

# Leveraging configuration interaction singles for qualitative descriptions of ground and excited states: state-averaging, linear-response, and spin-projection

Takashi Tsuchimochi<sup>1,2, a)</sup> and Benjamin Mokhtar<sup>3</sup>

<sup>1)</sup>Department of Applied Chemistry, Shibaura Institute of Technology, 3-7-5 Toyosu, Koto-ku, Tokyo 135-8548 Japan

<sup>2)</sup>Institute for Molecular Science, 38 Nishigonaka, Myodaiji, Okazaki 444-8585 Japan

<sup>3)</sup>Graduate School of Engineering and Science, Shibaura Institute of Technology, 3-7-5 Toyosu, Koto-ku, Tokyo 135-8548 Japan

While configuration interaction singles (CIS) provides a computationally efficient description of excited states, it systematically overestimates excitation energies and performs poorly for strongly correlated systems, partially due to the lack of orbital relaxation and the strong ground-state bias of Hartree–Fock orbitals. To address these limitations, we present a unified variational framework that extends CIS by incorporating orbital optimization (state-specific and state-averaged), linear-response orbital relaxation via double-CIS schemes, and spin-symmetry breaking and restoration. In spin-projected state-averaged formulations, standard multistate parametrizations are no longer valid because the projection operator breaks the unitary invariance of orbital rotations and induces nonorthogonal couplings among states. By formulating a rigorous state-averaged objective in the projected subspace, we derive analytic gradients and Hessians and enable robust optimization using a trust-region augmented Hessian algorithm. Benchmark calculations show that spin projection alone significantly exacerbates the CIS overestimation in weakly correlated systems, whereas combining spin projection with state averaging or double-CI corrections substantially reduces errors, particularly for Rydberg excitations. We further demonstrate that state averaging and spin projection provide complementary and essential benefits in strongly correlated regimes, as illustrated by the bond dissociation of hydrogen fluoride.

## I. INTRODUCTION

Accurate yet computationally affordable descriptions of electronically excited states remain a long-standing challenge in quantum chemistry. High-level wavefunction methods such as equation-of-motion coupled-cluster theory with singles and doubles or triples (EOM-CCSD/CCSDT)<sup>1–3</sup> and other related methods<sup>4,5</sup> provide reliable excitation energies for a wide range of molecular systems. However, their steep computational cost severely restricts their applicability to large molecules, extended excited-state manifolds, and the exploration of excited-state potential energy surfaces. This limitation has motivated sustained efforts to develop low-cost excited-state methods that capture essential physical effects while retaining the favorable scaling and conceptual simplicity of mean-field-based approaches.

Among such methods, configuration interaction singles (CIS) represents one of the most fundamental low-cost approaches to excited states. Formally equivalent to the Tamm–Dancoff approximation (TDA) to time-dependent Hartree–Fock (TDHF), CIS provides a variationally stable and computationally efficient description of excited states in terms of single excitations from a Hartree–Fock (HF) reference. Despite these appealing features, CIS is well known to systematically overestimate excitation energies.<sup>6–8</sup> This deficiency originates from two fundamental limitations: the absence of dynamical correlation

and, more importantly for many classes of excitations, the lack of orbital relaxation. Because excitation energies are differences between ground- and excited-state energies, their errors arise from imbalances in correlation energy between the two states. In particular, the missing differential correlation energy is typically on the order of the correlation energy associated with the electron pairs involved in the excitation. Dynamical correlation effects can often be treated perturbatively by including double excitations with respect to CIS states, leading to the widely used CIS(D) approach.<sup>9</sup> While treatments of dynamical correlation have a long history and are now relatively well established, the incorporation of orbital relaxation effects for excited states has attracted renewed interest in recent years.<sup>10–18</sup>

Because HF orbitals are optimized solely for the ground state, CIS excited states are inherently biased toward the ground-state electronic structure, leading to a systematic imbalance between ground and excited states. This bias is particularly severe for excitations that involve substantial redistribution of electron density such as charge-transfer<sup>8,19</sup> and core excited states,<sup>20–23</sup> and for near-degenerate electronic states.<sup>24,25</sup> Consequently, CIS, or more broadly, single-reference linear-response methods including TDDFT, also fails qualitatively in strongly correlated regimes, such as bond dissociation and conical intersections,<sup>26–28</sup> where the single-reference picture breaks down.

Orbital relaxation therefore plays a central role in improving CIS-based excited-state methods. One rigorous way to incorporate orbital relaxation is full state-specific orbital optimization for each excited state. This philos-

---

<sup>a)</sup>Electronic mail: [tsuchimochi@gmail.com](mailto:tsuchimochi@gmail.com)

ophy underlies excited-state mean-field theory (ESMF), introduced by Neuscamman and co-workers,<sup>29–32</sup> which can be viewed as a state-specific, orbital-optimized CIS (SSCIS) method. In ESMF, or equivalently SSCIS, orbitals and CI coefficients are optimized simultaneously for a targeted excited state, yielding substantial improvements in excitation energies through balanced orbital relaxation especially when combined with perturbative treatments of dynamical correlation.<sup>29,33,34</sup> However, this approach introduces a highly nonlinear optimization problem that is variationally unstable and potentially difficult to converge in practice. Moreover, the resulting excited states are state-specific and nonorthogonal, complicating the description of multiple excited states and the evaluation of transition properties.

An alternative strategy is to incorporate orbital relaxation effects within a linear-response-like framework while retaining the orthogonality and conceptual simplicity of CIS. A pioneering work by Subotnik and co-workers introduced a single-shot perturbative orbital relaxation correction to CIS using the HF Hessian as an approximation to the CIS Hessian.<sup>8</sup> While this approach improved charge-transfer excitation energies, it was found to be insufficient in general.<sup>19,35</sup> More recently, double CIS (DCIS) proposed by one of us provides a systematic framework,<sup>36</sup> in which CIS is effectively performed on top of CIS to incorporate first-order orbital relaxation effects. DCIS can be interpreted as a linear-response correction within the TDA framework and has been shown to systematically lower CIS excitation energies for charge-transfer states, achieving improved accuracy through error cancellation between ground and excited states, at modest additional computational cost. Importantly, this approach preserves orthonormality between states, enabling straightforward evaluation of transition properties.

To mitigate the strong ground-state bias inherent in HF-based excited-state methods, state averaging provides yet another route. By optimizing orbitals with respect to an average energy over several electronic states, state-averaged approaches reduce the preferential stabilization of the ground state and yield a more balanced description of excited states. State averaging has a long history in multiconfigurational self-consistent field (MCSCF) theories<sup>37,38</sup>, and also has been combined with other methods,<sup>39,40</sup> particularly for photochemical applications and near-degenerate electronic structures. When applied to CIS-based methods, state-averaged CIS (SACIS) should provide a natural framework for introducing orbital relaxation “on average” across multiple states, as a variationally stable alternative to state-specific orbital optimization. However, whether and to what extent this strategy mitigates the systematic overestimation of CIS excitation energies remains to be systematically assessed.

Strong static correlation presents an additional challenge for CIS-based excited-state methods. In such regimes, the HF reference becomes qualitatively incor-

rect, leading to severe failures of CIS. Spin-unrestricted formulations can partially alleviate this problem by allowing symmetry breaking, but they introduce spin contamination that complicates the interpretation of excited states. Spin-projection techniques<sup>41,42</sup> offer a systematic remedy by restoring spin symmetry while retaining the flexibility of broken-symmetry determinants. Spin-extended CIS (ECIS),<sup>24,25</sup> built upon a spin-projected unrestricted HF (SUHF) reference (also known as projected HF),<sup>43</sup> provides a natural extension of CIS for strongly correlated systems and has been shown to yield qualitatively correct excitation energies in prototypical near-degenerate cases.<sup>24,25</sup> However, its adequacy across a broad range of excitations and correlation regimes remains largely unexplored.

It is also important to note that, despite these conceptual advances, orbital-optimization in CIS and ECIS is expected to pose significant numerical challenges just like MCSCF. The simultaneous optimization of orbitals and CI coefficients leads to strongly coupled, nonlinear equations, for which convergence is far from trivial. Conventional schemes based on (quasi-)Newton methods<sup>44,45</sup> or DIIS (direct inversion of the iterative subspace)<sup>46,47</sup> would often suffer from slow convergence, numerical instability, or convergence to saddle points, particularly when the initial trial states are far from convergence due to the presence of near-degeneracies or multiple competing solutions. Robust optimization algorithms are therefore essential for making these methods practically viable. One promising approach is the trust-region augmented Hessian (TRAH) method,<sup>48,49</sup> which provides enhanced robustness against such instabilities and enables reliable convergence for both state-specific and state-averaged formulations.

In this paper, we present a unified framework for state-specific and state-averaged orbital-optimized CIS and ECIS, including their linear-response extensions based on DCIS. We derive analytic gradients and Hessians of SACIS and SAECIS, formulate efficient and robust optimization strategies using TRAH, and systematically assess the performance of these methods for both weakly and strongly correlated systems. Through benchmark calculations and detailed analyses of convergence behavior and potential energy surfaces, we aim to clarify the respective roles of orbital relaxation, state averaging, and spin projection in low-cost excited-state methods, and to delineate their regimes of applicability.

## II. THEORY

In this work, the occupied and virtual spin orbitals are indicated by  $i, j, k, l$  and  $a, b, c, d$ , respectively. In addition, we will indicate the  $\beta$  orbitals by a bar on each index, when we discuss spatial orbitals. To simplify the aufbau ( $|\Phi_0\rangle$ ) and non-aufbau, singly-excited determinants ( $|\Phi_i^a\rangle$ ), we use  $\mu$  to indicate the excitation manifold, i.e.,  $|\Phi_\mu\rangle = \hat{E}_\mu |\Phi_0\rangle \in \{\Phi_0, \Phi_i^a\}$  for generalized CIS. Capital

letters  $I, J, K, L$  are reserved for electronic states.

### A. State-specific (single-state) orbital-optimized ECIS

#### 1. Energy, gradients, and Hessian

We begin with the state-specific (single-state) formulation. Because the projected formalism reduces straightforwardly to standard (unprojected) CIS by removing the projection operator  $\hat{P}$ , we introduce spin projection from the outset and later recover the non-projected expressions as a special case.

An unperturbed CIS state is given by

$$|0\rangle = \hat{C}|\Phi_0\rangle = \sum_{\mu} c_{\mu}|\Phi_{\mu}\rangle \quad (1)$$

where

$$\hat{C} = \sum_{\mu} c_{\mu}\hat{E}_{\mu} \quad (2)$$

Spin-restricted CIS is obtained by imposing  $c_{ai} = c_{\bar{a}\bar{i}}$ . While this restriction enforces a pure singlet manifold, it can also reduce the flexibility needed to describe near-degeneracy and static correlation effects. Allowing spin-unrestricted single excitations and orbitals can provide the required flexibility, particularly for (near-)degenerate systems, but it typically results in spin-symmetry breaking when the  $\alpha$  and  $\beta$  orbitals are optimized independently. When spin contamination is substantial, the interpretation of excited states becomes difficult—a serious drawback in photochemical applications, where distinguishing spin manifolds is essential (e.g., for studying intersystem crossing and internal conversion).

In spin-extended CIS (ECIS), one can remove the spin contamination in Eq. (1) to fully realize its potential, by employing spin-projection through the projection operator  $\hat{P}$ . In the following, we primarily focus on the spin-extended formulation including  $\hat{P}$ ; however, the non-projected results (without the “E”) can be readily obtained by setting  $\hat{P} \rightarrow 1$ .

Here, our aim lies in constructing a fully variational ECIS wave function for a single state, whose energy is stationary with respect to both orbital rotations and CI amplitudes. Clearly, this can be achieved by introducing a set of variational parameters  $\lambda = (^o\lambda, ^c\lambda)^{\top}$ , where the superscripts  $o$  and  $c$  indicate the orbital-rotation and CI subspaces, respectively. Although the presence of  $\hat{P}$  somewhat complicates the parametrization, the state-specific case remains considerably simpler than the state-averaged formulation discussed later.

First, the orbital rotation part should be parametrized by the exponential of single excitations as usual,

$${}^o\hat{\lambda} = \sum_{ai} {}^o\lambda_{ai}\hat{E}_{ai}^- \quad (3)$$

with

$$\hat{E}_{pq}^- = \hat{E}_{pq} - \hat{E}_{qp} \quad (4)$$

$$\hat{E}_{pq} = a_p^\dagger a_q \quad (5)$$

Here, one only needs to take into account orbital rotations between the occupied and virtual spaces as in Eq. (3), because an ECIS wave function is invariant with respect to occupied-occupied and virtual-virtual rotations.

In passing, we should emphasize that for full optimization orbital rotations must be unrestricted and applied to the broken-symmetry CIS state  $|0\rangle$  before performing spin-projection  $\hat{P}$ , as  $\hat{P}e^{-{}^o\hat{\lambda}}|0\rangle$ . This corresponds to the so-called “variation-after-projection” scheme, which allows for full optimization of the ansatz. Conversely, if orbital rotations were to be applied after projection, i.e.,  $e^{-{}^o\hat{\lambda}}\hat{P}|0\rangle$ , the operator  ${}^o\hat{\lambda}$  would need to be spin-adapted, making it impossible to vary the degree of symmetry breaking in  $|0\rangle$  and  $|\Phi_{\mu}\rangle$ .

Second, the variational parametrization for the CI space is achieved by adding

$$|^c\lambda\rangle = {}^c\hat{\lambda}|0\rangle = \sum_{\mu} {}^c\lambda_{\mu}|\Phi_{\mu}\rangle \quad (6)$$

with

$${}^c\hat{\lambda} = \sum_{\mu} {}^c\lambda_{\mu}\hat{E}_{\mu} \quad (7)$$

which takes the same form as Eq.(1) but with the CI coefficients  $c$  replaced by  ${}^c\lambda_{\mu}$ . It is advantageous to keep the CI perturbation  ${}^c\lambda$  within the complement of the reference subspace created by  $c$ , especially in the presence of both spin-projection and orbital rotation; namely, we require  $\hat{P}e^{-{}^o\hat{\lambda}}|0\rangle \perp \hat{P}e^{-{}^o\hat{\lambda}}|^c\lambda\rangle$ . This condition would be easily satisfied independent of  ${}^o\lambda$  if spin-projection is not present, because of the unitarity of  $e^{-{}^o\hat{\lambda}}$ . However, as the spin-projection operator  $\hat{P}$  does not commute with (spin-unrestricted) orbital rotations, we generally have

$$e^{{}^o\hat{\lambda}}\hat{P}e^{-{}^o\hat{\lambda}} \neq 1, \quad (8)$$

and one needs to explicitly remove the CI-unperturbed ECIS wave function  $\hat{P}e^{-{}^o\hat{\lambda}}|0\rangle$  from  $\hat{P}e^{-{}^o\hat{\lambda}}|^c\lambda\rangle$  while still allowing for orbital rotation. This can be accomplished by introducing the following projector:

$$\hat{Q} = 1 - \frac{\hat{P}e^{-{}^o\hat{\lambda}}|0\rangle\langle 0|e^{{}^o\hat{\lambda}}\hat{P}}{\langle 0|e^{{}^o\hat{\lambda}}\hat{P}e^{-{}^o\hat{\lambda}}|0\rangle} \quad (9)$$

This definition of  $\hat{Q}$  correctly accounts for the change in the normalization (denominator) and ensures a clear separation between the parameter spaces  $\{{}^o\lambda_{ai}\}$  and  $\{{}^c\lambda_{\mu}\}$ .

These requirements lead us to the following parameterization of the state-specific, orbital-optimized ECIS (SSECIS),

$$|\tilde{0}[\lambda]\rangle = \hat{P}e^{-{}^o\hat{\lambda}} \left( |0\rangle + e^{{}^o\hat{\lambda}}\hat{Q}e^{-{}^o\hat{\lambda}}|^c\lambda\rangle \right) \quad (10)$$

which can be optimized by finding the stationary point of the following energy expectation value:

$$E[\boldsymbol{\lambda}] = \frac{\langle \tilde{0}[\boldsymbol{\lambda}] | \hat{H} | \tilde{0}[\boldsymbol{\lambda}] \rangle}{\langle \tilde{0}[\boldsymbol{\lambda}] | \tilde{0}[\boldsymbol{\lambda}] \rangle} \quad (11)$$

It is important to note that the norm of  $|\tilde{0}\rangle$  is generally not preserved, i.e.,  $\langle \tilde{0}[\boldsymbol{\lambda}] | \tilde{0}[\boldsymbol{\lambda}] \rangle \neq 1$ . Nevertheless, it is convenient to choose the unperturbed state  $|\tilde{0}[0]\rangle \equiv \hat{P}|0\rangle$  to be normalized,  $\langle 0 | \hat{P} | 0 \rangle = 1$ , simplifying the derivation. For example,

$$E[0] = \langle 0 | \hat{H} \hat{P} | 0 \rangle \equiv E_0 \quad (12)$$

A stationary point of Eq. (11) is where the first derivatives of the energy with respect to  ${}^o\lambda_{ai}$  and  ${}^c\lambda_\mu$  around

$\boldsymbol{\lambda} = \mathbf{0}$  are both sufficiently small. They are, respectively,

$${}^o g_{ai} = \left. \frac{\partial E[\boldsymbol{\lambda}]}{\partial {}^o\lambda_{ai}} \right|_{\boldsymbol{\lambda}=\mathbf{0}} = \langle 0 | [\hat{E}_{ai}^-, (\hat{H} - E_0)\hat{P}] | 0 \rangle \quad (13)$$

$${}^c g_\mu = \left. \frac{\partial E[\boldsymbol{\lambda}]}{\partial {}^c\lambda_\mu} \right|_{\boldsymbol{\lambda}=\mathbf{0}} = \langle 0 | (\hat{H} - E_0)\hat{P} | \Phi_\mu \rangle + \langle \Phi_\mu | (\hat{H} - E_0)\hat{P} | 0 \rangle \quad (14)$$

If the spin-projection is neglected, this reduces exactly to the excited state mean-field theory<sup>29–32</sup>.

While gradient minimization often converges to the desired state if the initial state  $|0\rangle$  is sufficiently close to the target, it is still advantageous to incorporate the second derivatives especially for the ground state optimization when the HF state is not qualitatively accurate. Recently, one of us has reported the electronic Hessian for the (non-projected) CIS energy<sup>36</sup>. Here, we generalize that result to ECIS by including the spin-projection operator:

$$\begin{aligned} {}^{oo} H_{ai,bj} &= \left. \frac{\partial^2 E}{\partial {}^o\lambda_{ai} \partial {}^o\lambda_{bj}} \right|_{\boldsymbol{\lambda}=\mathbf{0}} \\ &= \frac{1}{2} \langle 0 | [\hat{E}_{ai}^-, [\hat{E}_{bj}^-, (\hat{H} - E_0)\hat{P}]] | 0 \rangle + \frac{1}{2} \langle 0 | [\hat{E}_{bj}^-, [\hat{E}_{ai}^-, (\hat{H} - E_0)\hat{P}]] | 0 \rangle - {}^o g_{ai} \langle 0 | [\hat{E}_{bj}^-, \hat{P}] | 0 \rangle - {}^o g_{bj} \langle 0 | [\hat{E}_{ai}^-, \hat{P}] | 0 \rangle \end{aligned} \quad (15)$$

$$\begin{aligned} {}^{oc} H_{ai,\mu} &= \left. \frac{\partial^2 E}{\partial {}^o\lambda_{ai} \partial {}^c\lambda_\mu} \right|_{\boldsymbol{\lambda}=\mathbf{0}} \\ &= \langle 0 | [\hat{E}_{ai}^-, (\hat{H} - E_0)\hat{P}] | \Phi_\mu \rangle + \langle \Phi_\mu | [\hat{E}_{ai}^-, (\hat{H} - E_0)\hat{P}] | 0 \rangle - {}^o g_{ai} (\langle 0 | \hat{P} | \Phi_\mu \rangle + \langle \Phi_\mu | \hat{P} | 0 \rangle) - {}^c g_\mu \langle 0 | [\hat{E}_{ai}^-, \hat{P}] | 0 \rangle \end{aligned} \quad (16)$$

$$\begin{aligned} {}^{cc} H_{\mu\nu} &= \left. \frac{\partial^2 E}{\partial {}^c\lambda_\mu \partial {}^c\lambda_\nu} \right|_{\boldsymbol{\lambda}=\mathbf{0}} \\ &= \langle \Phi_\mu | \hat{Q}(\hat{H} - E_0)\hat{P}\hat{Q} | \Phi_\nu \rangle + \langle \Phi_\nu | \hat{Q}(\hat{H} - E_0)\hat{P}\hat{Q} | \Phi_\mu \rangle \end{aligned} \quad (17)$$

For real orbitals, these can be recast as

$$\begin{aligned} {}^{oo} H_{ai,bj} &= \mathcal{P}(ai)\mathcal{P}(bj)(2 {}^{oo} A_{ai,bj} + {}^{oo} B_{ai,bj} + {}^{oo} B_{bj,ai}) \\ &\quad + 2 {}^o g_{ai} \mathcal{P}(bj) P_{bj} + 2 {}^o g_{bj} \mathcal{P}(ai) P_{ai} \end{aligned} \quad (18)$$

$$\begin{aligned} {}^{oc} H_{ai,\mu} &= -2\mathcal{P}(ai)({}^{oc} A_{ai,\mu} + {}^{oc} B_{ai,\mu} - {}^c g_\mu P_{ai}) \\ &\quad - 2 {}^o g_{ai} \langle 0 | \hat{P} | \Phi_\mu \rangle \end{aligned} \quad (19)$$

$${}^{cc} H_{\mu\nu} = 2 {}^{cc} A_{\mu\nu} - {}^c g_\mu \langle 0 | \hat{P} | \Phi_\nu \rangle - {}^c g_\nu \langle 0 | \hat{P} | \Phi_\mu \rangle \quad (20)$$

where  $\mathcal{P}(pq)$  is the anti-symmetrizer between  $p$  and  $q$ ,

and

$${}^{oo} A_{ai,bj} = \langle 0 | \hat{E}_{ia} (H - E_0) \hat{P} \hat{E}_{bj} | 0 \rangle, \quad (21)$$

$${}^{oc} A_{ai,\mu} = \langle 0 | \hat{E}_{ia} (\hat{H} - E_0) \hat{P} | \Phi_\mu \rangle, \quad (22)$$

$${}^{cc} A_{\mu\nu} = \langle \Phi_\mu | (\hat{H} - E_0) \hat{P} | \Phi_\nu \rangle, \quad (23)$$

$${}^{oo} B_{ai,bj} = \langle 0 | (\hat{H} - E_0) \hat{P} \hat{E}_{ai} \hat{E}_{bj} | 0 \rangle, \quad (24)$$

$${}^{oc} B_{ai,\mu} = \langle 0 | (\hat{H} - E_0) \hat{P} \hat{E}_{ai} | \Phi_\mu \rangle, \quad (25)$$

$$P_{pq} = \langle 0 | \hat{E}_{qp} \hat{P} | 0 \rangle. \quad (26)$$

Note that  ${}^{oo} A_{ai,bj}$  and  ${}^{oo} B_{ai,bj}$  resemble the familiar matrices that appear in the Hartree-Fock electronic Hessian. However, they contain  $\hat{P}$  and the de-excitation parts such as  ${}^{oo} A_{ia,jb}$  and  ${}^{oo} B_{ia,jb}$  that are nonzero, as opposed to Hartree-Fock. Especially, the density matrix-like quantity  $P_{pq}$  is asymmetric in the presence of  $\hat{P}$ .

The explicit working equations to evaluate these terms are given in the Supplemental Information.

Using the above results, quadratic convergence methods such as the Newton method and the trust-region augmented Hessian method can be formulated. We defer the discussion to Sec.III A where the latter method is discussed using more generalized state-averaged ECIS.

## 2. Treatment of other excited states: double CIS

One of the main disadvantages of SSCIS/SSECIS is that orbitals are optimized solely with respect to a particular CIS/ECIS state, which significantly deteriorates the description of other CIS/ECIS states. For instance, when the ground state is optimized using generalized CIS/ECIS, the other solutions, which are orthogonal to each other and thus correspond to excited states, exhibit excessively high energies, leading to a substantial overestimation of excitation energies. In the ground state SSE-CIS, the overestimation is even more pronounced. To avoid this issue, one could perform state-specific calculations independently for different states of interest in the same spirit as  $\Delta$  SCF; however, the use of different orbitals makes the optimized states non-orthogonal to each other, complicating the computation of inter-state quantities such as transition dipole moments.

To address this challenge, we exploit the recently proposed double CIS scheme.<sup>36</sup> DCIS aims to incorporate orbital relaxation effects in CIS by performing an additional CIS. Since DCIS involves the diagonalization of its Hamiltonian matrix, the resulting states are orthogonal to one another, in a similar way to CIS.

The wave function ansatz for DCIS is represented by the two-layered CIS:

$$|\Psi_{\text{DCIS}}\rangle = \hat{D}|0\rangle = \hat{D}\hat{C}|\Phi_0\rangle \quad (27)$$

$$\hat{D} = \sum_{pq} d_{pq} \hat{E}_{pq} \quad (28)$$

This scheme corresponds to a linear-response of the CIS state  $|0\rangle$  within the Tamm-Danoff approximation and treats both single excitations and de-excitations with respect to it. Note that we can utilize the orbital-optimized CIS for  $|0\rangle$ .

As described in Ref.<sup>36</sup>, the parameterization in Eq. (27) suffers from numerous redundancies, introducing numerical challenges. However, by reformulating it orthogonal, the same variational space can be constructed by the following parametrization without redundancies:

$$|\Psi_{\text{DCIS}}\rangle = \sum_{ai} d_{ai} \hat{E}_{ai}|0\rangle + \bar{d}_0 |\Phi_0\rangle + \sum_{ai} \bar{d}_{ai} |\Phi_i^a\rangle \quad (29)$$

As is expected, it is straightforward to extend it to spin-extended DCIS (EDCIS) simply by adding  $\hat{P}$ . The variational condition for DCIS/EDCIS is given by the follow-

ing equations:

$$\langle 0 | \hat{E}_{ia} (\hat{H} - E_{\text{EDCIS}}) \hat{P} | \Psi_{\text{DCIS}} \rangle = 0 \quad (30)$$

$$\langle \Phi_\mu | (\hat{H} - E_{\text{EDCIS}}) \hat{P} | \Psi_{\text{DCIS}} \rangle = 0 \quad (31)$$

Again, for the standard DCIS, one can remove  $\hat{P}$  from the equations. The EDCIS Hamiltonian matrix—the second derivative of the energy expectation value with respect to  $\{d_{ai}, \bar{d}_0, \bar{d}_{ai}\}$ —parallels the electronic Hessian matrix  $\mathbf{H}$ . Namely,

$$\begin{aligned} & \sum_{bj} (^{\text{oo}}A_{ai,bj} - \omega ^{\text{oo}}S_{ai,bj}) d_{bj} \\ & + \sum_{\nu} (^{\text{oc}}A_{ai,\nu} - \omega ^{\text{oc}}S_{ai,\nu}) \bar{d}_{\nu} = 0 \end{aligned} \quad (32)$$

$$\begin{aligned} & \sum_{bj} (^{\text{co}}A_{\mu,bj} - \omega ^{\text{co}}S_{\mu,bj}) d_{bj} \\ & + \sum_{\nu} (^{\text{cc}}A_{\mu\nu} - \omega ^{\text{cc}}S_{\mu\nu}) \bar{d}_{\nu} = 0 \end{aligned} \quad (33)$$

$$(34)$$

where  $\omega = E_{\text{EDCIS}} - E_0$  is the relaxation or excitation energy, and  $\mathbf{S}$  is the overlap matrix,

$${}^{\text{oo}}S_{ai,bj} = \langle 0 | \hat{E}_{ia} \hat{P} \hat{E}_{bj} | 0 \rangle \quad (35)$$

$${}^{\text{oc}}S_{ai,\mu} = \langle 0 | \hat{E}_{ia} \hat{P} | \Phi_\mu \rangle \quad (36)$$

$${}^{\text{cc}}S_{\mu\nu} = \langle \Phi_\mu | \hat{P} | \Phi_\nu \rangle \quad (37)$$

Then,  $\omega$  can be obtained by solving the generalized eigenvalue problem

$$\begin{pmatrix} {}^{\text{oo}}\mathbf{A} & {}^{\text{oc}}\mathbf{A} \\ {}^{\text{co}}\mathbf{A} & {}^{\text{cc}}\mathbf{A} \end{pmatrix} \begin{pmatrix} \mathbf{d} \\ \bar{\mathbf{d}} \end{pmatrix} = \omega \begin{pmatrix} {}^{\text{oo}}\mathbf{S} & {}^{\text{oc}}\mathbf{S} \\ {}^{\text{co}}\mathbf{S} & {}^{\text{cc}}\mathbf{S} \end{pmatrix} \begin{pmatrix} \mathbf{d} \\ \bar{\mathbf{d}} \end{pmatrix} \quad (38)$$

Using the ground state of orbital-optimized SSCIS or SSECIS for  $|0\rangle$ , the lowest eigenvalue of DCIS and EDCIS is zero ( $\omega = 0$ ) as a consequence of the generalized Brillouin condition. This indicates that no additional orbital relaxation effect occurs for the ground state. In contrast, the higher-lying DCIS/EDCIS solutions correspond to the relaxed excited states relative to the SSE-CIS ground state, since EDCIS is, in essence, a linear-response method that performs CIS on top of SSE-CIS. This property is therefore expected to lead to improved excitation energies when EDCIS is combined with an SSECIS reference.

Finally, the structure of Eq. 38 strongly suggests a close connection to a time-dependent formulation involving the  $\mathbf{B}$  matrix, where DCIS is obtained as its Tamm-Danoff approximation. Although such a relationship between DCIS and time-dependent CIS appears natural, we do not pursue this direction further in the present work.

## B. State-averaged (multistate) formalism

### 1. Energy, gradients, and Hessian

An alternative method for achieving a balanced description of ground and excited states is state-averaged (SA) orbital optimization. This approach is particularly beneficial for systems exhibiting energetic quasi-degeneracy among multiple states, such as conical intersections, as it can treat them on an equal footing. Consequently, the SA optimization has become a standard protocol in the study of photochemistry.

Since both CIS and ECIS diagonalize a truncated Hamiltonian and can thus determine multiple states simultaneously, it is conceptually straightforward to formulate their SA variants, which minimize the averaged energy of  $n$  states  $\{|0_I\rangle; I = 1, \dots, n\}$  or a weighted sum of their energies. For convenience, we choose the CI basis in which both the Hamiltonian and the overlap of the subspace are diagonal:

$$\langle 0_I | \hat{H} \hat{P} | 0_J \rangle = E_{0,I} \delta_{IJ}, \quad (39)$$

$$\langle 0_I | \hat{P} | 0_J \rangle = \delta_{IJ} \quad (40)$$

where  $E_{0,I}$  is the energy of  $\hat{P}|0_I\rangle$ .

It should first be noted that, for multistate calculations of standard MCSCF (without  $\hat{P}$ ) such as SA-CASSCF, one frequently employs the following parameterization:<sup>50</sup>

$$|\tilde{0}_I\rangle = e^{-\circ\hat{\lambda}} e^{-\hat{\mathcal{R}}} |0_I\rangle \quad (41)$$

where  $\hat{\mathcal{R}}$  is a linear combination of state-transfer operators,

$$\hat{\mathcal{R}} = \sum_{K>J, J \leq n} \mathcal{R}_{KJ} (|0_K\rangle\langle 0_J| - |0_J\rangle\langle 0_K|) \quad (42)$$

This formulation ensures the orthogonality.

However, it is important to point out that the presence of the projection operator  $\hat{P}$  no longer allows for this convenient parameterization, as broken-symmetry states  $\{|0_I\rangle\}$  are not orthonormal to each other by themselves: they only become so when projected by  $\hat{P}$ , as shown in Eq.(40). If  $\hat{P}$  is explicitly incorporated into  $\hat{\mathcal{R}}$ , so that each of  $|0_K\rangle$  and  $|0_J\rangle$  is projected and orthonormal, then the situation becomes significantly more complicated. This is because, as described in Section II A 1, the orbital rotation  $e^{-\circ\hat{\lambda}}$  must be applied to  $|0_I\rangle$  but not to  $\hat{P}|0_I\rangle$ , and these two requirements cannot be satisfied simultaneously.

Due to this issue, we instead parameterize the  $I$ th ECIS state as

$$|\tilde{0}_I[\circ\boldsymbol{\lambda}, ^c\boldsymbol{\lambda}^I]\rangle = \hat{P}e^{-\circ\hat{\lambda}}|0_I\rangle + \hat{P}\hat{Q}[\circ\boldsymbol{\lambda}]e^{-\circ\hat{\lambda}}|^c\boldsymbol{\lambda}^I\rangle \quad (43)$$

analogously to the SS scheme, Eq. (10). To ensure the orthogonality between different states under perturbation

of CI coefficients (but not orbital rotation), we introduce the following generalized projection operator

$$\hat{Q} = 1 - \sum_{IJ} \hat{P}e^{-\circ\hat{\lambda}}|0_I\rangle\langle N[\circ\boldsymbol{\lambda}]^{-1}]_{IJ} \langle 0_J|e^{\circ\hat{\lambda}}\hat{P} \quad (44)$$

where

$$(N[\circ\boldsymbol{\lambda}])_{IJ} = \langle 0_I | e^{\circ\hat{\lambda}} \hat{P} e^{-\circ\hat{\lambda}} | 0_J \rangle \quad (45)$$

is the state overlap Eq.(40) generalized with orbital rotation, and therefore is dependent on  $\circ\boldsymbol{\lambda}$ . Therefore,  $\hat{Q}$  explicitly projects out the residual CI space that is orthogonal to the CI-unperturbed states  $\{\hat{P}e^{-\circ\hat{\lambda}}|0_I\rangle\}$ . However, as is the case in the SS formalism, even the parametrization of Eq. (43) is still not sufficient to preserve the norm and orthogonality of the ECIS states under the perturbation of  $\circ\boldsymbol{\lambda}$ . In fact, it violates the orthogonality between different orbital-perturbed ECI states when  $\circ\boldsymbol{\lambda} \neq 0$ ,

$$\langle \tilde{0}_I | \tilde{0}_J \rangle \neq \delta_{IJ} \quad (46)$$

due to the first term in Eq. (43). This clearly indicates that the orbital perturbation in  $|\tilde{0}_J\rangle$  implicitly affects the energy of other states  $|\tilde{0}_I\rangle$ ,  $E_I$ , even though the latter is not an explicit function of the former, and vice versa. Therefore, the simple average of each energy expectation value,  $\frac{1}{n} \sum_I \langle \tilde{0}_I | \hat{H} | \tilde{0}_I \rangle / \langle \tilde{0}_I | \tilde{0}_I \rangle$ , is not a proper function to optimize for state-averaged ECIS. In other words, the energy expectation value does not represent the true orthogonal energy  $E_I$ , and it is necessary to properly account for the couplings between  $|\tilde{0}_I\rangle$  and  $|\tilde{0}_J\rangle$ ,

$$\mathcal{H}_{IJ}[\circ\boldsymbol{\lambda}, ^c\boldsymbol{\lambda}^I, ^c\boldsymbol{\lambda}^J] = \langle \tilde{0}_I | [\circ\boldsymbol{\lambda}, ^c\boldsymbol{\lambda}^I] | \hat{H} | \tilde{0}_J | [\circ\boldsymbol{\lambda}, ^c\boldsymbol{\lambda}^J] \rangle \quad (47)$$

$$\mathcal{N}_{IJ}[\circ\boldsymbol{\lambda}, ^c\boldsymbol{\lambda}^I, ^c\boldsymbol{\lambda}^J] = \langle \tilde{0}_I | [\circ\boldsymbol{\lambda}, ^c\boldsymbol{\lambda}^I] | \tilde{0}_J | [\circ\boldsymbol{\lambda}, ^c\boldsymbol{\lambda}^J] \rangle \quad (48)$$

Note that in standard non-projected methods such as SA-CIS, orthogonality is guaranteed, i.e.,  $\mathcal{N}_{IJ} = \delta_{IJ}$ , because of the unitarity of  $e^{-\circ\hat{\lambda}}$ . The emergence of the non-orthogonal property in spin-projection methods is attributed to the fact that  $e^{\circ\hat{\lambda}} \hat{P} e^{-\circ\hat{\lambda}} \neq \hat{P}$ . One could avoid this non-orthogonal issue by introducing an appropriate projector in our ansatz Eq.(43) to ensure orthogonality, but this would introduce significant complications. Nevertheless, we can simplify the formulation if we focus on the *averaged* energy rather than a weighted sum of energies.

The energy  $E_I$  can be defined only after explicit orthogonalization, which involves solving the generalized eigenvalue problem

$$\mathcal{H}\mathbf{V} = \mathcal{N}\mathbf{VE} \quad (49)$$

where  $\mathcal{H}$  and  $\mathcal{N}$  are the Hamiltonian matrix and overlap matrix in the subspace  $|\tilde{0}_I\rangle$ ; ( $I = 1, \dots, n$ ),  $\mathbf{V}$  is the eigenvector, and  $\mathbf{E}$  is the diagonal matrix composed of  $E_I$ . This allows us to express the averaged energy as

$$E_{\text{ave}} = \frac{\text{Tr}[\mathcal{N}^{-1} \mathcal{H}]}{n} \quad (50)$$

where we should stress that the numerator corresponds to the sum of the energies  $\sum_I E_I = \text{Tr}[\mathbf{E}]$ , as evident from Eq.(49). This way,  $E_{\text{ave}}$  can be written as an explicit function of the parameter set  $\boldsymbol{\lambda} = (^o\boldsymbol{\lambda}, ^c\boldsymbol{\lambda}^1, ^c\boldsymbol{\lambda}^2, \dots, ^c\boldsymbol{\lambda}^n)^\top$ .

The first derivatives are quite similar to those of the state-specific scheme. It is straightforward to find

$$^o g_{ai} \equiv \frac{\partial E_{\text{ave}}}{\partial ^o \lambda_{ai}} \Big|_{\boldsymbol{\lambda}=0} = \frac{1}{n} \sum_I ^o g_{ai}^{I* I} \quad (51)$$

$$\begin{aligned} ^c g_\mu^I &\equiv \frac{\partial E_{\text{ave}}}{\partial ^c \lambda_\mu^I} \Big|_{\boldsymbol{\lambda}=0} \\ &= \frac{1}{n} \left( \langle 0_I | (\hat{H} - E_{0,I}) \hat{P} | \Phi_\mu \rangle + \langle \Phi_\mu | (\hat{H} - E_{0,I}) \hat{P} | 0_I \rangle \right) \end{aligned} \quad (52)$$

where

$$^o g_{ai}^{I* K} = \langle 0_I | [\hat{E}_{ai}^-, (\hat{H} - E_{0,I}) \hat{P}] | 0_K \rangle \quad (53)$$

is the gradient-like coupling term. Here, the star on  $I$  indicates the reference energy is that for  $I$  (i.e.,  $E_{0,I}$ ).

The second derivatives are important not only for orbital optimization but also for geometry optimization (nuclear gradients), as will be discussed in a separate paper.<sup>51</sup> Again, in real orbitals,

$$\begin{aligned} ^{oo} H_{ai,bj} &= \frac{1}{n} \sum_I \left( \mathcal{P}(ai) \mathcal{P}(bj) (2^{oo} A_{ai,bj}^I + ^{oo} B_{ai,bj}^I \right. \\ &\quad \left. + ^{oo} B_{bj,ai}^I) + \sum_K (^o g_{ai}^{I* K} \mathcal{P}(bj) (P_{bj}^{IK} + P_{bj}^{KI}) \right. \\ &\quad \left. + ^o g_{bj}^{I* K} \mathcal{P}(ai) (P_{ai}^{IK} + P_{ai}^{KI})) \right) \end{aligned} \quad (54)$$

$$\begin{aligned} ^{oc} H_{ai,\mu}^I &= -\frac{2}{n} \mathcal{P}(ai) (^{oc} A_{ai,\mu}^I + ^{oc} B_{ai,\mu}^I - ^c g_\mu^I P_{ai}^I) \\ &\quad - \frac{2}{n} \sum_K ^o g_{ai}^{I* K} \langle 0_K | \hat{P} | \Phi_\mu \rangle \end{aligned} \quad (55)$$

$$\begin{aligned} ^{cc} H_{\mu\nu}^{IJ} &= \frac{\delta_{IJ}}{n} \left( 2^{cc} A_{\mu\nu}^I - \sum_K \left( ^c g_\mu^K \langle 0_K | \hat{P} | \Phi_\nu \rangle \right. \right. \\ &\quad \left. \left. + ^c g_\nu^K \langle 0_K | \hat{P} | \Phi_\mu \rangle + 2(E_{0,K} - E_{0,I}) \langle \Phi_\mu | \hat{P} | 0_K \rangle \langle 0_K | \hat{P} | \Phi_\nu \rangle \right) \right) \end{aligned} \quad (56)$$

where  $^{oo} A_{ai,bj}^I$ , etc., are similarly defined for the  $I$ th state, and  $P_{ai}^{IK}$  is the transition density matrix between  $I$  and  $K$ ,

$$P_{pq}^{IK} = \langle 0_I | \hat{E}_{qp} \hat{P} | 0_K \rangle \quad (57)$$

It is clear that in this basis there is no direct coupling between  $I$ th and  $J$ th states through their CI coefficients, and they are indirectly coupled with each other through the  $oc$ -block. In other words, the CI part of the Hessian

is block-diagonal. The structure of the Hessian matrix is therefore as follows:

$$\mathbf{H} = \begin{pmatrix} ^{oo}\mathbf{H} & ^{oc}\mathbf{H}^1 & ^{oc}\mathbf{H}^2 & \cdots & ^{oc}\mathbf{H}^n \\ ^{co}\mathbf{H}^1 & ^{cc}\mathbf{H}^{11} & \mathbf{0} & \cdots & \mathbf{0} \\ ^{co}\mathbf{H}^2 & \mathbf{0} & ^{cc}\mathbf{H}^{22} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ ^{co}\mathbf{H}^n & \mathbf{0} & \mathbf{0} & \cdots & ^{cc}\mathbf{H}^{nn} \end{pmatrix} \quad (58)$$

One caveat is that this Hessian contains a few zero eigenvalues because of the redundant parameterization introduced in Eq. (43).

We have summarized the detailed derivation and the working equations in the Supplemental Information.

### III. SCF ALGORITHM

In this section, we discuss two algorithms for orbital optimization in SACIS and SAECIS.

#### A. Newton method and trust-region augmented Hessian method

In the Newton method, the parameters are updated by solving

$$\mathbf{H}\boldsymbol{\lambda} = -\mathbf{g} \quad (59)$$

where

$$\mathbf{g} = (^o\mathbf{g}, ^c\mathbf{g}^1, \dots, ^c\mathbf{g}^n)^\top \quad (60)$$

However, this update is often not optimal when the states are (on average) far from the minimum, where the energy functional is not well approximated by a quadratic form.<sup>50</sup> Instead, the trust-region augmented Hessian (TRAH) algorithm diagonalizes the augmented Hessian,

$$\begin{pmatrix} 0 & \alpha\mathbf{g}^\top \\ \alpha\mathbf{g} & \mathbf{H} \end{pmatrix} \begin{pmatrix} 1 \\ \boldsymbol{\lambda}(\alpha) \end{pmatrix} = \mu \begin{pmatrix} 1 \\ \boldsymbol{\lambda}(\alpha) \end{pmatrix} \quad (61)$$

where  $\mu$  is the lowest eigenvalue of the augmented Hessian (AH) matrix and acts as a level shift. The scaling factor  $\alpha$  confines the update  $\boldsymbol{\lambda}$  within a trust radius  $h$ , thereby avoiding abrupt changes:

$$\frac{1}{\alpha^2} \|\boldsymbol{\lambda}\|^2 \leq h^2 \quad (62)$$

Further details of the TRAH method can be found in the literature.<sup>48,49,52</sup>

The TRAH algorithm has a two-level structure. In each macro iteration, the gradient  $\mathbf{g}$  and Hessian  $\mathbf{H}$  of the given CIS/ECIS states are constructed, and the AH is diagonalized using an iterative eigensolver such as the Davidson method.<sup>53</sup> The iterative scheme of Davidson

TABLE I: Number of Fock-like matrices  $N_{\text{Fock}}$  required for orbital optimization and CI iterations in SACIS and SAECIS. SAECIS further incurs a prefactor of  $N_g$  due to grid integration.

Method	Iteration type	Operation	SACIS	SAECIS
TRAH	Macro	${}^{\circ}\mathbf{g}^{I^*K}, {}^c\mathbf{g}^I, \mathbf{F}$	$\frac{n(n+1)}{2} + n + 1$	$\frac{n(n+1)}{2} + 2n + 1$
TRAH	Micro	$\mathbf{H}\boldsymbol{\lambda}$	$4n + 1$	$7n + 1$
DIIS	Macro	${}^{\circ}\mathbf{g}^{I^*I}, \mathbf{F}$	$n + 1$	$2n + 1$
DIIS	Micro	${}^c\mathbf{g}^I$	$n$	$n$

helps dynamically determine an appropriate value of  $\alpha$  satisfying Eq.(62). Upon convergence of the micro iterations, the resulting  $\boldsymbol{\lambda}$  is used to update the MO coefficient matrix as

$$\mathbf{C}_{\text{new}} = \mathbf{C}_{\text{old}} \exp(-\boldsymbol{\kappa}) \quad (63)$$

$$\boldsymbol{\kappa} = \begin{pmatrix} \mathbf{0} & -{}^{\circ}\boldsymbol{\lambda}^\top \\ {}^c\boldsymbol{\lambda} & \mathbf{0} \end{pmatrix} \quad (64)$$

while the CI coefficients are updated by adding  ${}^c\boldsymbol{\lambda}^I$ . As noted above, in SAECIS, each projected state  $\tilde{P}|0_I\rangle$  must subsequently be re-orthonormalized.

We note that an exact solution of Eq.(61) in each macro iteration is not required, particularly during the initial macro cycles, since the update  $\boldsymbol{\lambda}$  is only a quadratic approximation. Therefore, we relax the convergence criterion for the micro iterations to  $\|\mathbf{g}_{\text{micro}}\| \leq \beta \|\mathbf{g}\|$ , where  $\mathbf{g}_{\text{micro}}$  denotes the residual vector of the micro iterations and  $0 < \beta < 1$  is a hyper parameter. In this work, we set  $\beta = 0.2$ .

During the SCF optimization, the trust radius  $h$  is adaptively updated to ensure both robustness and efficiency. Specifically, we employ Fletcher's algorithm to control  $h$  at each macro step<sup>54</sup>. Although detailed procedures are described in other works<sup>48,49</sup>, we briefly review the algorithm here in order to introduce our modification.

In Fletcher's algorithm, two energy differences are computed:

$$\Delta E_{\text{actu}} = E_{\text{new}} - E_{\text{old}} \quad (65)$$

$$\Delta E_{\text{pred}} = \mathbf{g}^\top \boldsymbol{\lambda} + \frac{1}{2} \boldsymbol{\lambda}^\top \mathbf{H} \boldsymbol{\lambda} \quad (66)$$

where  $\Delta E_{\text{actu}}$  is the actual energy change and  $\Delta E_{\text{pred}}$  is the predicted change based on the quadratic model. Their ratio  $r = \Delta E_{\text{actu}} / \Delta E_{\text{pred}}$  is used to update  $h$  as follows:

- $r > 0.75$ : scale  $h$  by 1.2.
- $0.25 < r \leq 0.75$ : keep  $h$  unchanged.
- $0 \leq r \leq 0.25$ : scale  $h$  by 0.7.
- $r < 0$ : scale  $h$  by 0.7, reject the orbital update and repeat the micro iterations with the reduced  $h$ .

In the present work, however, even when the energy increases after the orbital update (i.e.,  $r < 0$ ), we do not reject the update. Such energy increases frequently occur when the current solution is near a saddle point, indicating the presence of lower-energy solutions. At a saddle point, without the trust radius constraint (i.e.,  $\alpha = 1$ ), the update  $\boldsymbol{\lambda}$  would become excessively large; however, the trust-region restriction effectively limits the step size and guides the optimization toward lower-energy states. A similar strategy has been adopted in the quadratically convergent algorithm for SUHF,<sup>55</sup> motivated by the same considerations.

Finally, we consider the computational cost of the TRAH algorithm in SACIS and SAECIS for  $n$  states. Because the augmented Hessian in Eq. (61) is diagonalized using the Davidson method, the rate-limiting steps are the construction of the gradient vector  $\mathbf{g}$  (in each macro iteration) and the evaluation of the sigma vector  $\mathbf{H}\boldsymbol{\lambda}$ . Nevertheless, the formal computational scaling of these operations remains  $O(N^4)$ , where  $N$  is the number of orbitals. This scaling is dominated by the construction of several Fock-like matrices, which involve contractions between AO integrals and pseudo-density matrices (see the Supporting Information).

In each macro iteration, the standard Fock matrix  $\mathbf{F}$  for the reference  $|\Phi_0\rangle$  is constructed in the current orbital basis. The evaluation of the CI gradients  ${}^c\mathbf{g}^I$  and the (transition) orbital gradients  ${}^{\circ}\mathbf{g}^{I^*K}$  requires  $n$  and  $n(n+1)/2$  Fock-like matrices, respectively. For SAECIS, there is an additional overhead of  $n$ . Since Eq. (52) is symmetric (assuming real orbitals), and thus only one of the two equivalent contractions is required for the sigma vector of standard CIS and ECIS. In contrast, the TRAH algorithm requires explicit contributions from both terms; therefore, SAECIS needs two Fock-like matrices per state rather than one. Finally, ECIS requires numerical grid integration for spin projection, increasing the cost by a factor of  $N_g$ , where  $N_g$  is the number of grid points.

For the sigma vector in each micro iteration,  $4n + 1$  ( $7n + 1$  for SAECIS) additional Fock-like matrices are constructed. Therefore, although the formal scaling remains  $O(N^4)$ , orbital optimization entails a significantly larger prefactor than single-shot CIS and ECIS calculations, as expected.

## B. DIIS combined with effective Fock

Although quadratic convergence is desirable for orbital optimization, it is also of interest to explore the use of DIIS as an alternative convergence acceleration scheme for SACIS and SAECIS. In conventional SCF procedures, DIIS updates the Fock matrix  $\mathcal{F}$  by minimizing its off-diagonal elements, which vanish at self-consistency according to the Brillouin theorem; the updated orbitals are then obtained by diagonalizing  $\mathcal{F}$ . For SACIS/SAECIS orbital optimization, while the off-diagonal elements are clearly identified with the residual vector, i.e.,  $\mathcal{F}_{ai} \equiv {}^0g_{ai}$ , there is no obvious definition for the diagonal blocks  $\mathcal{F}_{ij}$  and  $\mathcal{F}_{ab}$ .

To circumvent this issue, we define an effective Fock matrix as

$$\mathcal{F}_{ij} = F_{ij}, \quad (67)$$

$$\mathcal{F}_{ab} = F_{ab} + \delta_{ab} \epsilon_{\text{LS}}, \quad (68)$$

where  $F_{pq}$  is the standard Fock matrix constructed from the current HF-like determinant  $|\Phi_0\rangle$ , and  $\epsilon_{\text{LS}}$  is a level-shift parameter introduced to stabilize the orbital update. Diagonalization of the effective Fock matrix,

$$\mathcal{F}\mathbf{U} = \mathbf{U}\epsilon, \quad (69)$$

yields a unitary transformation that mixes occupied and virtual orbitals, and the MO coefficients are updated as

$$\mathbf{C}_{\text{new}} = \mathbf{C}_{\text{old}} \mathbf{U}. \quad (70)$$

This approach has been taken in the SCF of symmetry-projected HF including SUHF.<sup>42,43</sup>

The application of DIIS to SACIS and SAECIS then closely follows the standard DIIS protocol: the effective Fock matrices and corresponding error vectors from several iterations are transformed into a common orbital representation (the AO basis), and an optimal linear combination of the Fock matrices is determined by minimizing the norm of the error vector.

The computational cost of this approach is expected to be lower than that of the TRAH algorithm, since constructing the effective Fock matrix requires only  $n + 1$  Fock-like matrices for SACIS ( $n$  for the orbital gradients of each CIS state and one for the HF-like reference state) and  $2n + 1$  for SAECIS per macro iteration. For the micro iterations (standard CIS and ECIS calculations), the Davidson algorithm requires  $n$  Fock-like matrices to construct the sigma vector. The number of Fock-like matrices,  $N_{\text{Fock}}$ , required for each method is summarized in Table I.

However, it is well known that DIIS is generally ineffective when the initial orbitals are far from convergence. In addition, the effective Fock matrix is well defined only for converged CIS states, for which  ${}^c g_\mu = 0$  and the orbital gradient  ${}^0g_{ai}$  is available. Consequently, the use of DIIS implies a two-step optimization strategy for SACIS and SAECIS, in which the orbital and CI-coefficient optimizations are decoupled.

## C. Point-group symmetry

In spin-projection approaches, point-group symmetry is not explicitly enforced, and the converged wave function is therefore not guaranteed to retain the correct point-group symmetry, since spin-unrestricted methods allow the breaking of spatial symmetry in addition to spin symmetry. To explicitly account for point-group symmetry within the framework of symmetry breaking and restoration, a point-group symmetry projection may be desirable.<sup>43,56,57</sup> However, from our experiences in many cases this is not needed; spin-projection can resurrect the lost spatial symmetry if not all. This means each of the contaminated spin states possesses a different point-group symmetry than the one the designated spin state has. Therefore, we will not perform the point-group symmetry projection, but simply assign each excited state to the dominant point-group irreducible representation (irrep).

To assign the dominant irrep of each state  $|\Psi_I\rangle$ , we evaluate the overlap

$$\omega_\Gamma = \langle \Psi_I | \hat{P}_\Gamma | \Psi_I \rangle \quad (71)$$

where  $\hat{P}_\Gamma$  is the projector to the irrep  $\Gamma$ ,

$$\hat{P}_\Gamma = \sum_{\hat{R}} \chi_\Gamma(\hat{R}) \hat{R} \quad (72)$$

Here,  $\hat{R}$  are the symmetry operations of the point group under consideration, and  $\chi_\Gamma(\hat{R})$  are the characters. Noting that  $\sum_\Gamma \omega_\Gamma = 1$ ,  $\omega_\Gamma$  stands for the weight of the component of the irrep  $\Gamma$  in  $|\Psi_I\rangle$ . In all our calculations below,  $\omega_\Gamma$  is more than 90 % (in most cases 100%) for a certain  $\Gamma$  and almost zero for others, by which we properly assign the irrep to facilitate the direct comparisons with the reference excitations.

## IV. RESULTS

### A. Computational details

All CIS, ECIS, SACIS, SAECIS, DCIS, and EDCIS calculations were performed using the GELLAN suite of programs.<sup>58</sup> For complete-active-space SCF (CASSCF) and equation-of-motion coupled-cluster with singles and doubles (EOM-CCSD), we used Q-Chem.<sup>59</sup> The 6-31G basis set was used for hydrogen fluoride, while aug-cc-pVDZ was employed for all other molecules.

For the TRAH method, the initial value of the trust radius  $h$  was set to 0.3. In the effective Fock approach, simultaneous updates of multiple CIS/ECIS states were carried out using the block-Davidson algorithm with diagonal preconditioning, where the diagonal was approximated without explicit two-electron integrals. The level shift  $\epsilon_{\text{LS}}$  applied to the effective Fock matrix was set to 0.2–0.3. Convergence was assumed when the norm of the

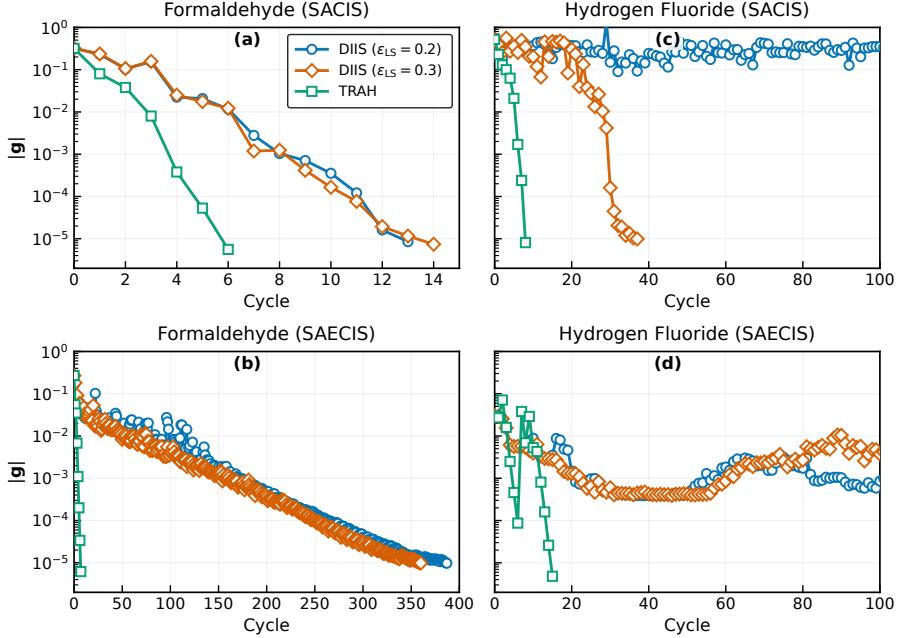


FIG. 1: Comparison between convergence behaviors of DIIS and TRAH. (a) Formaldehyde with SACIS. (b) Formaldehyde with SAECIS. (c) Hydrogen fluoride with SACIS. (d) Hydrogen fluoride with SAECIS.

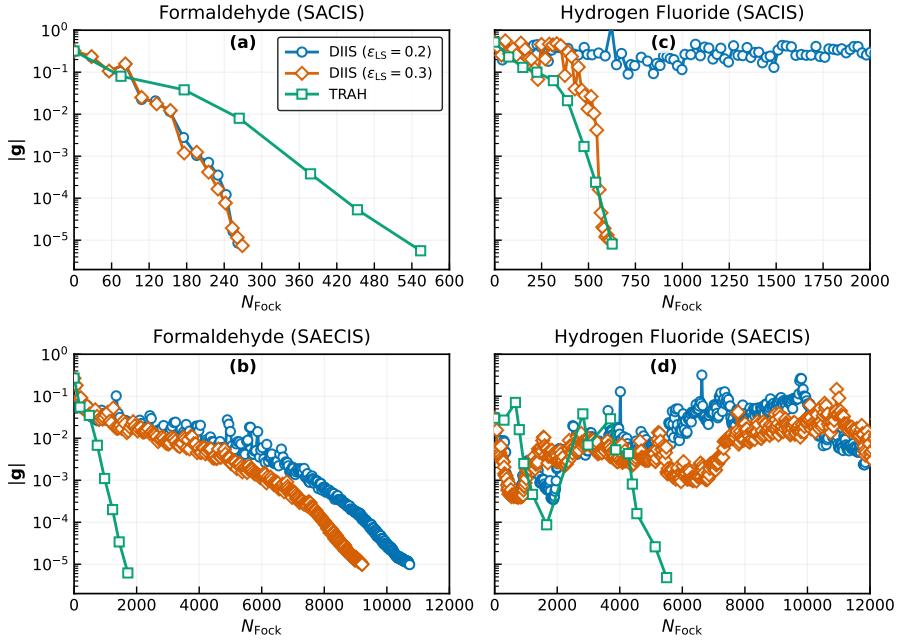


FIG. 2: Same as Figure 1 but as a function of  $N_{\text{Fock}}$ .

corresponding gradient fell below  $10^{-5}$ . For DIIS acceleration, we used eight error vectors for SACIS and five for SAECIS. The latter converges much more slowly than SACIS, and the DIIS inversion often suffers from linear dependence.

For SACIS and DCIS, spin-restricted calculations were enforced so that all computed states were singlets. In

contrast, for spin-projected methods, all parameters were optimized spin-unrestrictedly, and four grid points were used for numerical integration. To perform SACIS and SAECIS calculations, we first converged the corresponding CIS and ECIS states using restricted HF and SUHF orbitals, respectively, and subsequently used them as initial guesses.

For the benchmark calculations in Section IV C, we state-averaged over a fixed manifold of  $n$  states that includes prescribed numbers of roots from each irrep in the benchmark set (including the ground state). In all methods, however, point-group symmetry constraints were not enforced during the orbital optimization. Therefore, for the SA optimization of multiple states with designated symmetry, we employed DIIS and, in each macro iteration, first converged more than  $n$  CIS/ECIS roots in the micro step. From these converged roots, we then identified and selected the lowest-energy solutions with the desired irreps following the protocol described in Section III C, and used only this selected set to construct the SA orbital gradient for the subsequent DIIS update.

## B. Convergence behavior

We first examine the convergence performance of the TRAH method and DIIS for SACIS and SAECIS. The test systems are formaldehyde at its equilibrium geometry and hydrogen fluoride at a stretched bond length of  $R_{\text{H}-\text{F}} = 3.0$ . In all cases, the three lowest states ( $n = 3$ ), including the ground state, were optimized. For formaldehyde, these correspond to the  $n \rightarrow \pi^*$  and  $n \rightarrow 3s$  excited states, while for hydrogen fluoride they correspond to doubly degenerate  $\pi \rightarrow \sigma^*$  states. Both TRAH and DIIS yield the same energies upon successful convergence.

Because the HF ground state of formaldehyde at equilibrium is stable, the two lowest CIS excited states are also expected to behave well. Using these states as the initial guess for SACIS, the effective Fock approach with DIIS converges rapidly for both  $\epsilon_{\text{LS}} = 0.2$  and 0.3, as shown in Figure 1(a). Depending on the level shift, convergence is achieved within 13–14 iterations. In contrast, TRAH converges even faster, within six macro-iterations. However, it should be noted that TRAH requires a larger number of Fock builds ( $N_{\text{Fock}} = 554$ ) than DIIS ( $N_{\text{Fock}} = 262$  for  $\epsilon_{\text{LS}} = 0.2$  and 269 for  $\epsilon_{\text{LS}} = 0.3$ ), because TRAH incurs additional overhead from the computation of  $\sigma$  vectors during micro-iterations (see Table I). Nevertheless, these numbers appear to be reasonable, given that the initial CIS (starting point of SACIS) requires  $N_{\text{Fock}} = 86$ . To enable a fairer comparison of computational cost, Figure 2(a) plots the gradient norm  $\|\mathbf{g}\|$  as a function of  $N_{\text{Fock}}$ .

For SAECIS, one might anticipate behavior similar to SACIS, since the system is only weakly correlated and both SUHF and ECIS converge stably, analogous to HF and CIS. Surprisingly, however, DIIS performs very poorly in this case. As shown in Figure 1(b), DIIS requires 350–400 iterations to reach convergence for SAECIS, corresponding to nearly 10,000 Fock builds (Figure 2(b)). In sharp contrast, TRAH converges within seven macro-iterations, requiring only  $N_{\text{Fock}} = 1719$  (the initial ECIS costs  $N_{\text{Fock}} = 146$ ). These results indicate that TRAH is far more suitable than DIIS for SAECIS.

When the system becomes strongly correlated and the HF reference orbitals are qualitatively incorrect, SACIS can induce substantial orbital changes during optimization. This situation arises for stretched hydrogen fluoride. Using HF orbitals, the initial ground- and excited-state energies are  $E_0 = -99.6243$  and  $E_1 = E_2 = -99.6144$  Hartree, which are significantly higher than the fully optimized SACIS energies,  $E_0 = -99.8585$  and  $E_1 = E_2 = -99.8575$  Hartree. Such large orbital relaxations pose a serious challenge for DIIS. Indeed, for  $\epsilon_{\text{LS}} = 0.2$ , the effective-Fock-based orbital updates are unstable and fail to converge altogether (Figure 1(c)). Increasing the level shift to  $\epsilon_{\text{LS}} = 0.3$  opens the occupied–virtual gap and partially stabilizes the optimization, allowing DIIS to converge. Nevertheless, large oscillations in energies and orbitals are observed during the first ten iterations, indicating intrinsic instability, and a total of 37 iterations are required for convergence. By contrast, TRAH converges smoothly without such oscillations. Although the total number of Fock builds is comparable for the two approaches, TRAH is clearly more reliable in this strongly correlated regime.

Figures 1(d) and 2(d) show that DIIS applied to SAECIS exhibits similar convergence difficulties, albeit for a different underlying reason. In this case, the initial guess lies close to a saddle point on the energy surface. When started from such orbitals, DIIS becomes trapped and fails to locate the lower-energy minimum. By contrast, TRAH initially follows the saddle-point solution but successfully detects the instability in the seventh macro-iteration, as evidenced by an increase in the average energy  $E_{\text{ave}}$ . After transitioning to the lower-energy solution, TRAH converges rapidly.

In summary, DIIS is effective in reducing computational cost for SACIS when the HF orbitals provide a reasonable reference. However, TRAH is considerably more robust and is particularly advantageous for SAECIS and strongly correlated systems. A promising practical strategy may therefore be to employ TRAH in the early stages to obtain a reliable approximate solution, and subsequently switch to DIIS to accelerate convergence at reduced computational cost.

## C. Benchmark calculations for excitation energies of weakly correlated systems

To quantitatively assess the performance of the proposed CIS-based methods, we evaluate their ability to reproduce vertical excitation energies for a diverse set of valence and Rydberg excited states using the benchmark dataset of Loos *et al.*<sup>60</sup>. This dataset comprises singlet excited states of various symmetries across a wide range of small organic molecules. Although none of the molecules in this benchmark exhibits strong static correlation, the availability of highly accurate CCSDT excitation energies as reference values allows for a reliable assessment of the overall accuracy of the methods in weakly

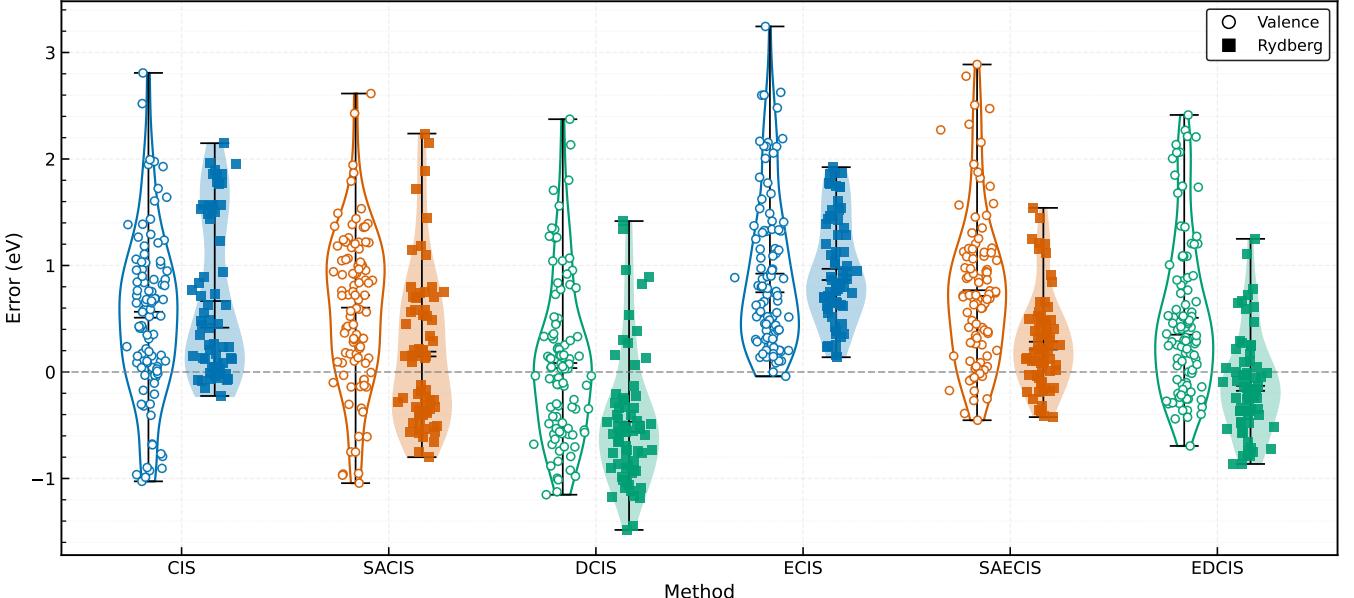


FIG. 3: Excitation energy error from EOM-CCSDT in eV.  $\circ$  and  $\blacksquare$  denote valence and Rydberg excitations.

correlated regimes. Among the test molecules, the two excitations for silylidene were not classified as either valence or Rydberg states and were therefore excluded from both categories.

Figure 3 summarizes the distributions of excitation energy errors for each method relative to the reference EOM-CCSDT values taken from Ref. [60]. It is evident that most methods systematically overestimate excitation energies. The general trend is that the overestimation observed for CIS/ECIS is partially mitigated by the introduction of state averaging (SACIS/SAECIS), and further reduced by the use of the double CI scheme (DCIS/EDCIS). For valence excitations, SACIS does not offer any improvement; in fact, both the ME and MAE deteriorate from 0.51 to 0.60 eV and from 0.73 to 0.78 eV, respectively, as summarized in Table II. This behavior is not unexpected, since SACIS does not incorporate dynamical correlation effects, which constitute the dominant source of error for valence excitations. In contrast, SACIS shows a clear improvement for Rydberg excitations. The mean overestimation is significantly reduced, with the ME decreasing from 0.67 to 0.19 eV. This improvement can be attributed to the nature of Rydberg states, which typically differ substantially from the mean field (HF) electronic environment, and therefore require strong orbital relaxation to balance the description of the ground and excited states. The MAE is also reduced, albeit more modestly, from 0.69 to 0.57 eV, reflecting the continued absence of dynamical correlation.

Similar behavior is expected for DCIS, which also seeks error cancellation between the ground and excited states by incorporating linear-response orbital relaxation effects within a CI framework. While this expectation is largely borne out, the energy lowering achieved by DCIS is more

pronounced than in SACIS, leading to an overall underestimation of Rydberg excitation energies. This difference arises because SACIS optimizes orbitals by averaging over all targeted states, so that no single state is fully optimized. In contrast, DCIS employs state-specific orbital optimization for each excited state, albeit at the first-order level. Consequently, the relaxation effects in DCIS are inherently stronger than in SACIS, particularly when the number of averaged states  $n$  exceeds two. Notably, the excitation energy lowering in DCIS is also observed for valence excitations, resulting in the overall improvement over CIS by 0.1 eV.

One of the most striking observations from this benchmark is that ECIS substantially overestimates excitation energies for all types of excited states considered. Both the ME and MAE of ECIS reach 0.93 eV, rendering its overall performance significantly worse than that of CIS. This deficiency can likely be attributed to an overly biased description of the reference SUHF state relative to the excited states. Although SUHF and ECIS were originally developed to remedy the qualitative failures of HF and CIS in systems with near-degeneracy, their benefits are highly system dependent. Indeed, as shown previously in Fig. 4(d), ECIS yields markedly improved excitation energies over CIS for stretched hydrogen fluoride, where strong static correlation is present. Near equilibrium geometries, however, its accuracy is comparable to that of CIS.

Taken together, these results indicate that for molecules that are only weakly correlated, spin projection provides little to no advantage over CIS and may even degrade excitation energies. It is worth noting that a time-dependent extension of SUHF was shown in the original work to substantially improve the Tamm–Dancoff

ECIS results by incorporating de-excitation channels as effective orbital-relaxation effects.<sup>24</sup> We expect that similar improvements would also be observed for the present benchmark, in close analogy to the systematic lowering of excitation energies found in TDHF and TDDFT.<sup>6,7,61</sup> However, time-dependent SUHF is not variational and is known to produce spurious zero eigenvalues,<sup>24</sup> which complicates the assignment and interpretation of excited states. For this reason, we do not pursue this direction further.

To remedy the inferior performance of ECIS, both state averaging and double CI schemes are effective for improving excitation energies. As shown clearly in Figure 3 and Table II, SAECIS recovers an accuracy comparable to SACIS for valence excitations (MAE = 0.81 eV), while offering a substantially improved description for Rydberg excitations (MAE = 0.39 eV). Again, EDCIS further ameliorates the systematic overestimation and shifts excitation energies downward, reducing the ME for both valence and Rydberg states. Overall, the accuracies of SAECIS and EDCIS are comparable, with EDCIS being slightly more accurate (MAE = 0.54 eV) than SAECIS (MAE = 0.63 eV).

Finally, we emphasize that the conclusions drawn here are specific to weakly correlated regimes. We expect that the proposed methods, particularly SAECIS and EDCIS, will perform more favorably for strongly correlated systems, where orbital relaxation and near-degeneracy effects play a more prominent role. A systematic assessment in such regimes, however, requires appropriate benchmark datasets, which are currently unavailable and will therefore be pursued in future work. In the absence of such a dataset, next we use the bond dissociation of hydrogen fluoride as a stringent qualitative test case for the proposed methods in a strongly correlated regime.

TABLE II: Error statistics (ME/MAE) for all, valence (V), and Rydberg (R) data sets in eV.

Method	All		V		R	
	ME	MAE	ME	MAE	ME	MAE
CIS	0.57	0.71	0.51	0.73	0.67	0.69
SACIS	0.44	0.69	0.60	0.78	0.19	0.57
DCIS	-0.14	0.61	0.08	0.56	-0.47	0.68
ECIS	0.93	0.93	0.92	0.92	0.97	0.97
SAECIS	0.56	0.63	0.77	0.81	0.28	0.39
EDCIS	0.24	0.54	0.51	0.65	-0.13	0.39

#### D. Potential energy curves of hydrogen fluoride

Because the proposed orbital-optimized approaches may be regarded as multiconfigurational SCF-type methods, it is essential to examine the quality of the resulting wave functions in the presence of strong electron correlation.

To this end, we computed the potential energy curves of hydrogen fluoride for the ground state and two  $\pi \rightarrow \sigma^*$  excited states, using the same system as in the previous section. We examined CIS, ECIS, SSECIS, SACIS, and SAECIS, and compared their results with exact full CI (FCI) reference data. For comparison with conventional multireference approaches, we also performed state-averaged CASSCF calculations using both a minimal active space of (6e,4o) and a moderately enlarged active space of (6e,7o).

Figure 4(a) shows the ground-state potential energy curves. As is well known, HF exhibits large errors as the bond is stretched. These errors arise from the lack of static correlation and clearly indicate the necessity of a multideterminantal description for bond dissociation. In contrast, SUHF incorporates static correlation through a superposition of nonorthogonal determinants and thereby recovers the qualitatively correct dissociation behavior.

Somewhat unexpectedly, SACIS is also capable of describing bond dissociation, despite the absence of explicit double excitations that would normally be required to represent the dissociation limit,

$$(c_0|\phi_{\sigma}^2\phi_{\sigma^*}^0\rangle - c_1|\phi_{\sigma}^0\phi_{\sigma^*}^2\rangle)|(\text{Core})\rangle,$$

where  $|(\text{Core})\rangle = |\phi_{\text{F}_{1s}}^2\phi_{\text{F}_{2s}}^2\phi_{\text{F}_{px}}^2\phi_{\text{F}_{py}}^2\rangle$  denotes the closed-shell, non-bonding orbitals. Inspection of the ground-state wave function at large bond distances reveals that SACIS instead favors a localized-orbital description reminiscent of valence bond theory. In this limit, the reference determinant  $|\Phi_0\rangle$  corresponds to the ionic configuration  $\text{H}^+\text{F}^-$ . Single excitations from  $\phi_{\text{F}_{pz}}$  to  $\phi_{\text{H}_{1s}}$  within CIS then yield the correct dissociation limit in the form of a charge-transfer singlet,

$$|\Psi_{S_0}^{\text{SACIS}}\rangle = \frac{1}{\sqrt{2}} \left( |\phi_{\text{H}_{1s}}^{\alpha}\phi_{\text{F}_{pz}}^{\beta}\rangle + |\phi_{\text{H}_{1s}}^{\beta}\phi_{\text{F}_{pz}}^{\alpha}\rangle \right) |(\text{Core})\rangle. \quad (73)$$

As shown in Figure 4(b), the absolute ground-state energy errors of SACIS relative to FCI are comparable to those of SUHF. However, the SACIS curve is more nearly parallel to the FCI reference. This is quantified by the non-parallelity error (NPE), defined as the difference between the maximum and minimum deviations from FCI, which is reduced from 13.9 mHartree for SUHF to 9.6 mHartree for SACIS, as summarized in Table III.

An even more striking feature of SACIS is its ability to correctly describe the exact degeneracy between the ground and excited states in the dissociation limit. Although this behavior was already observed at  $R_{\text{H}-\text{F}} = 3.0$  Å in the preceding section, Figs. 4(c) and 4(d) provide a more detailed analysis. These panels show the excited-state potential energy curves  $E_{S_1}$  (degenerate with  $E_{S_2}$ ) and the deviation of the excitation energy from the FCI reference,  $\Delta\Delta E$ , respectively. As the bond is stretched, most methods—including SACIS—recover the vanishing excitation energy of FCI, corresponding to exact degeneracy between the states ( $\Delta\Delta E = 0$ ). In SACIS, this

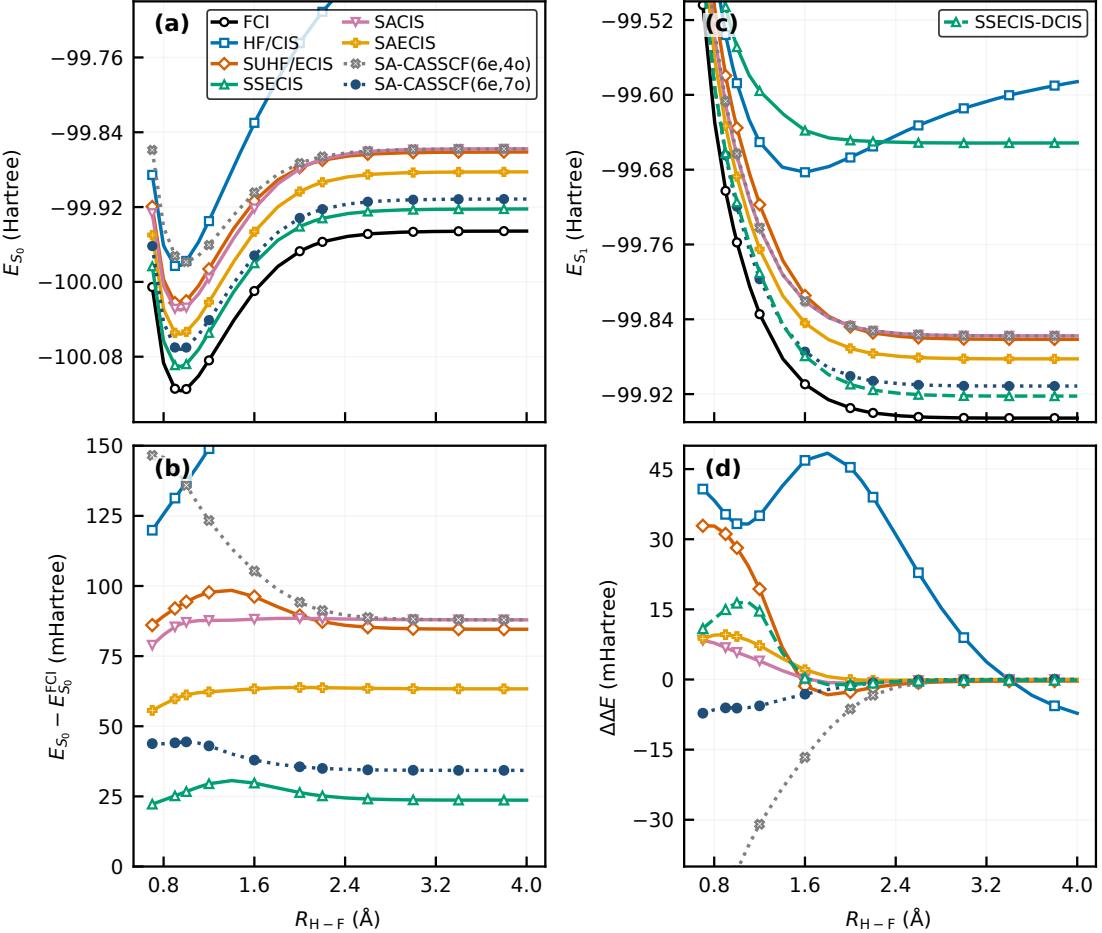


FIG. 4: Potential curves of hydrogen fluoride computed by different methods. (a) Ground state  $S_0$ . (b) Energy error from FCI for  $S_0$ . (c) First excited state  $S_1$ . (d) Error from the FCI excitation energy.

degeneracy arises from charge-transfer excitations from  $\phi_{\text{F}_{px}}$  and  $\phi_{\text{F}_{py}}$  into  $\phi_{\text{H}_{1s}}$ . While SACIS cannot be expected to describe all possible types of degeneracy, these results demonstrate that state averaging alone already captures the essential qualitative physics of bond dissociation and excited-state degeneracy for this system.

Having established the central role of state averaging, we next assess whether spin projection provides additional benefits for the excited-state potential energy surfaces. To this end, we examine ECIS and its state-averaged extension, SAECIS.

Near the equilibrium bond distance, where the electronic structure is only weakly correlated, ECIS offers little improvement over CIS in excitation energies. In this regime, the spin-projected SUHF reference remains strongly biased toward the closed-shell ground state, and spin projection alone cannot compensate for the absence of orbital relaxation. Consequently, ECIS excitation energies around equilibrium remain comparable to those obtained with CIS, consistent with the trends observed in the previous section.

As the bond is stretched and the system enters a strongly correlated regime, the role of spin projection becomes more pronounced. In this limit, ECIS substantially improves upon CIS by restoring the correct symmetry and near-degeneracy between configurations, thereby recovering qualitative features of the excited-state potential energy surfaces that are entirely missed by CIS. At the same time, the success of SACIS demonstrates that these qualitative features can already be captured through state averaging at the CIS level, without explicit spin projection. This indicates that while spin projection can provide a useful complementary mechanism for describing strongly correlated excited states, it is not strictly required when state averaging is properly employed.

This observation naturally motivates the combination of spin projection with state averaging. By optimizing orbitals with respect to an average energy over multiple states, SAECIS reduces the ground-state bias inherent in ECIS and provides a more balanced description of ground and excited states along the entire potential energy sur-

face.

As evidenced by both Figure 4 and Table III, SAECIS performs very similarly to SACIS for this system in a relative sense. Although SAECIS gains additional correlation energy in absolute terms, this effect is nearly constant along the potential energy curves and therefore does not qualitatively alter their shapes. This indicates that the dominant qualitative features of the electronic structure are already captured through state averaging at the CIS level, whereas spin projection provides only additional, but comparatively modest, corrections.

While state averaging provides a balanced and robust description of multiple states, it necessarily sacrifices variational optimality for any single state. It is therefore instructive to contrast SAECIS with a purely state-specific optimization strategy, in which the orbitals are optimized exclusively for the ground state. To this end, we next examine the behavior of SSECIS.

As expected, SSECIS recovers significantly more correlation energy than SAECIS; its total energy is approximately 10 mHartree lower than that of SS-CASSCF(6e,7o) over the entire potential energy curve (not shown). However, excited states obtained naïvely as higher-energy solutions of the underlying ECIS Hamiltonian are highly unstable (Figure 4(c)), leading to severe overestimation of excitation energies. As a consequence, the corresponding  $\Delta\Delta E$  values reach several hundred mHartree and are therefore omitted from Figure 4(d). This failure can be attributed to the strong bias of the optimized orbitals toward ground-state energy minimization.

Introducing the DCIS scheme as a linear-response treatment on top of SSECIS largely remedies this problem. The resulting excitation energies, shown as dashed curves in Figs. 4(c) and 4(d), exhibit substantial improvement, although the NPEs for  $S_1$  and  $\Delta\Delta E$  remain larger than those obtained with SACIS and SAECIS (Table III).

Finally, SA-CASSCF(6e,4o) represents the minimal active space for this system, but it yields large ground-state errors, particularly near the equilibrium bond length, with an NPE of 58.6 mHartree. In contrast, the excited-state energies are reasonably described over the entire range of bond lengths, with an NPE of 7.9 mHartree. Expanding the active space to SA-CASSCF(6e,7o) significantly improves the ground-state description, resulting in overall accuracy comparable to that of SACIS and SAECIS. These results underscore the sensitivity of CASSCF to the choice of active space.

In summary, SACIS and SAECIS provide balanced and reliable descriptions of the ground- and excited-state potential energy surfaces for hydrogen fluoride. A particularly attractive feature of these methods is that they do not require a user-defined active space, unlike CASSCF, allowing them to be applied in a largely black-box manner.

TABLE III: Non-parallelity errors (NPE) in mHartree.

Method	NPE( $S_0$ )	NPE( $S_1$ )	NPE( $\Delta\Delta E$ )
HF/CIS	247.2	199.2	55.6
SUHF/ECIS	13.9	38.9	36.1
SSECIS	8.4	114.9 <sup>a</sup> , 21.2 <sup>b</sup>	113.4 <sup>a</sup> , 17.6 <sup>b</sup>
SACIS	9.6	5.6	9.2
SAECIS	8.2	7.0	9.8
SA-CASSCF(6e,4o)	58.6	7.9	52.5
SA-CASSCF(6e,7o)	10.2	4.1	7.2

<sup>a</sup> Second lowest solution of the underlying CIS Hamiltonian.

<sup>b</sup> Linear response using the EDCIS scheme.

## V. CONCLUSIONS

In this work, we have developed and systematically assessed a family of low-cost CIS-based excited-state methods that incorporate spin projection, orbital relaxation, and state averaging within a unified variational framework. By formulating both state-specific and state-averaged orbital-optimized CIS and ECIS, and by extending the double-CI approach to include spin projection, we have clarified how different physical mechanisms—error cancellation, orbital relaxation, and symmetry restoration—affect excitation energies across distinct correlation regimes. In addition, we have demonstrated that robust optimization strategies are indispensable for these methods: in particular, the TRAH method, realized with the aid of analytic Hessian derived in this work, enables reliable convergence for strongly coupled nonlinear optimization problems arising in state-averaged schemes, where conventional DIIS approaches often fail or converge very slowly.

Benchmark calculations for weakly correlated molecules reveal that state averaging and linear-response-type orbital relaxation are effective in reducing the systematic overestimation of CIS excitation energies, especially for Rydberg states. In contrast, spin projection alone, as realized in ECIS, does not generally improve excitation energies in this regime and may even degrade them due to an overly biased SUHF reference. However, combining spin projection with state averaging or double-CI corrections (SAECIS and EDCIS) restores a balanced description of ground and excited states, yielding excitation energies slightly better than their non-projected counterparts. These results highlight that the usefulness of spin projection is highly regime dependent and must be complemented by appropriate orbital-relaxation mechanisms.

The importance of state-averaging becomes more evident in strongly correlated systems, as illustrated by the potential energy curves of stretched hydrogen fluoride. In this case, conventional CIS fails qualitatively due to the breakdown of the HF reference, whereas SACIS correctly captured the essential physics of near-degeneracy and static correlation. Spin-projection variants, ECIS

and SAECIS, are also successful in describing excited-state potential energy surfaces in bond-breaking regimes, but SAECIS again is more robust as ECIS overestimates excitation energies around weakly correlated region (i.e., equilibrium bond distance).

Finally, we emphasize that all the methods proposed in this work remain mean-field approximations and no dynamical correlation is treated explicitly. Therefore, it is of interest to develop correlation methods like perturbation theory and combine them with these CIS variants to achieve further accuracy in ground and excited state calculations.

## ASSOCIATED CONTENT

### Supplemental information

Detailed derivation of matrix elements for SAECIS and EDCIS, and simplified equations for SACIS and DCIS; excitation energies computed with CIS, SACIS, SAECIS, DCIS, and EDCIS.

## ACKNOWLEDGMENTS

This work was supported by JST FOREST Program, Grant No. JPMJFR223U, and JSPS KAKENHI, Grant No. 25K01733 and 25K22247.

- <sup>1</sup>J. F. Stanton and R. J. Bartlett, “The equation of motion coupled-cluster method. a systematic biorthogonal approach to molecular excitation energies, transition probabilities, and excited state properties,” *The Journal of Chemical Physics* **98**, 7029–7039 (1993).
- <sup>2</sup>A. I. Krylov, “Equation-of-motion coupled-cluster methods for open-shell and electronically excited species: The hitchhiker’s guide to fock space,” *Annual Review of Physical Chemistry* **59**, 433–462 (2008).
- <sup>3</sup>R. J. Bartlett and M. Musiał, “Coupled-cluster theory in quantum chemistry,” *Rev. Mod. Phys.* **79**, 291–352 (2007).
- <sup>4</sup>O. Christiansen, H. Koch, and P. J. Årgensen, “The second-order approximate coupled cluster singles and doubles model cc2,” *Chemical Physics Letters* **243**, 409–418 (1995).
- <sup>5</sup>H. Koch, O. Christiansen, P. Jørgensen, A. M. Sanchez de MerÁjs, and T. Helgaker, “The cc3 model: An iterative coupled cluster approach including connected triples,” *The Journal of Chemical Physics* **106**, 1808–1818 (1997).
- <sup>6</sup>S. Hirata and M. Head-Gordon, “Time-dependent density functional theory within the tamm-dancoff approximation,” *Chemical Physics Letters* **314**, 291–299 (1999).
- <sup>7</sup>A. Dreuw and M. Head-Gordon, “Single-reference ab initio methods for the calculation of excited states of large molecules,” *Chemical Reviews* **105**, 4009–4037 (2005), pMID: 16277369.
- <sup>8</sup>J. E. Subotnik, “Communication: Configuration interaction singles has a large systematic bias against charge-transfer states,” *J. Chem. Phys.* **135**, 071104 (2011).
- <sup>9</sup>M. Head-Gordon, R. J. Rico, M. Oumi, and T. J. Lee, “A doubles correction to electronic excited states from configuration interaction in the space of single substitutions,” *Chemical Physics Letters* **219**, 21–29 (1994).
- <sup>10</sup>A. T. Gilbert, N. A. Besley, and P. M. Gill, “Self-consistent field calculations of excited states using the maximum overlap method (mom),” *J. Phys. Chem. A* **112**, 13164–13171 (2008).
- <sup>11</sup>T. Kowalczyk, T. Tsuchimochi, P.-T. Chen, L. Top, and T. Van Voorhis, “Excitation energies and Stokes shifts from a restricted open-shell Kohn-Sham approach,” *J. Chem. Phys.* **138**, 164101 (2013).
- <sup>12</sup>G. Levi, A. V. Ivanov, and H. J. Ångströmsson, “Variational density functional calculations of excited states via direct optimization,” *J. Chem. Theory Comput.* **16**, 6968 (2020).
- <sup>13</sup>D. Hait and M. Head-Gordon, “Orbital optimized density functional theory for electronic excited states,” *J. Phys. Chem. Lett.* **12**, 4517–4529 (2021).
- <sup>14</sup>F. Koskoski and P.-F. Loos, “State-specific configuration interaction for excited states,” *Journal of Chemical Theory and Computation* **19**, 2258–2269 (2023).
- <sup>15</sup>Y. Damour, A. Scemama, D. Jacquemin, F. Koskoski, and P.-F. Loos, “State-specific coupled-cluster methods for excited states,” *Journal of Chemical Theory and Computation* **20**, 4129–4145 (2024).
- <sup>16</sup>H. Tuckman and E. Neuscamman, “Excited-state-specific pseudoprojected coupled-cluster theory,” *J. Chem. Theory Comput.* **19**, 6160–6171 (2023).
- <sup>17</sup>J. Kussmann, Y. Lemke, A. Weinbrenner, and C. Ochsenfeld, “A constraint-based orbital-optimized excited state method (coox),” *Journal of Chemical Theory and Computation* **20**, 8461–8473 (2024).
- <sup>18</sup>E. Selenius, A. E. Sigurdarson, Y. L. A. Schmerwitz, and G. Levi, “Orbital-optimized versus time-dependent density functional calculations of intramolecular charge transfer excited states,” *J. Chem. Theory Comput.* **20**, 3809–3822 (2024).
- <sup>19</sup>X. Liu and J. E. Subotnik, “The variationally orbital-adapted configuration interaction singles (voa-cis) approach to electronically excited states,” *J. Chem. Theory Comput.* **10**, 1004–1020 (2014).
- <sup>20</sup>P. Norman and A. Dreuw, “Simulating x-ray spectroscopies and calculating core-excited states of molecules,” *Chemical Reviews* **118**, 7208–7248 (2018), pMID: 29894157.
- <sup>21</sup>J. Liu, D. Matthews, S. Coriani, and L. Cheng, “Benchmark calculations of k-edge ionization energies for first-row elements using scalar-relativistic core-valence-separated equation-of-motion coupled-cluster methods,” *Journal of Chemical Theory and Computation* **15**, 1642–1651 (2019), pMID: 30702889.
- <sup>22</sup>K. J. Oosterbaan, A. F. White, and M. Head-Gordon, “Non-orthogonal configuration interaction with single substitutions for core-excited states: An extension to doublet radicals,” *Journal of Chemical Theory and Computation* **15**, 2966–2973 (2019), pMID: 31017781.
- <sup>23</sup>S. M. Garner and E. Neuscamman, “Core excitations with excited state mean field and perturbation theory,” *The Journal of Chemical Physics* **153**, 154102 (2020).
- <sup>24</sup>T. Tsuchimochi and T. Van Voorhis, “Time-dependent projected hartree-fock,” *J. Chem. Phys.* **142**, 124103 (2015).
- <sup>25</sup>T. Tsuchimochi, “Spin-flip configuration interaction singles with exact spin-projection: Theory and applications to strongly correlated systems,” *The Journal of Chemical Physics* **143**, 144114 (2015).
- <sup>26</sup>B. G. Levine, C. Ko, J. Quenneville, and T. J. Martínez, “Conical intersections and double excitations in time-dependent density functional theory,” *Molecular Physics* **104**, 1039–1051 (2006).
- <sup>27</sup>S. L. Li, A. V. Marenich, X. Xu, and D. G. Truhlar, “Configuration interaction-corrected tamm-dancoff approximation: A time-dependent density functional method with the correct dimensionality of conical intersections,” *The Journal of Physical Chemistry Letters* **5**, 322–328 (2014).
- <sup>28</sup>J. T. Taylor, D. J. Tozer, and B. F. E. Curchod, “On the description of conical intersections between excited electronic states with lr-tddft and adc(2),” *The Journal of Chemical Physics* **159**, 214115 (2023).
- <sup>29</sup>J. A. R. Shea and E. Neuscamman, “Communication: A mean field platform for excited state quantum chemistry,” *The Journal of Chemical Physics* **149**, 081101 (2018).

- <sup>30</sup>J. A. R. Shea, E. Gwin, and E. Neuscamman, "A generalized variational principle with applications to excited state mean field theory," *Journal of Chemical Theory and Computation* **16**, 1526–1540 (2020).
- <sup>31</sup>L. N. Tran and E. Neuscamman, "Improving excited-state potential energy surfaces via optimal orbital shapes," *J. Phys. Chem. A* **124**, 8273–8279 (2020).
- <sup>32</sup>T. S. Hardikar and E. Neuscamman, "A self-consistent field formulation of excited state mean field theory," *The Journal of Chemical Physics* **153**, 164108 (2020).
- <sup>33</sup>R. Clune, J. A. R. Shea, and E. Neuscamman, "N5-scaling excited-state-specific perturbation theory," *Journal of Chemical Theory and Computation* **16**, 6132–6141 (2020).
- <sup>34</sup>R. Clune and E. Neuscamman, "An excitation matched local correlation approach to excited state specific perturbation theory," *The Journal of Chemical Physics* **163**, 094109 (2025).
- <sup>35</sup>X. Liu, Q. Ou, E. Alguire, and J. E. Subotnik, "Communication: An inexpensive, variational, almost black-box, almost size-consistent correction to configuration interaction singles for valence excited states," *J. Chem. Phys.* **138**, 221105 (2013).
- <sup>36</sup>T. Tsuchimochi, "Double configuration interaction singles: Scalable and size-intensive approach for orbital relaxation in excited states and bond-dissociation," *The Journal of Chemical Physics* **161**, 241102 (2024).
- <sup>37</sup>H.-J. Werner and W. Meyer, "A quadratically convergent mcsf method for the simultaneous optimization of several states," *The Journal of Chemical Physics* **74**, 5794–5801 (1981).
- <sup>38</sup>P. G. Szalay, T. Müller, G. Gidofalvi, H. Lischka, and R. Shepard, "Multiconfiguration self-consistent field and multireference configuration interaction methods and applications," *Chemical Reviews* **112**, 108–181 (2012).
- <sup>39</sup>A. Kazaryan, J. Heuver, and M. Filatov, "Excitation energies from spin-restricted ensemble-referenced kohn-sham method: A state-average approach," *The Journal of Physical Chemistry A* **112**, 12980–12988 (2008).
- <sup>40</sup>J. Nite and C. A. Jiménez-Hoyos, "Low-cost molecular excited states from a state-averaged resonating hartree-fock approach," *Journal of Chemical Theory and Computation* **15**, 5343–5351 (2019).
- <sup>41</sup>P.-O. Löwdin, "Quantum theory of many-particle systems. iii. extension of the hartree-fock scheme to include degenerate systems and correlation effects," *Phys. Rev.* **97**, 1509 (1955).
- <sup>42</sup>G. E. Scuseria, C. A. Jiménez-Hoyos, T. M. Henderson, K. Samanta, and J. K. Ellis, "Projected quasiparticle theory for molecular electronic structure," *J. Chem. Phys.* **135**, 124108 (2011).
- <sup>43</sup>C. A. Jiménez-Hoyos, T. M. Henderson, T. Tsuchimochi, and G. E. Scuseria, "Projected hartree-fock theory," *J. Chem. Phys.* **136**, 164109 (2012).
- <sup>44</sup>H.-J. Werner and P. J. Knowles, "A second order multiconfiguration scf procedure with optimum convergence," *The Journal of Chemical Physics* **82**, 5053–5063 (1985).
- <sup>45</sup>P. J. Knowles and H.-J. Werner, "An efficient second-order mc scf method for long configuration expansions," *Chemical Physics Letters* **115**, 259–267 (1985).
- <sup>46</sup>P. Pulay, "Convergence acceleration of iterative sequences. the case of scf iteration," *Chem. Phys. Lett.* **73**, 393–398 (1980).
- <sup>47</sup>P. Pulay, "Improved scf convergence acceleration," *J. Comput. Chem.* **3**, 556–560 (1982).
- <sup>48</sup>B. Helmich-Paris, "A trust-region augmented hessian implementation for restricted and unrestricted hartree-fock and kohn-sham methods," *The Journal of Chemical Physics* **154**, 164104 (2021).
- <sup>49</sup>B. Helmich-Paris, "A trust-region augmented hessian implementation for state-specific and state-averaged casscf wave functions," *The Journal of Chemical Physics* **156**, 204104 (2022).
- <sup>50</sup>T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular Electronic-Structure Theory* (Wiley, New York, 2000).
- <sup>51</sup>T. Tsuchimochi, "Conical intersection and analytical nuclear gradients of state-averaged configuration interaction singles variants," in preparation.
- <sup>52</sup>S. Høst, J. Olsen, B. Jansik, L. Thøgersen, P. Jørgensen, and T. Helgaker, "The augmented roothaan-hall method for optimizing hartree-fock and kohn-sham density matrices," *The Journal of Chemical Physics* **129**, 124106 (2008).
- <sup>53</sup>E. R. Davidson, "The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices," *J. Comput. Phys.* **17**, 87–94 (1975).
- <sup>54</sup>R. Fletcher, "Restricted step methods," in *Practical Methods of Optimization* (John Wiley & Sons, Ltd, 2000) Chap. 5, pp. 95–109.
- <sup>55</sup>M. Uejima and S. L. Ten-no, "Quadratically convergent self-consistent field of projected hartree-fock," *The Journal of Chemical Physics* **153**, 164103 (2020).
- <sup>56</sup>C. A. Jiménez-Hoyos, R. Rodríguez-Guzmán, and G. E. Scuseria, "Multi-component symmetry-projected approach for molecular ground state correlations," *J. Chem. Phys.* **139**, 204102 (2013).
- <sup>57</sup>M. Uejima and S. Ten-no, "Analytic energy gradient of projected hartree-fock within projection after variation," *The Journal of Chemical Physics* **146**, 104106 (2017).
- <sup>58</sup>GELLAN, A Hierarchical Quantum Chemistry Program; Kobe University, Kobe.
- <sup>59</sup>E. Epifanovsky, A. T. B. Gilbert, X. Feng, J. Lee, Y. Mao, N. Mardirossian, P. Pokhilko, A. F. White, M. P. Coons, A. L. Dempwolff, Z. Gan, D. Hait, P. R. Horn, L. D. Jacobson, I. Kaliman, J. Kussmann, A. W. Lange, K. U. Lao, D. S. Levine, J. Liu, S. C. McKenzie, A. F. Morrison, K. D. Nanda, F. Plasser, D. R. Rehn, M. L. Vidal, Z.-Q. You, Y. Zhu, B. Alam, B. J. Albrecht, A. Aldossary, E. Alguire, J. H. Andersen, V. Athavale, D. Barton, K. Begam, A. Behn, N. Bellonzi, Y. A. Bernard, E. J. Berquist, H. G. A. Burton, A. Carreras, K. Carter-Fenk, R. Chakraborty, A. D. Chien, K. D. Closser, V. Cofer-Shabica, S. Dasgupta, M. de Wergifosse, J. Deng, M. Diedenhofen, H. Do, S. Ehler, P.-T. Fang, S. Fatehi, Q. Feng, T. Friedhoff, J. Gayvert, Q. Ge, G. Gidofalvi, M. Goldey, J. Gomes, C. E. González-Espinoza, S. Gulania, A. O. Gumina, M. W. D. Hanson-Heine, P. H. P. Harbach, A. Hauser, M. F. Herbst, M. Hernández Vera, M. Hodecker, Z. C. Holden, S. Houck, X. Huang, K. Hui, B. C. Huynh, M. Ivanov, Ádám Jász, H. Ji, H. Jiang, B. Kaduk, S. Kähler, K. Khistyae, J. Kim, G. Kis, P. Klunzinger, Z. Koczor-Benda, J. H. Koh, D. Kosenkov, L. Koulias, T. Kowalczyk, C. M. Krauter, K. Kue, A. Kunitsa, T. Kus, I. Ladjánszki, A. Landau, K. V. Lawler, D. Lefrancois, S. Lehtola, R. R. Li, Y.-P. Li, J. Liang, M. Liebenthal, H.-H. Lin, Y.-S. Lin, F. Liu, K.-Y. Liu, M. Loipersberger, A. Luenser, A. Manjanath, P. Manohar, E. Mansoor, S. F. Manzer, S.-P. Mao, A. V. Marenich, T. Markovich, S. Mason, S. A. Maurer, P. F. McLaughlin, M. F. S. J. Menger, J.-M. Mewes, S. A. Mewes, P. Morgante, J. W. Mullinax, K. J. Oosterbaan, G. Paran, A. C. Paul, S. K. Paul, P. Fabijan, Z. Pei, S. Prager, E. I. Proynov, A. Rák, E. Ramos-Cordoba, B. Rana, A. E. Rask, A. Rettig, R. M. Richard, F. Rob, E. Rossomme, T. Scheele, M. Scheurer, M. Schneider, N. Sergueev, S. M. Sharada, W. Skomorowski, D. W. Small, C. J. Stein, Y.-C. Su, E. J. Sundstrom, Z. Tao, J. Thirman, G. J. Tornai, T. Tsuchimochi, N. M. Tubman, S. P. Veccham, O. Vydrov, J. Wenzel, J. Witte, A. Yamada, K. Yao, S. Yeganeh, S. R. Yost, A. Zech, I. Y. Zhang, X. Zhang, Y. Zhang, D. Zuev, A. Aspuru-Guzik, A. T. Bell, N. A. Besley, K. B. Bravaya, B. R. Brooks, D. Casanova, J.-D. Chai, S. Coriani, C. J. Cramer, G. Cserey, I. DePrince, A. Eugene, J. DiStasio, Robert A., A. Dreuw, B. D. Dunietz, T. R. Furlani, I. Goddard, William A., S. Hammes-Schiffer, T. Head-Gordon, W. J. Hehre, C.-P. Hsu, T.-C. Jagau, Y. Jung, A. Klamt, J. Kong, D. S. Lambrecht, W. Liang, N. J. Mayhall, C. W. McCurdy, J. B. Neaton, C. Ochsenfeld, J. A. Parkhill, R. Peverati, V. A. Rassolov, Y. Shao, L. V. Slipchenko, T. Stauch, R. P. Steele, J. E. Subotnik, A. J. W. Thom, A. Tkatchenko, D. G. Truhlar, T. Van Voorhis, T. A. Wesolowski, K. B. Whaley, I. Woodcock, H. Lee, P. M. Zimmerman, S. Faraji, P. M. W. Gill, M. Head-Gordon, J. M. Herbert, and A. I. Krylov, "Software for the

- frontiers of quantum chemistry: An overview of developments in the q-chem 5 package," *The Journal of Chemical Physics* **155**, 084801 (2021).
- <sup>60</sup>P.-F. Loos, F. Lipparini, D. A. Matthews, A. Blondel, and D. Jacquemin, "A Mountaineering Strategy to Excited States: Revising Reference Values with EOM-CC4," *Journal of Chemical Theory and Computation* **18**, 4418–4427 (2022), publisher: American Chemical Society.
- <sup>61</sup>M. Shao, F. H. da Jornada, C. Yang, J. Deslippe, and S. G. Louie, "Structure preserving parallel algorithms for solving the bethe-salpeter eigenvalue problem," *Linear Algebra and its Applications* **488**, 148–167 (2016).