Orbital-invariant spin-extended approximate coupled-cluster for multi-reference systems

Tsuchimochi, Takashi

Ten-no, L. Seiichiro

The Journal of Chemical Physics

Orbital-invariant spin-extended approximate coupled-cluster for multi-reference systems

Cite as: J. Chem. Phys. **149**, 044109 (2018); https://doi.org/10.1063/1.5036542
Submitted: 16 April 2018 . Accepted: 02 July 2018 . Published Online: 25 July 2018

Takashi Tsuchimochi, and Seiichiro L. Ten-no

Orbital-invariant spin-extended approximate coupled-cluster for multi-reference systems

Takashi Tsuchimochi[1,2,a)] and Seiichiro L. Ten-no[1,2,b)]
[1]*Graduate School of Science, Technology, and Innovation, Kobe University, Kobe, Hyogo 657-0025, Japan*
[2]*Graduate School of System Informatics, Kobe University, Kobe, Hyogo 657-0025, Japan*

(Received 16 April 2018; accepted 2 July 2018; published online 25 July 2018)

We present an approximate treatment of spin-extended coupled-cluster (ECC) based on the spin-projection of the broken-symmetry coupled-cluster (CC) *ansatz*. ECC completely eliminates the spin-contamination of unrestricted CC and is therefore expected to provide better descriptions of dynamical and static correlation effects, but introduces two distinct problems. The first issue is the emergence of non-terminating amplitude equations, which are caused by the de-excitation effects inherent in symmetry projection operators. In this study, we take a minimalist approach and truncate the Taylor series of the exponential *ansatz* at a certain order such that the approximation safely recovers the traditional CC without spin-projection. The second issue is that the nonlinear equations of ECC become underdetermined, although consistent, yielding an infinitude of solutions. This problem arises because of the redundancies in the excitation manifold, as is common in other multi-reference approaches. We remove the linear dependencies in ECC by employing an orthogonal projection manifold. We also propose an efficient solver for our method, in which the components are usually sparse but not diagonal-dominant. It is shown that our approach is rigorously orbital-invariant and provides more accurate results than its configuration interaction and linearized CC analogues for chemical systems. *Published by AIP Publishing.* https://doi.org/10.1063/1.5036542

I. INTRODUCTION

Treating quasi-degeneracies accurately is one of the most challenging tasks in electronic structure theory, as many single-reference (SR) methods are not suitable for capturing static correlations in general and thus typically break down. The failures of SR methods are ascribed to their deficient starting point, Hartree–Fock (HF). Although a number of approaches have attempted to approximate static correlation effects in an SR framework,[1–9] multi-reference (MR) methods are usually more reliable and accurate. The fundamental idea is to handle static correlations at some multi-configuration level such as the complete active space self-consistent field (CASSCF)[10–12] and add dynamical correlations using perturbation theory, configuration interaction (CI), or coupled-cluster (CC).[13–16]

Thus far, there have been various attempts to develop MRCC approaches. Many Hilbert-space methods are based on the Jeziorski–Monkhorst *ansatz*, where each determinant in the reference space associates with its own cluster operator.[17] In the state-universal coupled-cluster (SUCC) formulation, all the amplitudes and electronic states are simultaneously determined by diagonalizing the effective Hamiltonian, but its application is severely hampered by the intruder state problem. The state-specific approaches such as the Brillouin–Wigner (BW) MRCC[18–21] and Mukherjee (Mk) MRCC[22–27] avoids the intruder state problem by considering only one (lowest) state; however, they in turn suffer from the redundancy problem where the number of amplitude equations is less than the number of amplitudes. In these state-specific approaches, the redundancy problem is solved by sufficiency conditions,[18–20,22,23] which unfortunately bring about other undesired consequences such as the improper residuals[28] and the lack of orbital invariance.[29,30] Finally, the internally contracted (*ic*) approaches resolve these issues, by employing a single cluster operator along with a multi-configuration *vacuum*,[31–33] although the prefactor is significantly large and grows rapidly with the size of the active space. The properties and capabilities of these methods as well as other variants of MRCC are well documented in the recent review papers.[34,35] With these difficulties and complications, it should be fair to point out that most MRCC methods are not straightforward to formulate. They are also cost-intensive and require *a priori* specification of an active space where electrons are strongly correlated. Hence, it is important to develop efficient and black-box MRCC schemes in an orbital-invariant fashion. In the present work, we aim at formulating such a method based on symmetry projection.

Symmetry-projected HF (PHF) appears to be a good compromise between HF and CASSCF.[36,37] By writing a symmetry-projection operator as

$$\hat{P} = \int w(\Omega)\hat{R}(\Omega)\, d\Omega, \tag{1}$$

its wave function comprises a few non-orthogonal broken-symmetry determinants, which are systematically generated

a)Electronic mail: tsuchimochi@gmail.com
b)Electronic mail: tenno@garnet.kobe-u.ac.jp

by some unitary rotation operator $\hat{R}(\Omega)$ specific to the symmetry that is to be restored. In the orthogonal orbital picture, a PHF wave function is also expressed as a linear combination of determinants through all orders of excitations with respect to some symmetry-adapted determinant.[38] Among the symmetries that Fermionic wave functions can break, spin symmetry has been of great interest in quantum chemistry because spontaneous symmetry breaking in spin-unrestricted HF (UHF) is usually associated with quasi-degeneracies in the system. As a result, spin-projected UHF (SUHF) brings a large amount of static correlations as well as physically correct spin states, all in a black-box manner. Many studies have shown its improved accuracy over both restricted HF (RHF) and UHF,[37,39–41] suggesting that it is a more suitable starting point.

We have therefore proposed several post-SUHF methods to include the residual dynamical correlation effect. These include spin-extended perturbation theory[42] and CI with singles and doubles (ECISD).[43,44] It was shown that these approaches generically outperform their SR analogues. We have further explored a generalization to coupled electron-pair approximation (ECEPA) including spin-extended linearized CC (ELCC) and averaged quadratic CC (EAQCC).[45] This builds on ECISD to furnish approximate size-consistent components in the dynamical correlation by writing the quadratic $\hat{P}\hat{T}^2$ elements as products of the correlation energy and linear $\hat{P}\hat{T}$ terms, where $\hat{T}$ is the excitation operator. ECEPA methods were shown to improve both the correlation energy and molecular properties over ECISD.[45] However, as the exclusion-principle-violating terms are treated in an average manner, they are plagued by instabilities; for average CEPA to be a good approximation, the reference state has to be well separated from the interacting space, but this is often not the case in SUHF and ECISD. The same problem is also present in the CAS case, unless the active space is large enough.[46]

From this point of view, it is obviously desirable to develop CC from SUHF, a theory we call spin-extended CC (ECC). The wave function *ansatz* for ECC is given by

$$|\Psi\rangle = \hat{P}e^{\hat{T}}|\Phi_0\rangle, \tag{2}$$

where $\hat{T}$ is the excitation operator and $|\Phi_0\rangle$ is an underlying broken-symmetry determinant of SUHF. There have recently been efforts in similar directions by other research groups.[47–51] In particular, the spin-projected unrestricted CC developed by Scuseria and coworkers uses the same *ansatz*.[51] There are, however, two main difficulties in formulating the rigorous ECC. The first obvious problem is that the corresponding equations do not terminate at finite order because $\hat{R}(\Omega)$ is unitary and $\hat{P}$ is thus an n_e-body operator, where n_e is the number of electrons in the system. In the pioneering work of Qiu *et al.*, $\hat{P}$ was written as a polynomial of particle-hole excitations with respect to some reference RHF-type determinant,[52] and then combined with a cluster operator followed by variational optimization of amplitudes.[50] However, such normal-ordering of $\hat{P}$ loses the de-excitations from the CC *ansatz*, which can play a significant role in describing static correlations. In addition, the variational CC approach results in the same complexity as full CI (FCI). Very recently, Qiu *et al.* proposed a truncation scheme to incorporate the de-excitation effects approximately on the basis of the excitation rank of so-called disentangled cluster operators.[51] The method essentially approximates $e^{\hat{V}_1(\Omega)}e^{\hat{T}}$, where $e^{\hat{V}_1(\Omega)}$ is the de-excitation component in $\hat{R}(\Omega)$ and thus no longer produces pure spin-states in general, although the spin-contamination introduced was shown to be satisfactorily small for a model Hamiltonian.

The second difficulty, which was not realized in previous work,[47,51] is the *redundancy* problem inherent to many other MR methods.[31–33,53,54] In essence, the emergence of nonlinear parametrization in Eq. (2) gives rise to linearly dependent t-amplitudes and equations because of the non-orthogonality of the excitation manifold. Such redundancies are common in ECI and ELCC, but do not pose any problem in these methods because they are strictly variational or quasi-variational in the sense that the CI coefficients or t-amplitudes are optimized with respect to a well-defined energy functional.[44,45] Therefore, these linear methods provide only one unique (lowest) energy, even though there are infinitely many sets of amplitudes that satisfy the underlying generalized eigenvalue problem. However, when ECC is solved in a non-variational, projective manner, its correlation energy is arbitrary up to linear dependencies. As in various *ic* MRCC approaches,[32,33] we need to orthogonalize the projection manifold by performing singular-value-decomposition (diagonalization) of some metric of the nonlinear ECC system.

In light of the latter problem, a well-defined metric that is, ideally, strictly positive-semidefinite is indispensable. While it is possible to systematically include only the important components of high powers of $\hat{R}(\Omega)e^{\hat{T}}$, such approximations can make it unclear which metric the nonlinear redundant problem is to be solved under. In fact, in the course of our study, we have found several cases in which approximating terms in $\hat{R}(\Omega)e^{\hat{T}}$ causes severe convergence problems. This instability is likely to be attributable to an inconsistent metric; as a result, the nonlinear ECC equations become unstable or even unsolvable. For this reason, we wish to avoid introducing approximations to $\hat{P}$ at all costs.

To this end, we will simply truncate the Taylor expansion while securing the (numerical) exactness of the projection operator, thus giving a metric with the desirable properties (symmetric, positive-semidefinite). This permits us to remove the redundancies in the equations correctly and to study the improvements that the quadratic and higher-order terms bring about, especially when the t-amplitudes are optimized in the presence of $\hat{P}$. As will be shown, the divergence behavior of ECEPA can be rectified in the truncated ECCSD method, while offering improved performance over ECISD, just like in the SR case.

The resultant residual equations are still non-orthogonal and non-diagonal-dominant, and cannot be solved by iterative diagonalization, unlike linearized schemes. In this paper, following our work on ECI,[44] we propose a robust converger for solving the nonlinear ECC equations. The use of our scheme in conjunction with the direct inversion of iterative subspace (DIIS)[55–57] gives stable and fast convergence of the t-amplitudes.

We organize this article as follows. As the appearance of redundancies is general, regardless of the approximations introduced to the ECC *ansatz* (as far as it is nonlinear), we first address this issue in Sec. II. Then, in Sec. III, we rewrite the ECCSD wave function as a superposition of non-orthogonal pseudo-CI wave functions in terms of normal-ordering, where the effective coefficients are approximated based on the truncation of the cluster expansion. Sec. IV describes how the truncated ECCSD equations can be solved in practice. Sec. VI confirms the orbital invariance of the method and reports the convergence behavior of the residuals. We also demonstrate the performance of the method for the H4 and H8 models and O_3 as pilot examples. Finally, we discuss the size-consistency problem of the method and the structure of metric eigenvalues.

II. LINEAR DEPENDENCIES

A. Background

Throughout this work, we will adopt the conventional notation for orbitals, namely, p, q, r, s, ... for general, i, j, k, l, ... for occupied, and a, b, c, d, ... for virtual spin-orbitals. Greek subscripts are used to represent excited determinants, whereas $|\Phi_0\rangle$ denotes the reference broken-symmetry determinant.

For ECCSD in which $\hat{T}$ is truncated at doubles, we have

$$\hat{T} = \hat{T}_1 + \hat{T}_2, \tag{3a}$$

$$\hat{T}_1 = \sum_{ia} t_i^a \hat{a}_i^a, \tag{3b}$$

$$\hat{T}_2 = \frac{1}{4}\sum_{ijab} t_{ij}^{ab} \hat{a}_{ij}^{ab}, \tag{3c}$$

where $\hat{a}_{ij\cdots}^{ab\cdots}$ are the excitation operators written with creation ($\hat{a}^p$) and annihilation ($\hat{a}_p$) operators

$$\hat{a}_{ij\cdots}^{ab\cdots} = \hat{a}^a\hat{a}^b\ldots\hat{a}_j\hat{a}_i. \tag{4}$$

In this work, we only consider a UHF-type determinant for $|\Phi_0\rangle$ and the collinear excitations to preserve the $\hat{S}_z$ quantum number, e.g., i and a share the same spin in t_i^a. For later use, Eqs. (3) are written in the shorthand notation

$$\hat{T} = \sum_{\mu}^{M_{\rm SD}} t_\mu \hat{E}_\mu, \tag{5}$$

where μ runs over the singles and doubles spaces and $M_{\rm SD}$ is the dimension. The excitation operators $\hat{a}_i^a$ and $\hat{a}_{ij}^{ab}$ are merged into the row vector $\hat{\mathbf{E}}$, where each component generates excited (broken-symmetry) determinants from $|\Phi_0\rangle$, i.e.,

$$|\Phi_\mu\rangle = \hat{E}_\mu|\Phi_0\rangle. \tag{6}$$

As in regular CC, the correlation energy $E_{\rm c}$ and t-amplitudes may fulfill the following equations:

$$\langle\Phi_0|\hat{H}_N|\Psi\rangle = E_{\rm c}\langle\Phi_0|\Psi\rangle, \tag{7a}$$

$$\langle\Phi_0|\hat{E}_\mu^\dagger\hat{H}_N|\Psi\rangle = E_{\rm c}\langle\Phi_0|\hat{E}_\mu^\dagger|\Psi\rangle, \tag{7b}$$

where $\hat{H}_N \equiv \hat{H} - E_{\rm SUHF}$ with

$$E_{\rm SUHF} = \frac{\langle\Phi_0|\hat{P}^\dagger\hat{H}\hat{P}|\Phi_0\rangle}{\langle\Phi_0|\hat{P}^\dagger\hat{P}|\Phi_0\rangle} = \langle\Phi_0|\hat{H}\hat{P}|\Phi_0\rangle. \tag{8}$$

We note that $\hat{P}$ is idempotent, Hermitian, and commutable with $\hat{H}$. Henceforth, we always assume the normalization of an SUHF wave function, but not of $|\Phi_0\rangle$. Note that the Hamiltonian is not similarity-transformed in Eqs. (7), and therefore, the left hand side includes unlinked terms. These are mostly expected to cancel out with the right hand side. As we have already mentioned, the above equation is non-terminating because $\hat{P}$ involves $n_{\rm e}$ de-excitations.

Furthermore, the above set of equations does not provide the proper residuals because of the redundancies. The redundancy problem in ECC is essential because the excitation manifold of Eq. (6) is non-orthogonal after the projection operator $\hat{P}$ is applied, and certain linear combinations will create the null basis. In other words, such linear combinations of excitation operators can create a mixed state that has no component of the target symmetry space of $\hat{P}$. This is shown in Fig. 1, where $|\Phi_0\rangle$ is set to an RHF determinant for simplicity, but two excited determinants $|\Phi_{\rm A}\rangle$ and $|\Phi_{\rm B}\rangle$ generate the same configuration state function once the singlet projection operator $\hat{P}^{\rm S}$ is enacted and are therefore redundant. A linear combination of the two bare excited determinants (without projection), $|\Phi_{\rm A}\rangle - |\Phi_{\rm B}\rangle$, is a pure triplet state, which is destroyed upon the action of $\hat{P}^{\rm S}$.

Therefore, although the nonlinear system of ECCSD as given by Eq. (7b) consists of $M_{\rm SD}$ equations and variables t_μ, it is underdetermined (overparametrized). A consequence of this is an infinite number of solutions that fulfill *both* Eqs. (7a) and (7b), i.e., $E_{\rm c}$ is arbitrary. This is somewhat different from the redundancy problem that appears in *state*-specific MRCC approaches based on the Jeziorski–Monkhorst *ansatz*,[17] which can be solved by invoking appropriate sufficiency conditions,[18,19,22,23] at the cost of orbital invariance.[30] Our redundancy problem arises from the non-orthogonal character of the excitation manifold, and, in this respect, is rather similar to the situation in *ic* MR methods.[31–33,53,54]

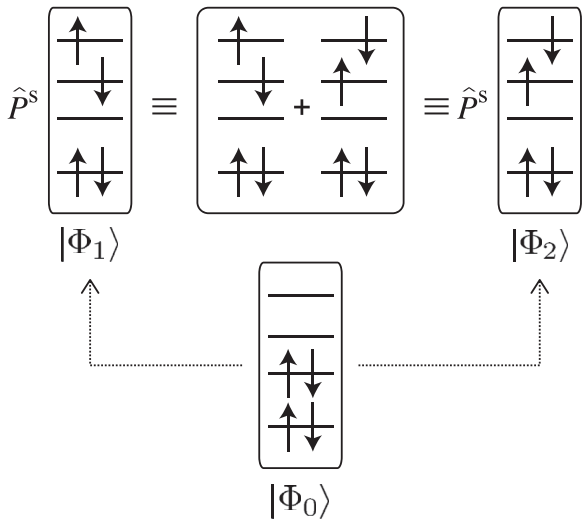

FIG. 1. Redundancy in the singlet-projected excitation manifold. For simplicity, the reference state is RHF, but excited determinants become identical when spin-projected.

One possible method of removing the redundancies is to make use of the property of spin-projection. In previous work, we numerically found that the number of linear dependencies in ECISD is always associated with the number of single excitations for tested singlet systems.[44] Further mathematical analyses should make it possible to identify the redundant excitations analytically for spin-projection, which can then be explicitly removed from the equations and amplitudes. Nevertheless, it is questionable whether this approach retains unitary invariance because there are multiple choices for the removal of such excitations; namely, either $|\Phi_A\rangle$ or $|\Phi_B\rangle$ can be removed. Ideally, any procedure should ensure a unique solution that is orbital-invariant. In the following, we outline such a protocol.

B. Intermediate normalized *ansatz*

Before eliminating the redundancies in Eqs. (7), we need a clear separation between the reference and excitation manifold. As $|\Phi_\mu\rangle$ overlaps with $|\Phi_0\rangle$ through $\hat{P}$, the original ECC *ansatz* $\hat{P}e^{\hat{T}}|\Phi_0\rangle$ is not intermediate normalized. This is not a desirable characteristic when we wish to solve our redundancy problem because the dimensions of the redundant parameters t_μ and redundant basis $\{\Phi_0, \Phi_\mu\}$ are inconsistent. Accordingly, as previously suggested in ECEPA theory,[45] the projection of the reference SUHF space is first performed, restricting the linear dependence problem only to the excitation manifold.

To this end, we prepare the projector $\hat{Q}$, which eliminates the SUHF reference $\hat{P}|\Phi_0\rangle$, as

$$\hat{Q} = 1 - \hat{P}|\Phi_0\rangle\langle\Phi_0|\hat{P}, \tag{9}$$

and redefine the ECC wave function as

$$|\Psi'\rangle = \left(\hat{P} + \hat{Q}e^{\hat{T}}\right)|\Phi_0\rangle = |\Psi\rangle + \hat{P}|\Phi_0\rangle(1 - \langle\Phi_0|\Psi\rangle), \tag{10}$$

where we have conveniently defined $\hat{Q} = \hat{P}\hat{Q} = \hat{Q}\hat{P}$. One could certainly put $\hat{Q}$ in front of $\hat{T}$ to form $e^{\hat{Q}\hat{T}}$, but we do not choose this option because it would complicate the algebra with terms like $\hat{Q}\hat{T}\hat{Q}\hat{T}$, which are obviously much more difficult to deal with (recall that $\hat{P}$ has de-excitation components). Rather, it is reasonable to set $|\Psi\rangle$ as above by recognizing $\hat{Q}e^{\hat{T}}|\Phi_0\rangle$ as the purely dynamical correlation space orthogonal to SUHF. Then, $|\Psi'\rangle$ is intermediate normalized (again, $\hat{P}|\Phi_0\rangle$ itself is normalized), and the corresponding t-amplitude equations are separated from the energy definition, by choosing $\langle\Phi_\mu|\hat{Q}^\dagger$ as the projection manifold

$$E_c = \langle\Phi_0|\hat{H}_N\hat{Q}e^{\hat{T}}|\Phi_0\rangle = \langle\Phi_0|\hat{H}_N\hat{P}e^{\hat{T}}|\Phi_0\rangle, \tag{11a}$$

$$\begin{aligned} r_\mu[\mathbf{t}] &:= \langle\Phi_\mu|\hat{Q}^\dagger\left(\hat{H}_N - E_c\right)\left(\hat{P} + \hat{Q}e^{\hat{T}}\right)|\Phi_0\rangle \\ &= \langle\Phi_\mu|\left(\hat{H}_N - E_c\right)\hat{P}e^{\hat{T}}|\Phi_0\rangle + \langle\Phi_\mu|\left(\hat{H}_N - E_c\right)\hat{P}|\Phi_0\rangle \\ &\quad\times\left(1 - \langle\Phi_0|\hat{P}e^{\hat{T}}|\Phi_0\rangle\right) \\ &= 0, \end{aligned} \tag{11b}$$

where we have used the fact that $\langle\Phi_0|\hat{H}_N\hat{P}|\Phi_0\rangle = 0$. In the intermediate normalized form, the redundancies in t_μ are directly mapped to those in $\hat{Q}|\Phi_\mu\rangle$, and the metric becomes

$$S_{\mu\nu} = \langle\Phi_\mu|\hat{Q}^\dagger\hat{Q}|\Phi_\nu\rangle = \langle\Phi_\mu|\hat{P}|\Phi_\nu\rangle - \langle\Phi_\mu|\hat{P}|\Phi_0\rangle\langle\Phi_0|\hat{P}|\Phi_\nu\rangle. \tag{12}$$

We note that Eq. (12) is not the only choice for the metric, and this is where the arbitrary nature of the ECC solution comes in Refs. 32, 33, and 58. The singles space is also redundantly present in the doubles space (so-called spectator excitations). The former can be projected out from the latter in a similar way as that presented above. This procedure changes the definition of the metric and gives different orthogonalized excitation operators, *vide infra*. In *ic*-MRCC, such a sequential projection of lower-excited functions removes some disconnected terms in the contraction between the metric and amplitudes. As a result, it partially recovers size-extensivity, known as core extensivity, i.e., the correct scaling of the correlation energy associated with the growth of inactive orbitals and electrons.[33,58] The full size-extensivity is perfectly retained if excitation operators are derived based on the generalized normal-ordering.[58–60] In our method, however, the full size-extensivity is particularly difficult, irrespective of the choice of metric, because of the symmetry projection operator, which destroys the full size-extensivity (see Sec. VI E). Hence, in the present work, we will employ Eq. (12), i.e., a mixture of singles and doubles. Our choice is therefore similar to that of Evangelista *et al.*[32] and version C of Hanauer and Köhn.[33] A comparison between different metric schemes for ECC will be investigated in future work.

C. Orthogonalization

We are now in a position to deal with the redundancies present in Eq. (11b). To remove the linear dependencies in t_μ, it is necessary to employ a set of orthogonalized excitation operators

$$\hat{\mathcal{E}}_\alpha = \sum_\mu \hat{E}_\mu X_{\mu\alpha}, \tag{13}$$

as suggested by other authors.[32,33,53,58,61–63] In analogy with these previous works, we adopt the canonical orthogonalization by using the singular-value-decomposition of $\mathbf{S}$. As $\mathbf{S}$ is positive-semidefinite, we have in general that

$$\sum_\nu S_{\mu\nu}U_{\nu\alpha} = U_{\mu\alpha}\lambda_\alpha, \tag{14a}$$

$$\sum_\nu S_{\mu\nu}U_{\nu\beta} = 0, \tag{14b}$$

where the subscript α is restricted to positive λ_α and β corresponds to zero singular values or those below some threshold η ($\lambda_\beta < \eta$). These spaces are called the range and null space, characterized as $\mathcal{R}(\mathbf{S})$ and $\mathcal{N}(\mathbf{S})$, whose dimensions are respectively $M_\mathcal{R}$ and $M_\mathcal{N} = M_{SD} - M_\mathcal{R}$. More often than not, $M_\mathcal{R} \gg M_\mathcal{N} > 0$. As several authors have pointed out,[32,63,64] in standard *ic*-MRCC methods, unlike in *ic*-MRCI,[65] the truncation threshold η must be carefully chosen because small values cause numerical instability while large values lead to degradation of results. However, we see no such difficulty in our method, and η can always be very small. We present some examples in Sec. VI F.

Using the singular values and vectors, the transformation matrix $\mathbf{X}$ is obtained as

$$X_{\mu\alpha} = U_{\mu\alpha}\lambda_\alpha^{-\frac{1}{2}}, \tag{15}$$

and one can produce an excitation basis

$$|\Xi_\alpha\rangle = \hat{\mathcal{E}}_\alpha|\Phi_0\rangle, \tag{16}$$

which is orthonormal *when* $\hat{\mathcal{Q}}$*-projected*,

$$\langle\Xi_{\alpha'}|\hat{\mathcal{Q}}|\Xi_\alpha\rangle = \delta_{\alpha\alpha'}. \tag{17}$$

Note that $\mathbf{U}_\alpha$ may not be defined uniquely because of possible degeneracies in λ_α. However, the final result is invariant with respect to the unitary rotation of degenerate singular vectors within the range (and the null space), as will be shown.

The orthonormalized cluster operator is obtained as

$$\hat{\tau} = \sum_\alpha^{M_\mathcal{R}} \tau_\alpha \hat{\mathcal{E}}_\alpha. \tag{18}$$

Clearly, the number of linearly independent parameters is only $M_\mathcal{R}$, and the same number of independent equations can be constructed with the orthonormal projection manifold $\langle\Xi_\alpha|\hat{\mathcal{Q}}^\dagger$. The resulting set of ECC equations is rewritten as

$$E_c = \langle\Phi_0|\hat{H}\hat{\mathcal{Q}}e^{\hat{\tau}}|\Phi_0\rangle, \tag{19}$$

$$\langle\Xi_\alpha|\hat{\mathcal{Q}}^\dagger\left(\hat{H} - E_c\right)\left(\hat{P} + \hat{\mathcal{Q}}e^{\hat{\tau}}\right)|\Phi_0\rangle = 0 \quad \forall\ \alpha \in \mathcal{R}(\mathbf{S}). \tag{20}$$

Solving Eqs. (19)–(20) gives unique E_c and $\boldsymbol{\tau}$.

In practice, we do not explicitly transform the excitation operators and amplitudes. Suppose that we have successfully obtained some $\boldsymbol{\tau}$ that satisfies Eq. (20). From Eq. (18), it is evident that there is the following one-to-one mapping between τ_α in the orthonormal basis and some $\tilde{t}_\mu$ in the original non-orthogonal basis:

$$\tilde{t}_\mu = \sum_\alpha^{M_\mathcal{R}} U_{\mu\alpha} s_\alpha^{-\frac{1}{2}} \tau_\alpha, \tag{21}$$

$$\tau_\alpha = s_\alpha^{\frac{1}{2}} \sum_\mu^{M_{\mathrm{SD}}} U_{\mu\alpha}^* \tilde{t}_\mu. \tag{22}$$

Hence, $\tilde{\mathbf{t}}$ is also uniquely determined, up to the choice of the metric $\mathbf{S}$. The chief difference between this particular $\tilde{\mathbf{t}}$ and other general $\mathbf{t}$ is that $\tilde{\mathbf{t}}$ is orthogonal to $\mathcal{N}(\mathbf{S})$, whereas $\mathbf{t}$ can spill into $\mathcal{N}(\mathbf{S})$, i.e.,

$$\tilde{\mathbf{t}} \perp \mathcal{N}(\mathbf{S}) \Leftrightarrow \tilde{\mathbf{t}} \in \mathcal{R}(\mathbf{S}), \tag{23}$$

$$\neg\ \mathbf{t} \perp \mathcal{N}(\mathbf{S}) \Leftrightarrow \mathbf{t} \in \mathcal{R}(\mathbf{S}) \oplus \mathcal{N}(\mathbf{S}). \tag{24}$$

This means that $\tilde{\mathbf{t}}$ is represented as a vector that is projected on to $\mathcal{N}(\mathbf{S})$ using the null-space projector $\boldsymbol{\mathcal{P}}$,

$$\tilde{t}_\mu = \sum_\nu^{M_{\mathrm{SD}}} \mathcal{P}_{\mu\nu} t_\nu, \tag{25}$$

where

$$\mathcal{P}_{\mu\nu} = \delta_{\mu\nu} - \sum_\beta^{M_\mathcal{N}} U_{\mu\beta} U_{\nu\beta}^*. \tag{26}$$

Substituting Eqs. (18) and (22) into Eq. (20), we have the proper residuals

$$\lambda_\alpha^{-\frac{1}{2}} \sum_\mu U_{\mu\alpha}^* r_\mu[\tilde{\mathbf{t}}] = 0, \tag{27}$$

where we remind the reader that $\mathbf{r}$ is the residual vector in the non-orthogonal basis and is defined in Eq. (11b).

As often $M_\mathcal{R} >> M_\mathcal{N}$, it is worth rewriting Eq. (27) in terms of $U_{\mu\beta}$. To do so, we multiply it by $S_{\nu\lambda}U_{\lambda\alpha}\lambda_\alpha^{-\frac{1}{2}}$ from the left and sum over α and λ to give

$$\tilde{r}_\nu[\tilde{\mathbf{t}}] \coloneqq \sum_{\mu\lambda} S_{\nu\lambda} S^+_{\lambda\mu} r_\mu[\tilde{\mathbf{t}}] = \sum_\mu \mathcal{P}_{\lambda\mu} r_\mu[\tilde{\mathbf{t}}] = 0, \tag{28}$$

where $\mathbf{S}^+$ is the Moore–Penrose pseudoinverse of $\mathbf{S}$. Both $\tilde{\mathbf{t}}$ and $\tilde{\mathbf{r}}$ are written in the same non-orthogonal basis restricted in $\mathcal{R}(\mathbf{S})$, and therefore, one can directly use the latter to perturbatively update the former. However, in this basis, the metric is not diagonal-dominant. In Sec. IV, we overcome this problem using a robust preconditioning scheme that is effective for non-diagonal-dominant metrics such as $\mathbf{S}$.

Finally, note that the scheme is invariant with respect to a unitary rotation between degenerate vectors in $U_{\mu\alpha}$ and in $U_{\mu\beta}$ because the null-space projection operator $\boldsymbol{\mathcal{P}}$ is uniquely defined in Eq. (26). Whether the method is unitary-invariant with respect to *orbital* rotations of $|\Phi_0\rangle$ is a separate question and will be discussed in Sec. III.

III. APPROXIMATE CCSD IN SPIN-PROJECTED MANIFOLD: TRUNCATION OF TAYLOR SERIES

While we have described how to manage the redundancies in the ECCSD parametrization, it is practically impossible to construct the exact residuals at polynomial cost when an exponential *ansatz* is associated with $\hat{P}$. Here, we consider how to approximate the *ansatz* with $\hat{P}$ intact.

By expanding $|\Psi\rangle$ (not intermediate normalized) in the Taylor series, we obtain

$$|\Psi\rangle = \hat{P}\Big(1 + \hat{T}_1 + \hat{T}_2 + \frac{1}{2}\hat{T}_1^2 + \frac{1}{6}\hat{T}_1^3 + \hat{T}_1\hat{T}_2 + \frac{1}{2}\hat{T}_2^2 + \frac{1}{2}\hat{T}_1^2\hat{T}_2 + \frac{1}{24}\hat{T}_1^4 + \cdots\Big)|\Phi_0\rangle. \tag{29}$$

As $\hat{H}$ is a two-body operator, at most quadruply excited determinants with respect to $|\Phi_0\rangle$ could interact with the bra projection $\langle\Phi_{ij}^{ab}|$ if $\hat{P}$ were absent. A similar relation also holds for the non-orthogonal case when the strings are normal-ordered.

First, let us write the projection operator as a numerical interaction,

$$\hat{P} \approx \sum_g^{N_{\mathrm{grid}}} w_g \hat{R}_g, \tag{30}$$

where g is the numerical grid for representing the Ω angle. The rotated creation and annihilation operators due to $\hat{R}_g$ are given by

$$\hat{b}^p(g) \equiv \hat{R}_g \hat{a}^p \hat{R}_g^{-1}. \tag{31}$$

For the explicit form of $\hat{R}_g$ as well as the simplification one can exploit for spin-projection, see Ref. 37. Therefore, the rotated determinant $|\Phi_0(g)\rangle \equiv \hat{R}_g|\Phi_0\rangle$ comprises $\hat{b}^i(g)$ instead of $\hat{a}^i$. Wick's theorem then allows us to write the rotated $\hat{T}$ operators as[43,44]

$$\hat{R}_g\hat{T}_1\hat{R}_g^{-1} = \sum_{kc} t_k^c\left(\mathcal{W}_k^c + \{\hat{b}_k^c\}_g\right), \tag{32}$$

$$\hat{R}_g\hat{T}_2\hat{R}_g^{-1} = \sum_{klcdf} t_{kl}^{cd}\left(\frac{1}{2}\mathcal{W}_k^c\mathcal{W}_l^d + \mathcal{W}_l^d\{\hat{b}_k^c\}_g + \frac{1}{4}\{\hat{b}_{kl}^{cd}\}_g\right), \tag{33}$$

where we have introduced the concept of normal-ordering $\langle\Phi|\{\dots\}_g|\Phi_0(g)\rangle = 0$ and the one-body contraction tensor $\boldsymbol{\mathcal{W}}$, composed of only nonzero contractions,

$$\mathcal{W}_i^j = \frac{\langle\Phi_0|\hat{a}^j\hat{b}_i|\Phi_0(g)\rangle}{\langle\Phi_0|\Phi_0(g)\rangle}, \tag{34a}$$

$$\mathcal{W}_a^i = \frac{\langle\Phi_0|\hat{a}^i\hat{a}_a|\Phi_0(g)\rangle}{\langle\Phi_0|\Phi_0(g)\rangle}, \tag{34b}$$

$$\mathcal{W}_i^a = \frac{\langle\Phi_0|\hat{b}^a\hat{b}_i|\Phi_0(g)\rangle}{\langle\Phi_0|\Phi_0(g)\rangle}, \tag{34c}$$

$$\mathcal{W}_a^b = \frac{\langle\Phi_0|\hat{a}_a\hat{b}^b|\Phi_0(g)\rangle}{\langle\Phi_0|\Phi_0(g)\rangle}, \tag{34d}$$

besides the trivial Kronecker delta,

$$\delta_{ij} = \frac{\langle\Phi_0|\hat{a}^i\hat{a}_j|\Phi_0(g)\rangle}{\langle\Phi_0|\Phi_0(g)\rangle} = \frac{\langle\Phi_0|\hat{b}^i\hat{b}_j|\Phi_0(g)\rangle}{\langle\Phi_0|\Phi_0(g)\rangle}, \tag{34e}$$

$$\delta_{ab} = \frac{\langle\Phi_0|\hat{a}_a\hat{a}^b|\Phi_0(g)\rangle}{\langle\Phi_0|\Phi_0(g)\rangle} = \frac{\langle\Phi_0|\hat{b}_a\hat{b}^b|\Phi_0(g)\rangle}{\langle\Phi_0|\Phi_0(g)\rangle}. \tag{34f}$$

Similar to the orthogonal case, the normal-ordering in the non-orthogonal Wick theorem also requires connectivity between the products

$$\{\cdots\}_g\{\cdots\}_g = \sum^{\text{all}}\{\overbrace{\cdots}\,\overbrace{\}_g\{\cdots}\}_g. \tag{35}$$

The essential idea is then to write all operators, e.g., $\hat{T}$ and $\hat{H}$, in terms of normal-ordered strings. Using the generalized Wick theorem for normal-ordered products, we write $|\Psi\rangle$ as a pseudo-CI

$$|\Psi\rangle = \sum_g w_g\Bigg(\omega_0 + \sum_{kc}\omega_k^c\{\hat{b}_k^c\}_g + \frac{1}{(2!)^2}\sum_{klcd}\omega_{kl}^{cd}\{\hat{b}_{kl}^{cd}\}_g + \frac{1}{(3!)^2} \times\sum_{\substack{klm\\cde}}\omega_{klm}^{cde}\{\hat{b}_{klm}^{cde}\}_g + \frac{1}{(4!)^2}\sum_{\substack{klmn\\cdef}}\omega_{klmn}^{cdef}\{\hat{b}_{klmn}^{cdef}\}_g + \cdots\Bigg)|\Phi_0(g)\rangle, \tag{36}$$

where we have introduced some effective coefficients $\boldsymbol{\omega}$, which are anti-symmetrized. They depend on both $\mathbf{t}$ and g, whose dependencies are suppressed for the sake of simplicity. As the Hamiltonian is a two-body operator, we need only up to normal-ordered doubles for the energy and overlap matrix elements and normal-ordered quadruples for solving the ECCSD equation. Quintuple and higher excitations play no role in the normal-ordered representation. It should be clear that any projected wave functions can be written in this form. For example, ECISD is truncated at doubles in Eq. (36), and there are no higher excitations. Therefore, the previously developed ECISD code can be directly exploited to compute matrix elements up to normal-ordered double excitations for ECCSD.[44] One can also exactly compute matrix elements such as $\langle\Phi_{ij}^{ab}|\hat{H}\{\hat{b}_{klmn}^{cdef}\}_g|\Phi_0(g)\rangle$ at polynomial cost (see the supplementary material). We should emphasize that it is the t-amplitudes (or $\tilde{\mathbf{t}}$) that are subject to optimization, but not $\boldsymbol{\omega}$, as we use the same $\hat{T}$ for all gauge angles.

So far, we have made no approximation except that $\hat{T}$ is truncated at doubles; Eq. (36) is still equivalent to Eq. (29). However, this does not pave the way to solving Eq. (11) [or (28)], as all the complexities arising from $\hat{P}$ are shifted to the evaluation of the effective coefficients $\boldsymbol{\omega}$. These are represented as highly complicated tensor contractions of several orders of $\mathbf{t}$ and $\boldsymbol{\mathcal{W}}$, and the computational cost for evaluating them is formally $\mathcal{O}(N!)$, where N is some measure of system size. Hence, some approximation is required to arrive at a tractable method.

In passing, we should note that, as in unitary CC and most MRCC methods, the non-terminating equation may be truncated at the second-order of the Baker–Campbell–Hausdorff expansion of the *projected* Hamiltonian, i.e., $e^{-\hat{T}}\hat{P}\hat{H}\hat{P}e^{\hat{T}} = \hat{P}\hat{H}\hat{P} + [\hat{P}\hat{H}\hat{P},\hat{T}] + \frac{1}{2}[[\hat{P}\hat{H}\hat{P},\hat{T}],\hat{T}] + \cdots$. However, the resulting equations comprise a mixture of symmetry-projected and symmetry-broken components, and so the t-amplitudes rarely converge. Therefore, this truncation scheme will not be discussed any further.

Instead, we consider truncating the Taylor expansion in Eq. (29) to include excitations up to quadruples. The size-extensivity is obviously lost by this approximation, but we should emphasize that this is not an important issue, as it cannot be retained anyway because of the introduction of the spin-projection operator. The wave function *ansatz* is given by

$$|\Psi_{\text{EACCSD}}\rangle = \hat{P}\Big(1 + \hat{T}_1 + \hat{T}_2 + \frac{1}{2}\hat{T}_1^2 + \frac{1}{6}\hat{T}_1^3 + \hat{T}_1\hat{T}_2 + \frac{1}{2}\hat{T}_2^2 + \frac{1}{2}\hat{T}_1^2\hat{T}_2 + \frac{1}{24}\hat{T}_1^4\Big)|\Phi_0\rangle, \tag{37}$$

and hereinafter, we call this method spin-extended approximate CCSD (EACCSD). This specific choice of truncation is made based on several observations. First, $|\Psi_{\text{EACCSD}}\rangle$ is guaranteed to reduce to SRCCSD when $\hat{P} = 1$. Second, the first-order interacting space (FOIS) of projected doubles for ket states ($\langle\Phi_{ij}^{ab}|$) is completely spanned by the SUHF reference, projected singles, $\cdots$, quadruples, and thus the EACCSD problem covers the Hilbert space essential for exact ECCSD (see Appendix A). Higher excitations such as $\hat{T}_2^n$ components with $n > 2$ can play a role in re-optimizing the cluster amplitudes; while it is possible to treat these terms explicitly, the inclusion of $\hat{T}_2^3$ increases the computational cost from $\mathcal{O}(N^6)$ to $\mathcal{O}(N^8)$. Additionally, provided that SUHF is a good reference, the cluster amplitudes are expected to be sufficiently small, and it is plausible that the effect of $\hat{T}_2^3$ will be insignificant. Hence, it is advisable to neglect these terms for our approximation to be feasible. Finally, the size-extensive component of the exact ECCSD correlation energy will eventually become that of UCCSD, and therefore, the

role of T_2^3 and higher terms should diminish with system size.

With the EACCSD *ansatz*, in the normal-ordered pseudo-CI form [i.e., Eq. (36)], the triples and quadruples amplitudes take the following forms:

$$\frac{1}{(3!)^2}\omega_{klm}^{cde} = \frac{1}{4}t_{kl}^{cd}t_m^e + \frac{1}{6}t_k^c t_l^d t_m^e + \sum_{nf}\Big(\frac{1}{4}t_{kl}^{cd}t_m^e t_n^f + \frac{1}{6}t_k^c t_l^d t_m^e t_n^f + \frac{1}{4}t_{kl}^{cd}t_{mn}^{ef} + \frac{1}{4}t_{kn}^{cd}t_{lm}^{ef} + \frac{1}{2}t_k^c t_l^d t_{mn}^{ef} - \frac{1}{2}t_{kn}^{cd}t_l^f t_m^e - \frac{1}{2}t_{kl}^{cf}t_n^d t_m^e - \frac{1}{4}t_{kl}^{cd}t_n^e t_m^f - \frac{1}{2}t_k^c t_l^d t_m^f t_n^e\Big)\mathcal{W}_n^f, \quad (38)$$

$$\frac{1}{(4!)^2}\omega_{klmn}^{cdef} = \frac{1}{32}t_{kl}^{cd}t_{mn}^{ef} + \frac{1}{8}t_{kl}^{cd}t_m^e t_n^f + \frac{1}{24}t_k^c t_l^d t_m^e t_n^f. \quad (39)$$

Other ω coefficients, as well as the matrix elements required to implement the EACCSD methods, are summarized in the supplementary material.

Finally, we briefly mention the orbital invariance characteristic of EACCSD. It is well known that most MRCC approaches, including those based on the SR formalism[66–69] and all state-specific MRCC methods based on the Jeziorski–Monkhorst *ansatz*,[17–27] are not orbital-invariant with respect to active space rotations, which is considered a serious drawback.[30] Achieving invariance requires an unbiased treatment of orbitals in each subspace as well as the use of a single vacuum. On the other hand, as in traditional CC, EACCSD is strictly orbital-invariant because all the tensor contractions appear in the energy and residuals run over all the orbitals in the given subspace, either occupied or virtual spaces of $|\Phi_0\rangle$. A proof is given in Appendix C.

IV. SOLVER FOR THE APPROXIMATE ECCSD EQUATIONS

Having outlined our approach, we now describe how to solve the EACCSD equation in practice. For linearized methods such as ECISD and ELCCSD, one can resort to the Davidson diagonalization approach,[70] as the resulting problems are cast as a generalized eigenvalue problem (with a dressed Hamiltonian).[44,45] However, the EACCSD equation becomes a non-Hermitian and nonlinear problem, which is symbolically written as

$$\tilde{r}_\mu[\tilde{\mathbf{t}}] \equiv \sum_{\nu\lambda}\mathcal{P}_{\mu\nu}\Big(\mathcal{H}_{\nu\lambda}[\tilde{\mathbf{t}}] - E_c\mathcal{S}_{\nu\lambda}[\tilde{\mathbf{t}}]\Big)\tilde{t}_\lambda = 0, \quad (40)$$

where $\mathcal{H}$ and $\mathcal{S}$ are the effective Hamiltonian and overlap of ECCSD, both of which depend on $\tilde{\mathbf{t}}$. This equation must be solved iteratively. To do so, in Eq. (40), we separate $\mathcal{H}$ and $\mathcal{S}$ into their dominant parts $\mathcal{H}^0$ and $\mathcal{S}^0$ and the remaining parts. They do not have to be diagonal, but it may be assumed that $\mathcal{H}^0 - E_c\mathcal{S}^0$ is relatively easy to invert. For simplicity, we assume $\mathcal{H}^0$ and $\mathcal{S}^0$ contain only the linear ECISD components, which are considered dominant, because the contributions of the quadratic and higher-order terms are of size-consistent correction and are expected to be several orders of magnitude smaller than the linear part.

First-order perturbation theory suggests

$$\begin{aligned}\tilde{\mathbf{r}}[\tilde{\mathbf{t}}+\Delta\tilde{\mathbf{t}}] &\approx \mathcal{P}(\mathcal{H}-E_c\mathcal{S})\tilde{\mathbf{t}} + \mathcal{P}\Big(\mathcal{H}^0 - E_c\mathcal{S}^0\Big)\Delta\tilde{\mathbf{t}}\\ &\approx \mathcal{P}\mathbf{r}[\tilde{\mathbf{t}}] + \Big(\mathcal{H}^0 - E_c\mathcal{S}^0\Big)\mathcal{P}\Delta\tilde{\mathbf{t}}\\ &= 0, \end{aligned} \quad (41)$$

where we have assumed $\left[\mathcal{H}^0 - E_c\mathcal{S}^0, \mathcal{P}\right] \approx \mathbf{0}$. This is a reasonable assumption because we have set $\mathcal{H}^0 \approx \mathbf{H}^{\mathrm{ECISD}}$, where $\mathbf{H}^{\mathrm{ECISD}}$ is the ECISD Hamiltonian, and the variational nature of ECISD necessarily insists that $\left[\mathbf{H}^{\mathrm{ECISD}}, \mathcal{P}\right] = \mathbf{0}$. In other words, $\mathbf{H}^{\mathrm{ECISD}}$ naturally shares the same null-space as $\mathbf{S}$; otherwise, the ECISD correlation energy would become negative infinity, which is unphysical. The same reasoning applies to $\mathcal{S}^0$ (indeed, in what follows, we choose it to be $\mathbf{S}$, which certainly satisfies the commutativity). Then, noting that $\mathcal{P}^2 = \mathcal{P}$, we obtain

$$\Delta\tilde{\mathbf{t}} = \Big(\mathcal{H}^0 - E_c\mathcal{S}^0\Big)^{-1}\mathcal{P}\mathbf{r}[\tilde{\mathbf{t}}] \approx \mathcal{P}\Big(\mathcal{H}^0 - E_c\mathcal{S}^0\Big)^{-1}\mathbf{r}[\tilde{\mathbf{t}}]. \quad (42)$$

This allows us to use $\mathbf{r}$ instead of $\tilde{\mathbf{r}}$. Our numerical tests suggest that the use of $\tilde{\mathbf{r}}$ in place of $\mathbf{r}$ makes no difference; both become zero at convergence.

In practice, the $(k+1)$th amplitudes $\tilde{t}_\mu^{(k+1)} = \tilde{t}_\mu^{(k)} + \Delta\tilde{t}_\mu^{(k)}$ are extrapolated to minimize the norm of the residual vector $||\tilde{\mathbf{r}}^{(k+1)}||$ by DIIS (assuming that $\tilde{\mathbf{r}}$ is approximately a linear function of $\tilde{\mathbf{t}}$). Then, it is just a matter of an appropriate choice for $\mathcal{H}^0$ and $\mathcal{S}^0$ to carry out the inversion that appears in Eq. (42).

The simplest approach is diagonal preconditioning, whereby one takes $\mathcal{H}_{\mu\nu}^0 = \delta_{\mu\nu}H_{\mu\nu}^{\mathrm{ECISD}}$ and $\mathcal{S}_{\mu\nu}^0 = \delta_{\mu\nu}S_{\mu\nu}$. This approach is usually stable and easy to implement but suffers from slow convergence because the off-diagonal couplings can make a significant contribution in ECISD and ECCSD. This is a considerable drawback because the computational cost of constructing $\mathbf{r}^{(k)}$ at each cycle scales as $\mathcal{O}(N^6)$, and we need to reduce the number of these steps as much as possible.

As is well known, appropriate preconditioning is important for fast convergence in iterative methods. In this paper, we follow our previous method with some further modification.[44] Let us write the Hamiltonian as

$$\hat{H}_N = (E_g - E_{\mathrm{SUHF}}) + \hat{F}_{N,g} + \hat{V}_{N,g}, \quad (43)$$

where E_g, $\hat{F}_{N,g}$, and $\hat{V}_{N,g}$ are the normal-ordered zero-, one-, and two-body operators at each projection-grid point,

$$E_g = \frac{\langle\Phi_0|\hat{H}|\Phi_0(g)\rangle}{\langle\Phi_0|\Phi_0(g)\rangle}, \quad (44)$$

$$\hat{F}_{N,g} = \sum_{pq} f_{pq}(g)\{\hat{a}_q^p\}_g, \quad (45)$$

$$\hat{V}_{N,g} = \frac{1}{4}\sum_{pqrs}\langle pq||rs\rangle\{\hat{a}_{rs}^{pq}\}_g, \quad (46)$$

with the transition Fock matrix $\mathbf{f}(g)$ defined as

$$f_{pq}(g) = h_{pq} + \sum_{rs}\langle pr||qs\rangle\frac{\langle\Phi_0|\hat{a}_s^r|\Phi_0(g)\rangle}{\langle\Phi_0|\Phi_0(g)\rangle}. \quad (47)$$

As the contribution of $\hat{V}_{N,g}$ to the ECISD Hamiltonian elements is generally weak, we ignore this part for $\mathcal{H}^0$. Integrating over grid points, we define the effective Fock-like matrix as follows:

$$\mathcal{F}_{\mu\nu} = \sum_g^{N_{\rm grid}} w_g \langle \Phi_\mu | \Big(\big(E_g - E_{\rm SUHF} \big) + \hat{F}_{N,g} \Big) | \Phi_\nu(g) \rangle. \quad (48)$$

Although $\mathcal{F}$ may not be spin-adapted, it holds a one-particle picture and is a good approximation to the ECISD Hamiltonian matrix, which is all we need for our purpose. The zeroth order $\mathcal{H}^0$ then has the following form:

$$\mathcal{H}^0_{\mu\nu} = \mathcal{F}_{\mu\nu} - \mathcal{F}_{\mu 0} S_{0\nu} - S_{\mu 0} \mathcal{F}_{0\nu}, \quad (49)$$

where the last two terms come from the $\hat{Q}$ projection (the subscript 0 indicates the reference determinant). For $\mathcal{S}^0$, it is natural to take the whole $\mathbf{S}$ because of its structural resemblance to $\mathcal{H}^0$. These definitions exactly parallel the adoption of the orbital energy differences for preconditioning in SRCC with the canonical orbitals. Unfortunately, there is no orbital basis in which $\mathcal{H}^0$ is diagonal for symmetry-projected methods, and hence the inversion in Eq. (42) requires some attention.

By writing

$$\Delta \tilde{t}^{(k)}_\mu = \sum_\nu \mathcal{P}_{\mu\nu} \Delta t^{(k)}_\mu, \quad (50)$$

we solve the linear equation

$$\rho^{(k)}_\mu := \sum_\nu \Big(\mathcal{H}^0_{\mu\nu} - E^{(k)}_{\rm c} S_{\mu\nu} \Big) \Delta t^{(k)}_\nu + r^{(k)}_\mu = 0, \quad (51)$$

at each macro iteration k. Eq. (51) is solved with the diagonal preconditioning scheme (and with DIIS). For the sake of clarity, the mth micro-iteration is performed as

$$\Delta t^{(k[m+1])}_\nu = \Delta t^{(k[m])}_\nu - \frac{\rho^{(k[m])}_\nu}{\mathcal{H}^0_{\nu\nu} - E^{(k)}_{\rm c} S_{\nu\nu}}. \quad (52)$$

The superscript $(k[m])$ denotes the mth micro-cycle of the kth macro-cycle. Once $\Delta\mathbf{t}^{(k)}$ is obtained, we perform the null-space projection given by Eq. (50).

Note that the computational scaling for solving Eq. (51) is $\mathcal{O}(o^2v^3)$, which comes from the contraction of some guess vector $\Delta\mathbf{t}^{(k[m])}$ against $\mathcal{H}^0$. While solving Eq. (51) significantly improves the convergence performance of the macro-iterations in Eq. (40), a large number of micro-iterations are needed to converge Eq. (52) because $\mathcal{H}^0$ has essentially the same structure as the ECISD Hamiltonian (i.e., not diagonal-dominant). Hence, we previously concluded that the two-step scheme is not likely to be efficient in terms of the overall computational performance. However, it is noteworthy that the accuracy of $\Delta\mathbf{t}$ is relatively unimportant in the first few macro-cycles because $||\tilde{\mathbf{r}}^{(k)}||$ itself is large. We can therefore *approximate* $\Delta\mathbf{t}$ in such a way that Eq. (51) is only solved accurately enough to qualitatively "scan" the structure of $\mathcal{H}^0 - E^{(k)}_{\rm c}\mathbf{S}$, which is all that is needed for preconditioning. By contrast, when we are close to convergence, $||\tilde{\mathbf{r}}^{(k)}||$ is small and so is $\|\Delta\mathbf{t}^{(k)}\|$, thus necessitating an accurate solution for Eq. (51). These requirements can be simultaneously satisfied by proposing the following adaptive criterion for $\boldsymbol{\rho}^{(k)}$:

1 : Initialize $\tilde{t}^{(1)}_\mu = 0$
Macro Iteration
2 : **Do** $k = 1$, $k_{\rm max}$
3 : $E^{(k)}_{\rm c} = \mathcal{H}^{(k)}_{0\nu} \tilde{t}^{(k)}_\nu$
4 : $\sigma^{(k)}_0 = 1 - \mathcal{S}^{(k)}_{0\nu} \tilde{t}^{(k)}_\nu$
5 : $r^{(k)}_\mu = \Big(\mathcal{H}^{(k)}_{\mu\nu} - E^{(k)}_c \mathcal{S}^{(k)}_{\mu\nu} \Big) \tilde{t}^{(k)}_\nu + \sigma^{(k)}_0 \Big(\mathcal{H}^{(k)}_{\mu 0} - E^{(k)}_c \mathcal{S}^{(k)}_{\mu 0} \Big)$
6 : Null-space projection: $\tilde{\mathbf{r}}^{(k)} \leftarrow \boldsymbol{\mathcal{P}} \mathbf{r}^{(k)}$
7 : **If** $||\tilde{\mathbf{r}}^{(k)}|| < \epsilon_{\rm thres}$ **then**
8 : **Exit** on Convergence
9 : **end if**
10 : Extrapolate $\tilde{\mathbf{t}}^{(k)}$ to minimize $||\tilde{\mathbf{r}}^{(k)}||$ by DIIS
Perturbative update of $\tilde{\mathbf{t}}$
11 : Initialize $\Delta\mathbf{t}^{(k[1])} = \mathbf{0}$
12 : **Do** $m = 1$, $m_{\rm max}$
13 : $\rho^{(k[m])}_\mu = \Big(\mathcal{H} - E^{(k)}_c \mathbf{S} \Big)_{\mu\nu} \Delta t^{(k[m])}_\nu + r^{(k)}_\mu$
14 : **If** $||\boldsymbol{\rho}^{(k[m])}|| < \delta_{\rm Pert} ||\tilde{\mathbf{r}}^{(k)}||$ **then**
15 : Set $\Delta\mathbf{t}^{(k)} \leftarrow \Delta\mathbf{t}^{(k[m])}$; **Exit**
16 : **end if**
17 : Extrapolate $\Delta\mathbf{t}^{(k[m])}$ to minimize $||\boldsymbol{\rho}^{(k[m])}||$ by DIIS
18 : $\Delta t^{(k[m+1])}_\mu = \Delta t^{(k[m])}_\mu - \dfrac{\rho^{(k[m])}_\mu}{\mathcal{H}^0_{\mu\mu} - E^{(k)}_c S_{\mu\mu}}$
19 : **end do**
20 : $\mathbf{t}^{(k+1)} = \mathbf{t}^{(k)} + \Delta\mathbf{t}^{(k)}$
21 : Null-space projection: $\tilde{\mathbf{t}}^{(k)} \leftarrow \boldsymbol{\mathcal{P}} \mathbf{t}^{(k)}$
22 : **end do**

FIG. 2. Pseudo code of $\mathbf{t}$ with adaptive preconditioning.

$$||\boldsymbol{\rho}^{(k[m])}|| < \delta_{\rm Pert} \, ||\tilde{\mathbf{r}}^{(k)}||, \quad (53)$$

where $\delta_{\rm Pert}$ is some loose threshold. We call this the adaptive Fock preconditioning scheme. The pseudo-code of our solver is summarized in Fig. 2.

V. COMPUTATIONAL SCALING

The proposed method retains the formal scaling order of CCSD, $\mathcal{O}(N^6)$, at each grid point. By factoring the contraction terms, for the macro-steps, the computational cost of EACCSD formally scales as $\mathcal{O}(\frac{15}{2}o^3v^3 + \frac{9}{8}o^2v^4)$ in our current implementation (where o and v are the numbers of occupied and virtual orbitals), whereas that of ECISD scales as $\mathcal{O}(o^3v^3 + \frac{1}{8}o^2v^4)$. The additional $\mathcal{O}(o^2v^4)$ burden comes from the 6th term of Eq. (38), which describes the entanglement between two T_2 amplitudes through $\mathcal{W}^f_n$, and this is likely to be the limiting step for large basis calculations. We note that, if the projection operator is absent, $\mathcal{W}^q_p \rightarrow \delta_{pq}$ and there is no entanglement between t-amplitudes.

For the micro-steps of the adaptive Fock preconditioning scheme, the necessary contraction $\mathcal{H}^0\Delta\mathbf{t}$ has a computational cost that scales as $\mathcal{O}(3o^3v^2 + 3o^2v^3)$, whereas the diagonal preconditioning requires only $\mathcal{O}(o^2v^2)$.

VI. RESULTS AND DISCUSSION

In the following, we report several EACCSD calculations. All of these calculations used the SUHF orbitals, and the t-amplitudes were determined by solving Eq. (40) perturbatively with DIIS, as outlined in Sec. IV. For spin-extended methods, frozen orbitals were spin-adapted and doubly occupied to ensure the spin-symmetry of a wave function in post-SUHF calculations. Finally, we set $\eta = 10^{-12}$ and $N_{\mathrm{grid}} = 4$ throughout.

A. Unitary invariance and uniqueness in correlation energy

It is important for a method to fulfill orbital invariance because otherwise energy and other properties depend on the choice of orbitals. Here, orbital invariance is characterized as the nature whereby the computed energy is unchanged when orbitals are unitarily transformed within each orbital subspace. A mathematical proof for the orbital invariance of EACCSD is given in Appendix C. In this section, we present numerical examples. Furthermore, it is also our purpose to present a numerical proof that the removal of the linear dependencies ($\mathcal{N}(\mathbf{S})$) is essential for a unique solution. We deem it important to provide these results before discussing the performance of EACCSD in more detail.

The orbital invariance of EACCSD can be most easily checked by employing two different orbital sets, namely (a) semi-canonical[71,72] and (b) corresponding-pair orbitals[73,74] of $|\Phi\rangle$. In addition to the standard initial guess $\tilde{\mathbf{t}}^{(1)} = \mathbf{t}^{(1)} = \mathbf{0}$, UCCSD amplitudes were also tested as an initial guess, in which case the null-space projection was first performed as necessary. The diagonal preconditioning scheme was used with the maximum DIIS subspace size set to 10 and a very tight convergence criterion, $||\tilde{\mathbf{r}}|| < 10^{-8}$ ($||\mathbf{r}|| < 10^{-8}$ for the results without the null-space projection) to ensure high precision in the evaluated energy. We use the H_2O molecule at equilibrium ($R_{O-H} = 0.9929$ Å and $\angle HOH = 109.57°$) with a 6-31G basis as our showcase example.

Table I presents the computed energies under these different conditions. As expected, if the null-space projection is not carried out, i.e., if we allow the t-amplitudes to spill into $\mathcal{N}(\mathbf{S})$, the arbitrariness of the EACCSD energy is manifest. While all such energies are close to each other under a unitary transformation, using different guess amplitudes can yield significant fluctuations in energy and clearly illustrates the problem of linear dependencies. This arbitrariness makes the method completely useless because molecular properties such as vibrational frequencies are much more sensitive. We note that the orbital invariance of the EACCSD energy guarantees the definitive energy when evaluated with the UCCSD amplitudes in one shot, even if the null-space projection is not performed. It is the linearly dependent residual equations (7) that are improper and violate the uniqueness.

When the null-space of $\mathbf{S}$ is projected out of $\mathbf{t}$, the orbital invariance and uniqueness of energy are rigorously achieved, as is clear from Table I.

TABLE I. Computed energies of $H_2O + 76$*Hartree* using two different orbital sets: (a) semi-canonical and (b) corresponding-pair orbitals.

		Amplitude space	
Orb.	Initial guess	$\mathcal{R}(\mathbf{S}) \oplus \mathcal{N}(\mathbf{S})$	$\mathcal{R}(\mathbf{S})$
(a)	Zero	−0.120 872 594 0	−0.120 870 965 7
(b)	Zero	−0.120 875 918 1	−0.120 870 965 7
(a)	UCCSD	−0.120 122 363 0	−0.120 870 965 7
(b)	UCCSD	−0.120 122 454 4	−0.120 870 965 7

B. Convergence behavior

In this section, we demonstrate the validity of our iterative solver and the consequence of approximating $\hat{P}$ by neglecting some grid-dependent terms. These tests should identify an optimal adaptive criterion for micro-iterations. We use the same water molecule and basis set as in Sec. VI A, but similar behaviors are observed in other systems. The maximum size of the DIIS subspace was set to 10. All calculations used the corresponding-pair orbitals and gave the same energy at convergence.

Figure 3 illustrates the norm of the EACCSD residual vector $\tilde{r}_\mu$ at the kth iteration. The diagonal preconditioning scheme (plotted as "Diag") is stable, although it converges slowly. This behavior is similar to the Davidson diagonalization of ECISD with the same preconditioning scheme.[44]

If we iterate Eq. (52) for a few cycles at each macro-iteration, the convergence performance is substantially improved. The results are plotted as "Fock" in Fig. 3 for different adaptive criteria δ_{Pert}. While there is almost no gain with $\delta_{\mathrm{Pert}} = 1$, the number of cycles to achieve $||\tilde{\mathbf{r}}|| < 10^{-6}$ decreases to 17, 11, and 10 for $\delta_{\mathrm{Pert}} = 0.5$, 0.1, and 0.01, respectively. In spite of a slight increase in the computational cost due to the micro-iterations, the total CPU time required to complete a calculation typically decreases by a factor of about 3–6 for $\delta_{\mathrm{Pert}} = 0.1$ compared to the diagonal preconditioning scheme. Table II summarizes the total number of micro- and macro-iterations required to reach the final convergence $||\tilde{\mathbf{r}}|| < 10^{-6}$. The average number of micro-iterations required at each macro-step is 1.1, 4.3, 7.6, and 13.4 for $\delta_{\mathrm{Pert}} = 1.0$, 0.5, 0.1, and 0.01, respectively. These numbers do not significantly change

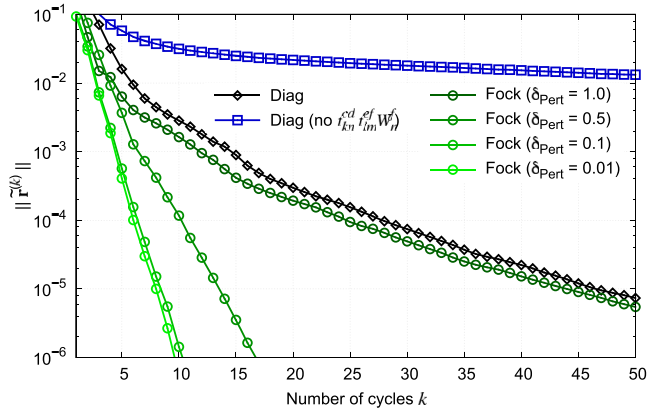

FIG. 3. Convergence behaviors of residual vector for N_2 at equilibrium. "Diag" and "Fock" denote diagonal preconditioning (practically equivalent to $\delta_{\mathrm{Pert}} > 1.0$) and adaptive micro-iteration schemes. "no $t_{kn}^{cd} t_{lm}^{ef} \mathcal{W}_n^f$" indicates that computationally expensive terms in Eq. (54) have been neglected.

TABLE II. Total number of micro- and macro-cycles required to reach the final convergence for the diagonal and adaptive Fock preconditioning schemes, with different thresholds δ_{Pert}.

	Diag	Fock			
δ_{Pert}	>1.0	1.0	0.5	0.1	0.01
Micro cycles ($\mathcal{O}(N^5)$)	0	80	73	84	134
Macro cycles ($\mathcal{O}(N^6)$)	72	70	17	11	10

for different molecules and basis sets. Based on these results, it is recommended that the adaptive Fock preconditioning scheme for perturbative updates uses $\delta_{\mathrm{Pert}} = 10^{-1}$. We additionally set the maximum number of micro-steps ($m_{\max}$ in Fig. 2) to 10.

In addition, the adaptive preconditioning is much more compatible with DIIS than is the diagonal preconditioning. DIIS is an extrapolation technique in which the new trial vector is expanded as a linear combination of the current and previous vectors $\tilde{\mathbf{t}}^{(k)}$. Therefore, the greater the differences between the $\tilde{\mathbf{t}}^{(k)}$, the more effective the DIIS technique, because it can explore a large space efficiently with a relatively small DIIS dimension. By contrast, if the previous vectors are all similar to one another, as in the diagonal preconditioning scheme, then the extrapolated vectors will be also similar, thus requiring a large subspace size to achieve good performance. This is demonstrated in Fig. 4, which shows the convergence behavior of both preconditioning schemes with DIIS dimensions of 10, 5, and 3 ($\delta_{\mathrm{Pert}} = 0.1$ for the adaptive Fock preconditioning scheme). Clearly, the performance of the adaptive preconditioning does not depend on the subspace size, whereas that of the diagonal preconditioning is prominently size-dependent.

As EACCSD is computationally more intensive than ECISD for each cycle, we wish to decrease its cost. The chief difference in cost between the two arises from the presence of the 6th term of $\boldsymbol{\omega}_3$ [Eq. (38)] in EACCSD. The Hamiltonian contraction with this term is

$$\frac{1}{4}\frac{\langle\Phi_{ij}^{ab}|\hat{H}\{\hat{b}_{klm}^{cde}\}_g|\Phi_0(g)\rangle}{\langle\Phi_0|\Phi_0(g)\rangle}t_{kn}^{cd}t_{lm}^{ef}\mathcal{W}_n^f = \mathcal{P}(ab)\mathcal{W}_l^i\mathcal{W}_m^j\mathcal{W}_b^d\bar{\mathcal{V}}_{ak}^{ec}t_{kn}^{cd}t_{ml}^{ef}\mathcal{W}_n^f + \cdots, \quad (54)$$

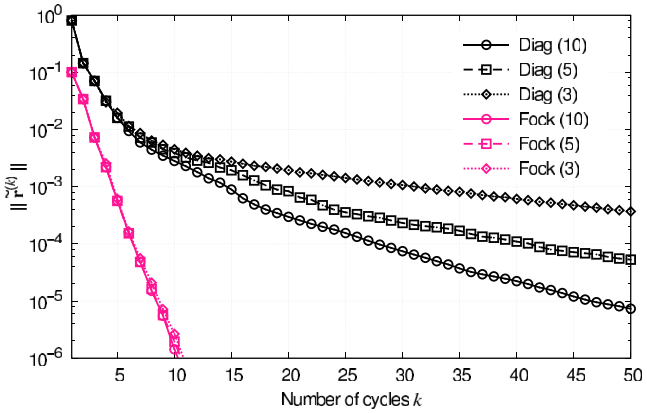

FIG. 4. Dependence of convergence behavior on the DIIS subspace size (numbers in parentheses).

where $\mathcal{P}(pq)$ is the antisymmetrizer that permutes p and q, and $\bar{\mathcal{V}}_{rs}^{pq}$ are the transformed two-electron integrals through $\boldsymbol{\mathcal{W}}$, which thus depend on g (see the supplementary material). This term evidently requires $\mathcal{O}(o^2v^4)$ operations to calculate the intermediate $X_{en}^{ad} = \sum \bar{\mathcal{V}}_{ak}^{ec}t_{kn}^{cd}$. It is tempting to avoid computing this term in order to reduce the computational cost. However, neglecting the term essentially amounts to approximating $\omega_{klm}^{cde}(g)$ in $\hat{R}_g\hat{T}^2$ at each grid point, which would cause numerical instability by losing the exactness of $\hat{P}$. To demonstrate this, we plot $||\tilde{\mathbf{r}}||$ without Eq. (54) in Fig. 3. Convergence is not achieved after hundreds of macro-cycles, strongly indicating the equations are ill-conditioned or inconsistent. While stable (but slow) convergence is occasionally observed in some other molecules, the H_2O result in Fig. 3 clearly implies that the approximation can introduce numerical difficulties.

C. H4 and H8 systems

The H4 and H8 systems are MR models that have previously been studied with several MRCC schemes, including the state-specific approaches such as BWCC[18–21] and MkCC[22–27] and the state-universal approach of Jeziorski and Monkhorst.[17] These systems illustrate the transition states of chemical reactions, as can be seen in the structures depicted in Fig. 5. In the H4 model, the bond distances between the nearest hydrogen atoms are fixed to 2 bohrs, and its structure changes from square to linear as the angle $\alpha\pi$ varies with α ranging from 0 to $\frac{1}{2}$. In the H8 model, four hydrogen molecules with $R_{\mathrm{H-H}} = 2$ bohrs are placed on a plane to form an octagonal configuration, and two of these molecules are horizontally deviated by α, reducing the spatial symmetry from D_{8h} to D_{2h}. In both cases, the high spatial symmetry at $\alpha = 0$ requires a two-configuration wave function for a proper description. Hence, all the genuine MRCC methods in previous studies employed a $(2e, 2o)$ CAS.[75] In this study, we conducted EACCSD calculations starting from SUHF, with the same basis sets employed in Ref. 75 (DZP and DZ for H4 and H8) to enable direct comparisons.

Figures 6 and 7 present the energy errors from FCI in mHartree as a function of α for the H4 and H8 systems, respectively. Table III lists the non-parallelity errors (NPE) as the difference between the maximum and minimum errors. In both systems, the SRCCSD energy largely deviates from the exact energy around $\alpha = 0$, as it fails to capture the strong static correlation accurately. All the MRCCSD methods improve the description of both the static and dynamic correlation effects reasonably well, except for SUCCSD in the H4 case, which tends to overestimate the correlation energy (by ~3 mHartree) as α increases. When triples are included, MRCCSDT methods generically reduce the errors of MRCCSD. However, Evangelista *et al.*[75] showed that they are much less accurate than SRCCSDT methods when a system has little MR character (i.e., large α region), although this is not shown in the present figures.

For these small systems, ECISD turns out to be more accurate than MRCCSD, but comparing the two systems shows that its accuracy deteriorates as the number of electrons increases from H4 to H8 because of the lack of size-consistency and

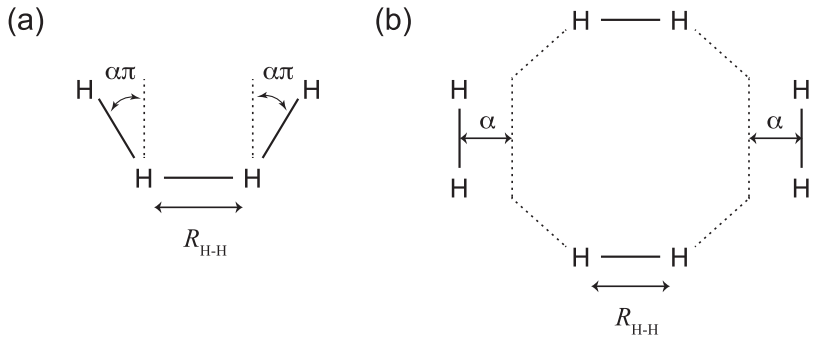

FIG. 5. Structures of (a) H4 and (b) H8 systems. The bond distance of each hydrogen molecule is R_{H-H} = 2 bohrs.

size-extensivity. Although ECISD with Davidson's size-consistent correction[76,77] (we used the variant proposed by Pople[78]), i.e., ECISD + Q, and ELCCSD are designed to fix this issue, they are both significantly less accurate than ECISD. However, a substantial improvement over these methods can be attained with EACCSD, with NPEs of only 66 and 76 μHartree for H4 and H8, respectively. As shown in Table III, the NPEs of BWCCSD (644 and 1906 μHartree) and MkCCSD (677 and 2019 μHartree) are one order of magnitude larger than those of EACCSD. It is also intriguing that EACCSD is much more accurate than MRCCSDT. However, we should point out that the reference wave functions are different for MRCC and EACCSD; while the former uses CASSCF (2*e*, 2*o*), the latter uses SUHF, which is more flexible. To faithfully evaluate the *dynamical* correlation effect that EACCSD captures, we also performed EACCSD calculations with a CASSCF (2*e*, 2*o*) reference, which is a subset of SUHF. In this case, the NPEs of EACCSD for H4 and H8 drastically increase to 666 and 1513 μHartree, which seem more reasonable when compared with the values of genuine MRCCSD. This result indicates that these systems are inherently more MR than a two-configuration description, and an active space of (2*e*, 2*o*) may not be appropriate. We should stress that the choice of orbitals, i.e., whether they are obtained by SUHF or CASSCF (2*e*, 2*o*), does not influence the computational effort in EACCSD. Occasionally, the former may require slightly more rotational grid points N_{grid} to carry out the numerical integration of spin-projection accurately, but this results in the increase in the computational cost that scales only linearly with N_{grid}. Therefore, there is no exponential increase in the cost when enlarging the "active" space of SUHF, a superiority of our method to other MRCC approaches tested here.

D. Ozone

The ozone is a challenging system for many SR methods. Its MR character requires descriptions of both dynamical and static correlation effects for accurate geometry and vibrational frequencies. Recently, we carried out calculations on this molecule to evaluate the performance of SUHF and ECISD analytical gradients.[79] This work revealed the questionable applicability of SUHF; while the bond length is drastically improved over RHF, some of the vibrational frequencies are severely underestimated. ECISD, on the other hand, yielded very similar results to MRCISD. However, compared to both traditional CCSD and MR-CCSD, ECISD and MRCISD slightly overestimated the vibrational frequencies. We therefore concluded that the overestimation can be attributed to the size-inconsistent error associated with these methods. However, the basis set used in the previous work was of double zeta quality, which is usually insufficient to compare results with experiments. Moreover, ECEPA methods tend to produce significant overcorrelation compared with ECISD +Q for this system.

It is therefore interesting to revisit the ozone molecule and learn the potential that EACCSD has to offer for molecular properties. For this purpose, we used cc-pVTZ.

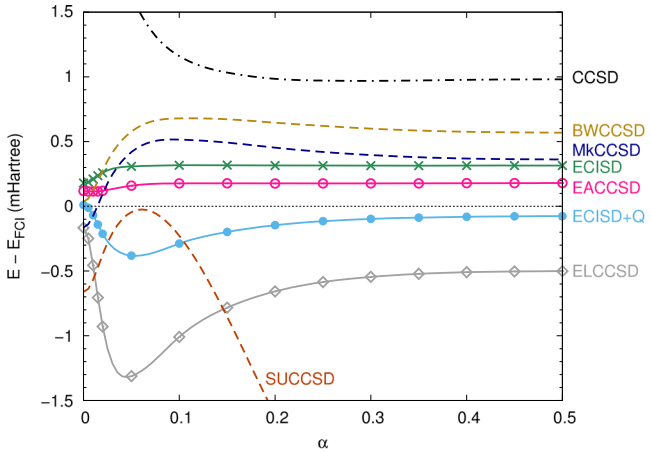

FIG. 6. Comparison of MR methods for the H4 system.

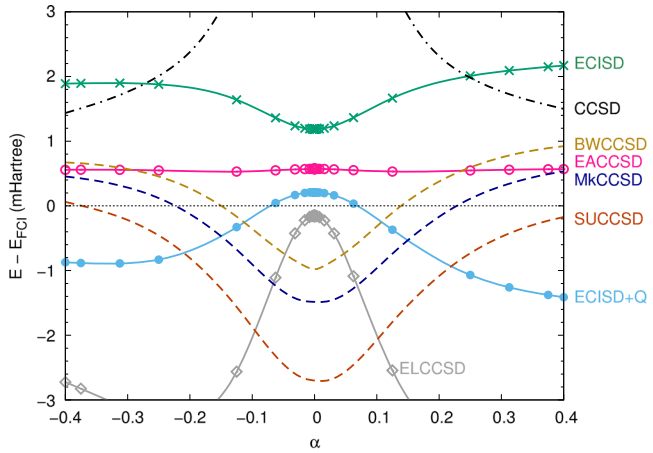

FIG. 7. Comparison of MR methods for the H8 system.

TABLE III. Non-parallelity errors with respect to FCI energy.

	NPE (μHartree)	
Method	H4	H8
SUHF	10 308	16 738
ECISD	138	982
ECISD+Q	392	1 622
ELCCSD	1 148	3 698
EACCSD	66	76
EACCSD[a]	666	1 513
BWCCSD	643	1 906
MkCCSD	677	2 019
SUCCSD	2 998	2 768
BWCCSDT	139	716
MkCCSDT	107	330
SUCCSDT	2 759	226

[a]CASSCF (2*e*, 2*o*) as the reference.

Although the size of the basis set may still not be large enough to allow a direct comparison against the experimental values, this moderate-size basis set has been extensively employed in many MRCC ozone studies and therefore should point out the relative accuracy of EACCSD in comparison with other methods.[29,33,80] As no analytical derivatives are yet available in EACCSD, we computed the gradients and Hessian using a finite difference scheme. The geometry optimization was considered to have converged when all the gradients in the Cartesian coordinates are less than 10^{-5} a.u., and the Hessian was computed with a step size of 10^{-3} Å. All electrons were correlated except for MRCCSD. The results of MkCCSD and BWCCSD were taken from Ref. 80, whereas those of *ic*-MRCCSD were obtained from Ref. 33. We summarize the computed geometries and frequencies in Table IV.

At the equilibrium geometry, an appropriate zeroth-order wave function of ozone is composed of two configurations, with either HOMO or LUMO of closed-shell HF being doubly occupied. As such, in MkCCSD and BWCCSD, a CASSCF (2*e*, 2*o*) model space was used. However, care must be taken in many MRCC approaches, because the open-shell singlet configuration $(\mathrm{HOMO})^1(\mathrm{LUMO})$,[1] whose contribution is zero with a C_{2v} geometry, is also indispensable for correctly describing the antisymmetric stretching frequency, $\omega(1b_2)$.[80,81] Furthermore, it was pointed out that, for MkCCSD and BWCCSD, which are not unitary-invariant, their values of $\omega(1b_2)$ depend rather sensitively on a small orbital rotation between these active orbitals; changes can be on the order of hundreds of wavenumbers.[80] Hence, while the reported values obtained with the natural orbitals of CASSCF (2*e*, 2*o*) are quite reasonable, they can deteriorate substantially when the active orbitals employed are prepared in a different manner. This clearly exposes the problems inherent in these methods. By contrast, *ic*-MRCCSD and EACCSD are free from this kind of problem, as they are orbital-invariant.

TABLE IV. Optimized geometries and frequencies of ozone.

			Freq. (cm^{-1})		
	$R_{\mathrm{O-O}}$ (Å)	∠OOO (°)	$1a_1$	$2a_1$	$1b_2$
CCSD	1.246	117.6	1287	770	1279
CCSD(T)	1.271	117.0	1163	723	1073
MkCCSD[a,b]	1.266	116.3	1180	739	1289
BWCCSD[a,b]	1.260	116.6	1218	748	1331
ic-MRCCSD[a,c]	1.266	116.6	1184	738	1243
SUHF	1.265	115.1	751	987	423
ECISD	1.246	116.8	1256	778	1478
EAQCC	1.349	115.0	1027	560	930
EACCSD	1.259	116.6	1214	751	1356
Exp.	1.272	116.8	1135	716	1089

[a]CASSCF(2*e*, 2*o*) reference, and 1*s* orbitals are frozen.
[b]The natural orbitals of CASSCF(2*e*, 2*o*) are used for the active orbitals. Taken from Ref. 80.
[c]Scheme "A". Taken from Ref. 33.

As in our previous study, SUHF was unable to predict reasonable frequencies and ECISD overestimated the $1b_2$ frequency. For this system, the instability of ELCCSD became manifest. In the vicinity of the equilibrium geometry, the correlation energy is considerably overestimated and does not converge in most cases. EAQCC, one of the ECEPA methods, mitigates this problem by appropriately treating the exclusion-principle-violating terms and thereby reducing the energy shift in the dressed Hamiltonian (see Ref. 45 for details), but the computed geometry is notably worse. This is most likely because the approximation to the quadratic $\hat{T}^2$ effect is inaccurate.

Including the quadratic term explicitly, EACCSD is immune to such instabilities. We obtain systematic improvements with EACCSD over both EAQCC and ECISD in most aspects. The bond length of EACCSD (1.259 Å) is in better agreement with the experimental value (1.272 Å), although slightly worse than other genuine MRCC methods. We note that, while the results of EACCSD and BWCCSD are very similar, this resemblance may be simply coincidental, given the orbital invariance issue in the latter method. That EACCSD does not correctly predict the ordering of $\omega(1a_1)$ and $\omega(1b_2)$ is consistent with other MRCC results. Hanauer and Köhn[33] showed that triple excitations are needed for *ic*-MRCC to be able to reproduce the correct ordering. Evangelista *et al.*[29] suggested that the inclusion of perturbative triples in MkCCSD could improve the description of $\omega(1b_2)$, although the possible sensitivity of their results to active orbital rotations was not addressed. Similarly, we expect that EACCSD(T) and EACCSDT should be able to give more accurate results than EACCSD.

E. Size-consistency

Size-consistency is a very difficult requirement to satisfy with SUHF-based methods. This is simply because a spin-projection operator does not generally allow for the separation of an SUHF wave function,

$$\hat{P}|\Phi\rangle = \hat{P}(|\Phi_{\mathrm{X}}\rangle \otimes |\Phi_{\mathrm{Y}}\rangle) \neq \hat{P}|\Phi_{\mathrm{X}}\rangle \otimes \hat{P}|\Phi_{\mathrm{Y}}\rangle, \qquad (55)$$

for subsystems X and Y. However, if X is symmetry-adapted and the symmetry-breaking occurs locally in Y, only the latter is subject to the spin projection. In such a case, SUHF is written as

$$\hat{P}|\Phi\rangle = \hat{P}|\Phi_X\rangle \otimes \hat{P}|\Phi_Y\rangle = |\Phi_X\rangle \otimes \hat{P}|\Phi_Y\rangle \quad (56)$$

and is separable. ECC formally satisfies size-consistency,

$$\hat{P}e^{\hat{T}}|\Phi\rangle = e^{\hat{T}_X}|\Phi_X\rangle \otimes \hat{P}e^{\hat{T}_Y}|\Phi_Y\rangle, \quad (57)$$

as $\hat{T}_X$ is symmetry-adapted. Here, we assume the intermediate normalization without loss of generality. This separability holds for EACCSD energy, but not for its residuals. The exact separability of residuals is only retained with the exponential *ansatz*. Furthermore, even if this problem was circumvented with the rigorous ECC theory, the current orthogonalization procedure might not guarantee Eq. (57), because **St** contains disconnected terms. Nonetheless, we expect the error due to size-inconsistency to be smaller in EACCSD than in ECISD and ECEPA because of the presence of the $\hat{T}_2^2$ term.

To numerically assess the non-separability of the EACCSD wave function, we computed the size-consistency errors of the non-interacting Be_2 and Be_3 molecules with a 6-31G basis, where each atom is either symmetry-adapted (**A**) or symmetry-broken (**B**). Here, the size-consistency error is defined as the difference between the energy of the supermolecule and the sum of energies of individual atoms, and all combinations of **A** and **B** are considered (except for **AAA**, which results in SR calculations). The results are presented in Table V. Although SUHF and ELCCSD give large errors for the **BB** and **BBB** systems, they are strictly size-consistent for **AB** and **AAB**. Furthermore, partial size-consistency is achieved for **ABB** in the sense that their energies are the sum of the energies of fragments **A** and **BB**. Unsurprisingly, ECISD, ECISD+Q, and EAQCC do not possess these properties. EACCSD is not exactly size-consistent for **AB** and **AAB**, either. In particular, the increase in error from **BB** to **ABB** is rather large, unlike the behavior of most other methods. Nevertheless, overall, the size-consistency error of EACCSD is small for these test cases. As expected, EACCSD drastically reduces the error in the ill-behaved ECISD energy. We expect this improvement to be further facilitated by eliminating unlinked terms via truncation schemes based on the Baker–Campbell–Hausdorff expansion.

F. Analysis of singular values

In *ic* approaches, one typically sets a numerical threshold η to truncate small singular values of the metric. The energy obtained noticeably depends on η. A tight threshold is therefore desired so that only those energies relevant to redundancies are discarded and a reliable energy is obtained, but such a threshold often causes convergence difficulties, because small singular values trigger large amplitudes. A loose threshold yields discontinuous potential curves, because the magnitudes of singular values change rather significantly across the reaction coordinates, and hence a fixed threshold changes the number of variables included in the cluster operator. In particular, many singular values that are larger than a reasonable threshold at equilibrium smoothly become numerical noise as a bond is stretched, and eventually vanish to zero. This makes it hard to choose an appropriate value for η in *ic* methods such as canonical transformation theory[64] and *ic*-MRCC.[32,65]

In EACCSD, one also takes the singular-value-decomposition of **S** to remove linear dependencies. However, the singular values are always well-behaved and the above problem never occurs. In other words, singular values do not turn to numerical noise and vice versa. In Fig. 8, we show the distribution of singular values of **S** for the double dissociation of H_2O. We have used the same geometry as in Sec. VI A for equilibrium and symmetrically stretched the two OH bonds from $1R_e$ to $3R_e$ with $R_e = 0.9929$ Å. Across the internuclear distance, there is always the same number of zero singular values, a feature common in *ic*-MRCC.[32,65] However, the singular values of **S** in EACCSD evidently lie in a certain narrow range and the distinction between nonzeros and zeros becomes even clearer as the bonds are stretched. This is a completely opposite characteristic to that seen in *ic*-MRCC[32,65] and can perhaps be ascribed to the physical simplicity of the spin-projection operator. In all the applications presented in this paper, we have seen the same behavior of **S** and have experienced no difficulties in choosing η and obtaining convergence.

However, we should note that EACCSD faces a different difficulty due to the black-box nature of SUHF active space. Namely, it will become increasingly difficult to diagonalize **S** to find $U_{\mu\beta}$ as the system size grows because **S** has the same dimension as the singles and doubles spaces. One could resort to an iterative diagonalizer[70] by finding the zero eigenvalues of **S**, but this will eventually fail as M_N increases. In

TABLE V. Size-consistency errors in mHartree for Be_2 and Be_3 supermolecules. Each atom in the system is labeled as **A** or **B**, indicating the use of symmetry-adapted or symmetry-broken orbitals.

System	SUHF	ECISD	ECISD+Q	ELCCSD	EAQCC	EACCSD
AB	0.000	5.178	−0.120	0.000	0.325	0.001
BB	8.255	3.523	−3.044	−5.551	−1.663	0.146
AAB	0.000	16.057	0.165	0.000	0.979	0.003
ABB	8.255	14.086	−2.442	−5.551	−0.973	0.862
BBB	19.606	12.476	−6.713	−14.376	−4.105	1.175

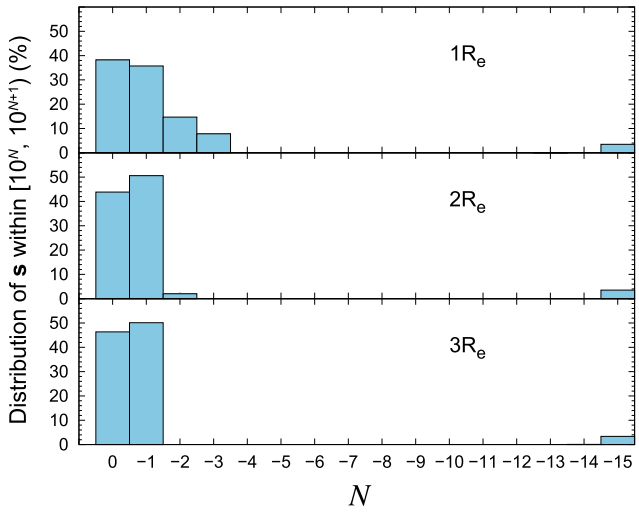

FIG. 8. Distribution of singular values for the symmetric dissociation of H_2O with $R_e = 0.9929$ Å. Each bar indicates the percentage of singular values λ_α with $10^N \le \lambda_\alpha < 10^{N+1}$.

particular, in our approach, $M_{\mathcal{N}}$ almost certainly scales as $\mathcal{O}(ov)$ for singlet systems, as discussed in our previous work.[44] This is clearly manifested in the computational cost of constructing the null-space projector $\boldsymbol{\mathcal{P}}$, which scales as $\mathcal{O}(o^3v^4)$, and the storage, which scales as $\mathcal{O}(o^3v^3)$. The size issue of $\mathcal{N}(\mathbf{S})$ therefore needs to be overcome before EACCSD can be used for larger applications. Dealing with this problem is not trivial and will be addressed in a forthcoming paper.

VII. CONCLUSIONS

In this paper, we have attempted to merge CC and spin-projected HF. Despite the simplicity of the ECC *ansatz*, the residual equations present two difficulties. First, they do not terminate naturally and require excitations of all electrons in the system. Second, the set of nonlinear equations is linearly dependent and possesses an infinitude of solutions. In the present work, we have avoided introducing approximations to a spin-projection operator $\hat{P}$ but rather truncated the exponential series of the wave operator at quadruples with respect to the underlying broken-symmetry determinant. This particular choice ensures that the method recovers the traditional CCSD when $\hat{P}$ is absent and that the computational scaling remains at $\mathcal{O}(N^6)$. To address the second issue, we have introduced an orthogonal excitation basis through the singular-value-decomposition of the metric $\mathbf{S}$ and the projection of its null space from the equations. This scheme was shown to correctly remove redundancies in the parametrization (t-amplitudes) in an orbital-invariant fashion. Furthermore, we have proposed an adaptive preconditioning scheme for updating t-amplitudes that introduces micro-iterations with an $\mathcal{O}(N^5)$ scaling but dramatically reduces the total number of macro-iterations with an $\mathcal{O}(N^6)$ scaling.

The current work focused on the performance of EACCSD for systems where genuine MRCCSD results are available. In all these systems, we have shown that EACCSD is more accurate (the H4 and H8 systems) or comparable (O_3) to MRCCSD. Moreover, unlike MRCC based on the Jeziorski–Monkhorst *ansatz*, our method was proven to be orbital-invariant, which is an important property. Although size-consistency is not strictly satisfied, EACCSD achieves remarkable improvements, in terms of correcting the size-inconsistency issue of SUHF, over previous CI-based models such as ECISD and ELCCSD.

Despite these initial encouraging results, the applicability of our method to larger systems is presently hampered by the need to diagonalize prohibitively large $\mathbf{S}$ to remove its null space. In principle, the same issue is shared by *ic* MR methods employing large active spaces (especially in combination with density matrix renormalization groups),[64] where approximating $\mathbf{S}$ usually causes significant errors in energy. We therefore need a robust and computationally efficient scheme to project out the null space. We will discuss this issue in a forthcoming paper.

SUPPLEMENTARY MATERIAL

See supplementary material describes the effective coefficients $\boldsymbol{\omega}$, matrix elements, and the total energies of H4 and H8.

ACKNOWLEDGMENTS

We would like to thank Gustavo E. Scuseria for stimulating discussions. This research was supported by MEXT as "Priority Issue on Post-K computer" (Development of new fundamental technologies for high-efficiency energy creation, conversion/storage and use) using the computational resources of the K computer provided by the RIKEN Advanced Institute for Computational Science through the HPCI System Research project (Project ID: hp170259). This work was partly supported by JSPS KAKENHI Grant-in-Aids for Scientific Research (A) (Grant No. JP18H03900) and Young Scientists (B) (Grant No. JP17K14438).

APPENDIX A: FIRST-ORDER INTERACTING SPACE OF SPIN-PROJECTED MANIFOLD

In principle, a spin-projected determinant interacts with essentially any projected n-tuply excited determinant $\hat{P}|\Phi_\mu^{(n)}\rangle$ through the Hamiltonian

$$\langle\Phi_\mu^{(n)}|\hat{H}\hat{P}|\Phi_0\rangle \neq 0, \tag{A1}$$

because the projected configuration space that each excitation level spans overlaps with that of other excitation levels. However, it can easily be shown that the first-order interacting space with respect to $\hat{P}|\Phi_\mu^{(n)}\rangle$ is completely spanned by the projected configuration spaces that consist of determinants with $n-2 \sim n+2$ excitations (not necessarily different by two electron substitutions). For example, the FOIS of SUHF is composed of the projected singles and doubles, and likewise that of the projected doubles is composed of SUHF and the projected singles to quadruples, with all possible spin combinations that maintain the same $\langle S_z\rangle$.

To see this, we first note that, while $|\Phi_0\rangle$ is a determinant in a broken-symmetry orbital basis ϕ_i^α and ϕ_i^β,

$$|\Phi_0\rangle = |\phi_{1_\alpha}\cdots\phi_{N_\alpha}\rangle|\phi_{1_\beta}\cdots\phi_{N_\beta}\rangle, \tag{A2}$$

the orbital transformation to some spin-adapted orbitals φ_P (capital letters indicate spatial orbitals of this basis) can be achieved by

$$\varphi_p = \sum_{q_\alpha}^{\text{all}} \mathcal{U}_{q_\alpha}^P \phi_{q_\alpha} = \sum_{q_\beta}^{\text{all}} \mathcal{U}_{q_\beta}^P \phi_{q_\beta}, \tag{A3}$$

where $\boldsymbol{\mathcal{U}}$ is a unitary matrix. The unitary group generator $\hat{\mathcal{E}}_Q^P$ in the orbital basis of $\{\varphi_P\}$ (represented as the creation and annihilation operators $\{\hat{c}^{P_\sigma}, \hat{c}_{P_\sigma}\}$) can be back-transformed to the broken-symmetry orbitals ϕ_P,

$$\begin{aligned}\hat{\mathcal{E}}_Q^P &= \hat{c}^{P_\alpha}\hat{c}_{Q_\alpha} + \hat{c}^{P_\beta}\hat{c}_{Q_\beta} \\ &= \sum_{t_\alpha u_\alpha} \mathcal{U}_{t_\alpha}^P \mathcal{U}_{u_\alpha}^{Q*} \hat{a}^{t_\alpha}\hat{a}_{u_\alpha} + \sum_{t_\beta u_\beta} \mathcal{U}_{t_\beta}^P \mathcal{U}_{u_\beta}^{Q*} \hat{a}^{t_\beta}\hat{a}_{u_\beta}.\end{aligned} \tag{A4}$$

As $\left[\hat{\mathcal{E}}_q^p, \hat{P}\right] = 0$ for spin-projection, the FOIS of an SUHF wave function $\hat{P}|\Phi_0\rangle$ is then found to be

$$\hat{\mathcal{E}}_Q^P\hat{\mathcal{E}}_S^R\hat{P}|\Phi_0\rangle = \hat{P}\hat{\mathcal{E}}_Q^P\hat{\mathcal{E}}_S^R|\Phi_0\rangle \in \hat{P}|\Phi_0\rangle \oplus \hat{P}|\Phi_i^a\rangle \oplus \hat{P}|\Phi_{ij}^{ab}\rangle, \tag{A5}$$

where both α and β spins are implicit in a, b, i, j. Similarly, the FOIS of projected doubles $\hat{P}|\Phi_{ij}^{ab}\rangle$ is shown to be spanned by the SUHF reference, projected singles, . . ., quadruples.

Finally, we mention that the nonorthogonality between different excitation ranks makes it impossible to distinguish them unless they are properly orthogonalized. This means, for example, that the importance of triples excitations in EACCSDT compared to that of singles and doubles excitations in EACCSD may not be clear. To avoid such ambiguous definitions of excitations in spin-extended methods, the sequential orthogonalization technique should be invoked, as proposed by the pioneering work of Hanauer and Kohn for ic-MRCC (version D).[33]

APPENDIX B: SINGULARITIES IN LINEARIZED COUPLED CLUSTER

Standard linearized CC methods have a tendency to diverge as the system size increases unless the active space is large enough. ELCCSD is usually solved as the dressed ECISD eigenvalue problem

$$\begin{pmatrix} 0 & \mathbf{H}_{0\nu} \\ \mathbf{H}_{\mu 0} & \mathbf{H}_{\mu\nu} + E_c\mathbf{S}_{\mu\nu} \end{pmatrix} \begin{pmatrix} c_0 \\ \mathbf{c}_\nu \end{pmatrix} = E_c \begin{pmatrix} 1 & \mathbf{0}_{0\nu} \\ \mathbf{0}_{\mu 0} & \mathbf{S}_{\mu\nu} \end{pmatrix} \begin{pmatrix} c_0 \\ \mathbf{c}_\nu \end{pmatrix}, \quad \text{(B1)}$$

in the intermediate normalization, i.e., with the basis $\{\hat{P}|\Phi_0\rangle, \hat{Q}|\Phi_\mu\rangle\}$. The equations for SRLCC and MRLCC can be written similarly. The correlation energy E_c is therefore formally determined by iteratively solving Eq. (B1). If the first excited state of ECISD is well separated from the ground state, to a good approximation, using the dressed Hamiltonian corresponds to shifting down all the CI excitation energies ε_i (i.e., the interacting space) by E_c [see Fig. 9(a)]. This procedure is, however, not stable if the interacting space is close to the reference space and, therefore, the low-lying excitation energies are small compared to the CI correlation energy. In this case, as shown in Fig. 9(b), these excited states (red) become intruders and go below the energy of the true ground state (green). Evidently, as the iteration proceeds, E_c will eventually overcorrect or diverge.

It is noteworthy that the excitation energies are size-intensive quantities, whereas the correlation energy of LCC is designed to be size-extensive. Therefore, the larger the system size, the more severely LCC is likely to overestimate the correlation energy. This issue is also common in MRLCC, and therefore a large (complete) active space is required to avoid instabilities.

APPENDIX C: ORBITAL INVARIANCE PROPERTY OF ECC

A method is said to be orbital-invariant if the energy does not change on a unitary rotation in a certain orbital space $\mathbb{P}$, which is given by

$$\hat{\tilde{a}}^{\bar{p}} = \sum_q^{\mathbb{P}} \hat{a}^q \mathcal{U}_q^{\bar{p}}, \quad \text{(C1)}$$

$$\hat{\tilde{a}}_{\bar{p}} = \sum_q^{\mathbb{P}} \hat{a}_q \mathcal{U}_q^{\bar{p}*}, \quad \text{(C2)}$$

where $\mathcal{U}_q^{\bar{p}}$ is again a unitary matrix and a bar indicates the transformed spin-orbitals.

To show the orbital invariance of SUHF and ECC, we first assume that the SUHF and ECC energies have converged and the residual equations have been solved. We then perform a unitary rotation within each orbital space to verify that the energy and residuals are unchanged.[30] In these methods, the orbital space can be naturally separated into occupied and virtual subspaces of the underlying broken-symmetry determinant $|\Phi_0\rangle = \hat{a}^{i_1 i_2 \cdots i_{n_e}}|\rangle$. Therefore, $\mathbb{P}$ spans either an occupied ($\mathcal{U}_{\rm occ}$) or virtual ($\mathcal{U}_{\rm vir}$) space.

As occupied orbital rotations in $|\Phi_0\rangle$ simply change the phase,

$$|\bar{\Phi}_0\rangle = |\Phi_0\rangle \det(\mathcal{U}_{\rm occ}), \quad \text{(C3)}$$

the unitary invariance of the SUHF energy is immediately obvious:

$$\bar{E}_{\rm SUHF} = \frac{\langle\bar{\Phi}_0|\hat{H}\hat{P}|\bar{\Phi}_0\rangle}{\langle\bar{\Phi}_0|\hat{P}|\bar{\Phi}_0\rangle} = \frac{\langle\Phi_0|\hat{H}\hat{P}|\Phi_0\rangle}{\langle\Phi_0|\hat{P}|\Phi_0\rangle} = E_{\rm SUHF}. \quad \text{(C4)}$$

It then follows that the residuals of SUHF, defined as

$$\begin{aligned} \mathbb{R}_{a_1 i_2 i_3 \cdots i_{n_e}}^{i_1 i_2 i_3 \cdots i_{n_e}} &= \langle|\hat{a}_{a_1 i_2 i_3 \cdots i_{n_e}} \hat{H}_N \hat{P} \hat{a}^{i_1 i_2 i_3 \cdots i_{n_e}}|\rangle \\ &= \langle\Phi_0|\hat{a}^{i_1}\hat{a}_{a_1}\hat{H}_N\hat{P}|\Phi_0\rangle, \end{aligned} \quad \text{(C5)}$$

are also invariant at convergence ($\mathbb{R}_{a_1 i_2 i_3 \cdots i_{n_e}}^{i_1 i_2 i_3 \cdots i_{n_e}} = 0$), because

$$\begin{aligned} \mathbb{R}_{\bar{a}_1 \bar{i}_2 \bar{i}_3 \cdots \bar{i}_{n_e}}^{\bar{i}_1 \bar{i}_2 \bar{i}_3 \cdots \bar{i}_{n_e}} &= \langle\bar{\Phi}_0|\hat{\tilde{a}}^{\bar{i}_1}\hat{\tilde{a}}_{\bar{a}_1}\hat{H}_N\hat{P}|\bar{\Phi}_0\rangle, \\ &= |\det(\mathcal{U}_{\rm occ})|^2 \sum_{i_1 a_1} \mathcal{U}_{i_1}^{\bar{i}_1}\mathcal{U}_{a_1}^{\bar{a}_1 *}\langle\Phi_0|\hat{a}^{i_1}\hat{a}_{a_1}\hat{H}_N\hat{P}|\Phi_0\rangle \\ &= \sum_{i_1 a_1} \mathcal{U}_{i_1}^{\bar{i}_1}\mathcal{U}_{a_1}^{\bar{a}_1 *}\mathbb{R}_{a_1 i_2 i_3 \cdots i_{n_e}}^{i_1 i_2 i_3 \cdots i_{n_e}} \\ &= 0. \end{aligned} \quad \text{(C6)}$$

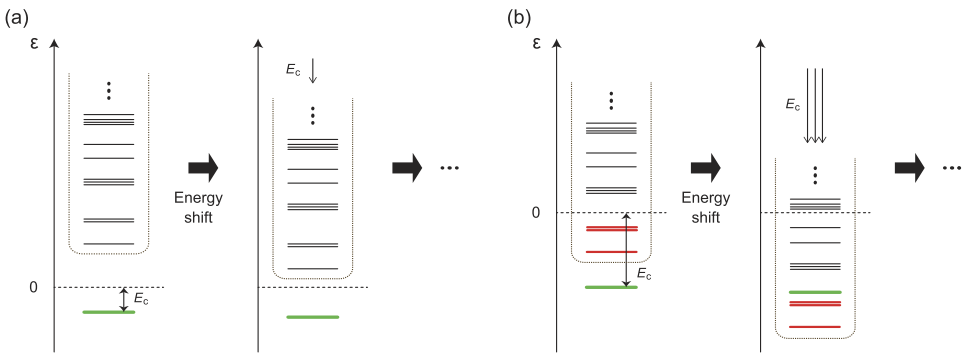

FIG. 9. Iterative procedure of linearized coupled cluster. (a) If the reference space (green) is sufficiently separated from the interacting space (surrounded by a dashed box), the excited states still appear well above the ground state after the energy shift. (b) If, however, the reference space is too close to the interacting space, the energy shift E_c overcorrects the spectrum and low-lying states (red) go below the true ground state.

Furthermore, Eq. (C6) clearly indicates that only i_1 and a_1 are relevant to the orbital rotation for $\mathbb{R}^{i_1 i_2 i_3 \cdots i_{n_e}}_{a_1 i_2 i_3 \cdots i_{n_e}}$, and all the inner indices can be excluded from the discussion; this is of course what is known as the Fock matrix

$$F^{i_1}_{a_1} \equiv \mathbb{R}^{i_1 i_2 i_3 \cdots i_{n_e}}_{a_1 i_2 i_3 \cdots i_{n_e}}. \tag{C7}$$

The ECC energy is also orbital-invariant because $\hat{T}$ is; for example,

$$\begin{aligned}
\hat{\bar{T}}_2 &= \frac{1}{4}\sum_{\bar{i}\bar{j}\bar{a}\bar{b}} \bar{t}^{\bar{a}\bar{b}}_{\bar{i}\bar{j}} \hat{\bar{a}}^{\bar{a}\bar{b}}_{\bar{i}\bar{j}} \\
&= \frac{1}{4}\sum_{\bar{i}\bar{j}\bar{a}\bar{b}} \left(\sum_{ijab} t^{ab}_{ij} \mathcal{U}^{\bar{a}*}_{a} \mathcal{U}^{\bar{b}*}_{b} \mathcal{U}^{\bar{i}}_{i} \mathcal{U}^{\bar{j}}_{j} \right) \left(\sum_{klcd} \hat{a}^{cd}_{kl} \mathcal{U}^{\bar{c}}_{c} \mathcal{U}^{\bar{d}}_{d} \mathcal{U}^{\bar{k}*}_{k} \mathcal{U}^{\bar{l}*}_{l} \right) \\
&= \frac{1}{4}\sum_{\substack{ijab \\ klcd}} t^{ab}_{ij} \hat{a}^{cd}_{kl} \delta_{ac}\delta_{bd}\delta_{ik}\delta_{jl} \\
&= \hat{T}_2,
\end{aligned} \tag{C8}$$

where we have used $\boldsymbol{\mathcal{U}}^{\dagger}_{\mathrm{occ}}\boldsymbol{\mathcal{U}}_{\mathrm{occ}} = \mathbf{1}$ and similarly for the virtual block. Obviously,

$$\bar{E}_{\mathrm{c}} = \langle \bar{\Phi}_0 | \hat{H}_N \hat{P} e^{\hat{\bar{T}}} | \bar{\Phi}_0 \rangle = \langle \Phi_0 | \hat{H}_N \hat{P} e^{\hat{T}} | \Phi_0 \rangle = E_{\mathrm{c}}. \tag{C9}$$

For the residuals, we first consider the property of the metric under a unitary rotation. For the doubles space, for example, this can be transformed as

$$\begin{aligned}
\bar{S}_{\bar{\mu}\bar{\nu}} &= \langle \bar{\Phi}_0 | \hat{\bar{a}}^{\bar{i}\bar{j}}_{\bar{a}\bar{b}} \, \hat{\bar{\mathcal{Q}}} \, \hat{\bar{a}}^{\bar{c}\bar{d}}_{\bar{k}\bar{l}} | \bar{\Phi}_0 \rangle \\
&= \sum_{\substack{ijab \\ klcd}} \langle \Phi_0 | \hat{a}^{ij}_{ab} \, \hat{\mathcal{Q}} \, \hat{a}^{cd}_{kl} | \Phi_0 \rangle \mathcal{U}^{\bar{i}}_{i} \mathcal{U}^{\bar{j}}_{j} \mathcal{U}^{\bar{a}*}_{a} \mathcal{U}^{\bar{b}*}_{b} \mathcal{U}^{\bar{c}}_{c} \mathcal{U}^{\bar{d}}_{d} \mathcal{U}^{\bar{k}*}_{k} \mathcal{U}^{\bar{l}*}_{l} \\
&= \sum_{\substack{ijab \\ klcd}} S^{ij,cd}_{ab,kl} \mathcal{U}^{\bar{i}}_{i} \mathcal{U}^{\bar{j}}_{j} \mathcal{U}^{\bar{a}*}_{a} \mathcal{U}^{\bar{b}*}_{b} \mathcal{U}^{\bar{c}}_{c} \mathcal{U}^{\bar{d}}_{d} \mathcal{U}^{\bar{k}*}_{k} \mathcal{U}^{\bar{l}*}_{l}.
\end{aligned} \tag{C10}$$

A similar relation holds for the singles space. As the singular-value-decomposition of $\mathbf{S}$ can be written as

$$S^{ij,cd}_{ab,kl} = \sum_{\gamma}^{\mathcal{R}\oplus\mathcal{N}} U^{ij}_{ab}(\gamma) \; \lambda_\gamma \; U^{kl}_{cd}(\gamma)^*, \tag{C11}$$

the singular vectors accordingly become

$$\bar{U}^{\bar{i}\bar{j}}_{\bar{a}\bar{b}}(\gamma) = \sum_{ijab} U^{ij}_{ab}(\gamma) \, \mathcal{U}^{\bar{i}}_{i} \mathcal{U}^{\bar{j}}_{j} \mathcal{U}^{\bar{a}*}_{a} \mathcal{U}^{\bar{b}*}_{b}, \tag{C12}$$

noting that the singular values remain unchanged, $\bar{\lambda}_\gamma = \lambda_\gamma$. Likewise, the unprojected residuals r_μ can be transformed as

$$\bar{r}^{\bar{a}\bar{b}}_{\bar{i}\bar{j}} = \sum_{ijab} r^{ab}_{ij} \, \mathcal{U}^{\bar{i}}_{i} \mathcal{U}^{\bar{j}}_{j} \mathcal{U}^{\bar{a}*}_{a} \mathcal{U}^{\bar{b}*}_{b}, \tag{C13}$$

where the inner indices have been suppressed, as in earlier equations. Hence, it is evident that, when the null-space is projected, $\tilde{\bar{r}}^{\bar{a}\bar{b}}_{\bar{i}\bar{j}}$ can also be written as a unitary transformation of $\tilde{r}^{ab}_{ij}$, and therefore remains at zero. This completes the proof of the unitary invariance of ECCSD (and EACCSD). In Sec. VI A, we have provided numerical evidence.

[1] T. V. Voorhis and M. Head-Gordon, Chem. Phys. Lett. **330**, 585 (2000).
[2] T. V. Voorhis and M. Head-Gordon, J. Chem. Phys. **113**, 8873 (2000).
[3] A. I. Krylov, Chem. Phys. Lett. **338**, 375 (2001).
[4] A. I. Krylov, Chem. Phys. Lett. **350**, 522 (2001).
[5] P. Piecuch, K. Kowalski, I. S. O. Pimienta, and M. J. McGuire, Int. Rev. Phys. Chem. **21**, 527 (2002).
[6] D. Kats and F. R. Manby, J. Chem. Phys. **139**, 021102 (2013).
[7] P. A. Limacher *et al.*, J. Chem. Theory Comput. **9**, 1394 (2013).
[8] T. Stein, T. M. Henderson, and G. E. Scuseria, J. Chem. Phys. **140**, 214113 (2014).
[9] I. W. Bulik, T. M. Henderson, and G. E. Scuseria, J. Chem. Theory Comput. **11**, 3171 (2015).
[10] B. O. Roos, P. R. Taylor, and P. E. Siegbahn, Chem. Phys. **48**, 157 (1980).
[11] P. Siegbahn, A. Heiberg, B. Roos, and B. Levy, Phys. Scr. **21**, 323 (1980).
[12] P. E. M. Siegbahn, J. Almlöf, A. Heiberg, and B. O. Roos, J. Chem. Phys. **74**, 2384 (1981).
[13] J. Čížek, J. Chem. Phys. **45**, 4256 (1966).
[14] R. J. Bartlett, Annu. Rev. Phys. Chem. **32**, 359 (1981).
[15] R. J. Bartlett and M. Musiał, Rev. Mod. Phys. **79**, 291 (2007).
[16] I. Shavitt and R. J. Bartlett, *Many-Body Methods in Chemistry and Physics: MBPT and Coupled-Cluster Theory* (Cambridge University Press, Cambridge, UK, 2009).
[17] B. Jeziorski and H. J. Monkhorst, Phys. Rev. A **24**, 1668 (1981).
[18] I. Hubač and P. Neogrády, Phys. Rev. A **50**, 4558 (1994).
[19] J. Mášik, I. Hubač, and P. Mach, J. Chem. Phys. **108**, 6571 (1998).
[20] J. Pittner, J. Chem. Phys. **118**, 10876 (2003).
[21] J. Pittner and O. Demel, J. Chem. Phys. **122**, 181101 (2005).
[22] U. S. Mahapatra, B. Datta, B. Bandyopadhyay, and D. Mukherjee, Adv. Quantum Chem. **30**, 163 (1998).
[23] U. S. Mahapatra, B. Datta, and D. Mukherjee, J. Chem. Phys. **110**, 6171 (1999).
[24] S. Chattopadhyay, U. S. Mahapatra, and D. Mukherjee, J. Chem. Phys. **111**, 3820 (1999).
[25] S. Chattopadhyay, U. S. Mahapatra, and D. Mukherjee, J. Chem. Phys. **112**, 7939 (2000).
[26] S. Chattopadhyay, D. Pahari, D. Mukherjee, and U. S. Mahapatra, J. Chem. Phys. **120**, 5968 (2004).
[27] K. Bhaskaran-Nair, O. Demel, and J. Pittner, J. Chem. Phys. **129**, 184105 (2008).
[28] L. Kong, Int. J. Quantum Chem. **109**, 441 (2009).
[29] F. A. Evangelista, E. Prochnow, J. Gauss, and H. F. Schaefer III, J. Chem. Phys. **132**, 074107 (2010).
[30] L. Kong, Int. J. Quantum Chem. **110**, 2603 (2010).
[31] T. Yanai and G. K.-L. Chan, J. Chem. Phys. **124**, 194106 (2006).
[32] F. A. Evangelista and J. Gauss, J. Chem. Phys. **134**, 114102 (2011).
[33] M. Hanauer and A. Köhn, J. Chem. Phys. **134**, 204111 (2011).
[34] D. I. Lyakh, M. Musiał, V. F. Lotrich, and R. J. Bartlett, Chem. Rev. **112**, 182 (2012).
[35] A. Köhn, M. Hanauer, L. A. Mück, T.-C. Jagau, and J. Gauss, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **3**, 176 (2012).
[36] G. E. Scuseria, C. A. Jiménez-Hoyos, T. M. Henderson, K. Samanta, and J. K. Ellis, J. Chem. Phys. **135**, 124108 (2011).
[37] C. A. Jiménez-Hoyos, T. M. Henderson, T. Tsuchimochi, and G. E. Scuseria, J. Chem. Phys. **136**, 164109 (2012).
[38] P.-O. Löwdin, Phys. Rev. **97**, 1509 (1955).
[39] C. A. Jiménez-Hoyos, R. Rodríguez-Guzmán, and G. E. Scuseria, J. Chem. Phys. **139**, 204102 (2013).
[40] T. Tsuchimochi and T. Van Voorhis, J. Chem. Phys. **142**, 124103 (2015).
[41] T. Tsuchimochi, J. Chem. Phys. **143**, 144114 (2015).
[42] T. Tsuchimochi and T. Van Voorhis, J. Chem. Phys. **141**, 164117 (2014).
[43] T. Tsuchimochi and S. Ten-no, J. Chem. Phys. **144**, 011101 (2016).
[44] T. Tsuchimochi and S. Ten-no, J. Chem. Theory Comput. **12**, 1741 (2016).
[45] T. Tsuchimochi and S. Ten-no, J. Chem. Theory Comput. **13**, 1667 (2017).
[46] P. G. Szalay and R. J. Bartlett, Chem. Phys. Lett. **214**, 481 (1993).
[47] T. Duguet, J. Phys. G: Nucl. Part. Phys. **42**, 025107 (2015).
[48] J. A. Gomez, M. Degroote, J. Zhao, Y. Qiu, and G. E. Scuseria, Phys. Chem. Chem. Phys. **19**, 22385 (2017).
[49] J. M. Wahlen-Strothman *et al.*, J. Chem. Phys. **146**, 054110 (2017).
[50] Y. Qiu, T. M. Henderson, and G. E. Scuseria, J. Chem. Phys. **146**, 184105 (2017).
[51] Y. Qiu, T. M. Henderson, J. Zhao, and G. E. Scuseria, J. Chem. Phys. **147**, 064111 (2017).
[52] Y. Qiu, T. M. Henderson, and G. E. Scuseria, J. Phys. Chem. **145**, 111102 (2016).

[53]H. Werner and E. Reinsch, J. Chem. Phys. **76**, 3144 (1982).
[54]T. Yanai and G. K.-L. Chan, J. Chem. Phys. **127**, 104107 (2007).
[55]P. Pulay, Chem. Phys. Lett. **73**, 393 (1980).
[56]P. Pulay, J. Comput. Chem. **3**, 556 (1982).
[57]G. E. Scuseria, T. J. Lee, and H. F. Schaefer III, Chem. Phys. Lett. **130**, 236 (1986).
[58]M. Hanauer and A. Köhn, J. Chem. Phys. **137**, 131103 (2012).
[59]D. Mukherjee, Chem. Phys. Lett. **274**, 561 (1997).
[60]W. Kutzelnigg and D. Mukherjee, J. Chem. Phys. **107**, 432 (1997).
[61]H.-J. Werner and P. J. Knowles, J. Chem. Phys. **89**, 5803 (1988).
[62]K. Andersson, P.-Å. Malmqvist, B. O. Roos, A. J. Sadlej, and K. Wolinski, J. Phys. Chem. **94**, 5483 (1990).
[63]L. Kong, K. R. Shamasundar, O. Demel, and M. Nooijen, J. Chem. Phys. **130**, 114101 (2009).
[64]E. Neuscamman, T. Yanai, and G. K.-L. Chan, J. Chem. Phys. **132**, 024106 (2010).
[65]M. Hanauer, "Internally contracted multireference coupled-cluster methods," Ph.D. thesis, Universität Mainz, 2013.
[66]N. Oliphant and L. Adamowicz, J. Chem. Phys. **94**, 1229 (1991).
[67]V. V. Ivanov and L. Adamowicz, J. Chem. Phys. **112**, 9258 (2000).
[68]P. Piecuch and L. Adamowicz, J. Chem. Phys. **100**, 5792 (1994).
[69]P. Piecuch, S. A. Kucharski, and R. J. Bartlett, J. Chem. Phys. **110**, 6103 (1999).
[70]E. R. Davidson, J. Comput. Phys. **17**, 87 (1975).
[71]N. C. Handy, J. A. Pople, M. Head-Gordon, K. Raghavachari, and G. W. Trucks, Chem. Phys. Lett. **164**, 185 (1989).
[72]W. J. Lauderdale, J. F. Stanton, J. Gauss, J. D. Watts, and R. J. Bartlett, Chem. Phys. Lett. **187**, 21 (1991).
[73]A. T. Amos and G. G. Hall, Proc. R. Soc. London, Ser. A **263**, 483 (1961).
[74]P. Karadakov, Int. J. Quantum Chem. **27**, 699 (1985).
[75]F. A. Evangelista, W. D. Allen, and H. F. Schaefer III, J. Chem. Phys. **125**, 154113 (2006).
[76]S. R. Langhoff and E. R. Davidson, Int. J. Quantum Chem. **8**, 61 (1974).
[77]E. R. Davidson, *The World of Quantum Chemistry* (Reidel, Dordrecht, 1974).
[78]I. A. Pople, R. Seeger, and R. Krishnan, Int. J. Quantum Chem. **12**(S11), 149 (1977).
[79]T. Tsuchimochi and S. Ten-no, J. Chem. Phys. **146**, 074104 (2017).
[80]F. A. Evangelista, W. D. Allen, and H. F. Schaefer III, J. Chem. Phys. **127**, 024102 (2007).
[81]X. Li and J. Paldus, J. Chem. Phys. **110**, 2844 (1999).