int. J. Quantum Chem.* **1997**, *62*, 137–151.

(92) Li, X.; Peris, G.; Planelles, J.; Rajadell, F.; Paldus, J. Externally corrected singles and doubles coupled cluster methods for open-shell systems. *J. Chem. Phys.* **1997**, *107*, 90–98.

(93) Peris, G.; Rajadell, F.; Li, X.; Planelles, J.; Paldus, J. Externally corrected singles and doubles coupled cluster methods for open-shell systems. II. Applications to the low lying doublet states of OH, NH₂, CH₃ and CN radicals. *Mol. Phys.* **1998**, *94*, 235–248.

(94) Stolarczyk, L. Z. Complete active space coupled-cluster method. Extension of single-reference coupled-cluster method using the CASSCF wavefunction. *Chem. Phys. Lett.* **1994**, *217*, 1–6.

(95) Xu, E.; Li, S. The externally corrected coupled cluster approach with four- and five-body clusters from the CASSCF wave function. *J. Chem. Phys.* **2015**, *142*, 094119.

- (96) Deustua, J. E.; Magoulas, I.; Shen, J.; Piecuch, P. Communication: Approaching exact quantum chemistry by cluster analysis of full configuration interaction quantum Monte Carlo wave functions. *J. Chem. Phys.* **2018**, *149*, 151101.
- (97) Magoulas, I.; Gururangan, K.; Piecuch, P.; Deustua, J. E.; Shen, J. Is Externally Corrected Coupled Cluster Always Better than the Underlying Truncated Configuration Interaction. 2021.
- (98) Li, X.; Paldus, J. Reduced multireference CCSD method: An effective approach to quasidegenerate states. *J. Chem. Phys.* **1997**, *107*, 6257–6269.
- (99) Li, X.; Paldus, J. Truncated version of the reduced multireference coupled-cluster method with perturbation selection of higher than pair clusters. *Int. J. Quantum Chem.* **2000**, *80*, 743–756.
- (100) Li, X.; Paldus, J. Reduced multireference coupled cluster method with singles and doubles: Perturbative corrections for triples. *J. Chem. Phys.* **2006**, *124*, 174101.
- (101) Li, X.; Paldus, J. Dissociation of N₂ triple bond: a reduced multireference CCSD study. *Chem. Phys. Lett.* **1998**, *286*, 145–154.
- (102) Li, X.; Paldus, J. Singlet-triplet splitting in methylene: An accurate description of dynamic and nondynamic correlation by reduced multireference coupled cluster method. *Collect. Czech. Chem. Commun.* **1998**, *63*, 1381–1393.
- (103) Li, X.; Paldus, J. Simultaneous handling of dynamical and nondynamical correlation via reduced multireference coupled cluster method: Geometry and harmonic force field of ozone. *J. Chem. Phys.* **1999**, *110*, 2844–2852.
- (104) Li, X.; Paldus, J. Reduced multireference coupled cluster method: Ro-vibrational spectra of N₂. *J. Chem. Phys.* **2000**, *113*, 9966–9977.
- (105) Li, X.; Paldus, J. Effect of spin contamination on the prediction of barrier heights by coupled-cluster theory: F + H₂ → HF + H reaction. *Int. J. Quantum Chem.* **2000**, *77*, 281–290.
- (106) Li, X.; Paldus, J. A truncated version of reduced multireference coupled-cluster method with singles and doubles and noniterative triples: Application to F₃ and Ni(CO)_n (*n* = 1, 2, and 4). *J. Chem. Phys.* **2006**, *125*, 164107.
- (107) Li, X.; Paldus, J. Singlet-triplet separation in BN and C₂: simple yet exceptional systems for advanced correlated methods. *Chem. Phys. Lett.* **2006**, *431*, 179–184.
- (108) Li, X.; Paldus, J. Reduced multireference coupled-cluster method: Barrier heights for heavy atom transfer, nucleophilic substitution, association, and unimolecular reactions. *J. Phys. Chem. A* **2007**, *111*, 11189–11197.
- (109) Li, X.; Paldus, J. Coupled-cluster approach to spontaneous symmetry breaking in molecules: The linear N₃ radical. *Int. J. Quantum Chem.* **2008**, *108*, 2117–2127.
- (110) Li, X.; Paldus, J. Full potential energy curve for N₂ by the reduced multireference coupled-cluster method. *J. Chem. Phys.* **2008**, *129*, 054104.
- (111) Yao, Y.; Giner, E.; Li, J.; Toulouse, J.; Umrigar, C. Almost exact energies for the Gaussian-2 set with the semistochastic heat-bath configuration interaction method. *J. Chem. Phys.* **2020**, *153*, 124117.
- (112) Kowalski, K.; Piecuch, P. The method of moments of coupled-cluster equations and the renormalized CCSD[T], CCSD(T), CCSD(TQ), and CCSDT(Q) approaches. *J. Chem. Phys.* **2000**, *113*, 18.
- (113) 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.
- (114) Sun, Q.; Berkelbach, T. C.; Blunt, N. S.; Booth, G. H.; Guo, S.; Li, Z.; Liu, J.; McClain, J. D.; Sayfutyarova, E. R.; Sharma, S.; Wouters, S.; Chan, G. K. PySCF: the Python-based simulations of chemistry framework. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2018**, *8*, e1340.
- (115) Chan, G. K.-L.; Head-Gordon, M. Highly correlated calculations with a polynomial cost algorithm: A study of the density matrix renormalization group. *J. Chem. Phys.* **2002**, *116*, 4462–4476.
- (116) Chan, G. K.-L. An algorithm for large scale density matrix renormalization group calculations. *J. Chem. Phys.* **2004**, *120*, 3172–3178.
- (117) Ghosh, D.; Hachmann, J.; Yanai, T.; Chan, G. K.-L. Orbital optimization in the density matrix renormalization group, with applications to polyenes and $\hat{I}^2$ -carotene. *J. Chem. Phys.* **2008**, *128*, 144117.
- (118) Sharma, S.; Chan, G. K.-L. Spin-adapted density matrix renormalization group algorithms for quantum chemistry. *J. Chem. Phys.* **2012**, *136*, 124121.
- (119) Sharma, S.; Holmes, A. A.; Jeanmairet, G.; Alavi, A.; Umrigar, C. J. Semistochastic Heat-Bath Configuration Interaction Method: Selected Configuration Interaction with Semistochastic Perturbation Theory. *J. Chem. Theory Comput.* **2017**, *13*, 1595–1604.
- (120) Holmes, A. A.; Tubman, N. M.; Umrigar, C. J. Heat-bath Configuration Interaction: An efficient selected CI algorithm inspired by heat-bath sampling. *J. Chem. Theory Comput.* **2016**, *12*, 3674–3680.
- (121) Smith, J. E. T.; Mussard, B.; Holmes, A. A.; Sharma, S. Cheap and Near Exact CASSCF with Large Active Spaces. *J. Chem. Theory Comput.* **2017**, *13*, 5468–5478.
- (122) Nielsen, I. M. B.; Janssen, C. L. Double-substitution-based diagnostics for coupled-cluster and Møller-Plesset perturbation theory. *Chem. Phys. Lett.* **1999**, *310*, 568–576.