The Journal of Chemical Physics

Systematic design of active spaces for multi-reference calculations of singlet–triplet gaps of organic diradicals, with benchmarks against doubly electron-attached coupled-cluster data

Cite as: J. Chem. Phys. **147**, 164120 (2017); https://doi.org/10.1063/1.4998256
Submitted: 29 July 2017 • Accepted: 04 October 2017 • Published Online: 31 October 2017

Samuel J. Stoneburner, Jun Shen, Adeayo O. Ajala, et al.

Systematic design of active spaces for multi-reference calculations of singlet–triplet gaps of organic diradicals, with benchmarks against doubly electron-attached coupled-cluster data

Samuel J. Stoneburner,[1] Jun Shen,[2] Adeayo O. Ajala,[2] Piotr Piecuch,[2,3,a)]
Donald G. Truhlar,[1,a)] and Laura Gagliardi[1,a)]
[1]*Department of Chemistry, Chemical Theory Center, and Minnesota Supercomputing Institute, University of Minnesota, 207 Pleasant Street Southeast, Minneapolis, Minnesota 55455-0431, USA*
[2]*Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, USA*
[3]*Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824, USA*

(Received 29 July 2017; accepted 4 October 2017; published online 31 October 2017)

Singlet–triplet gaps in diradical organic π-systems are of interest in many applications. In this study, we calculate them in a series of molecules, including cyclobutadiene and its derivatives and cyclopentadienyl cation, by using correlated participating orbitals within the complete active space (CAS) and restricted active space (RAS) self-consistent field frameworks, followed by second-order perturbation theory (CASPT2 and RASPT2). These calculations are evaluated by comparison with the results of doubly electron-attached (DEA) equation-of-motion (EOM) coupled-cluster (CC) calculations with up to 4-particle–2-hole (4*p*-2*h*) excitations. We find active spaces that can accurately reproduce the DEA-EOMCC(4*p*-2*h*) data while being small enough to be applicable to larger organic diradicals. *Published by AIP Publishing.* https://doi.org/10.1063/1.4998256

I. INTRODUCTION

Organic diradicals are of interest as reaction intermediates[1] and in a variety of applications including photochemical pathways,[2] molecular magnets,[3] magnetic resonance imaging,[4] spintronics,[5,6] nonlinear optics,[7] and photovoltaics.[8–12] One of the most important characteristics of diradical molecules is the energy gap between their lowest singlet and triplet states, ΔE_{ST}. The persistence of magnetic properties at room temperature typically requires a triplet ground state with ΔE_{ST} of at least a couple of kcal/mol[13] and the magnitude of ΔE_{ST} plays a direct role in singlet fission.[8] However, determining accurate values of ΔE_{ST} for diradicals remains a challenge, even when high-level *ab initio* methods are employed.[14–19] This is because diradicals feature low-lying open-shell singlet states with nearly degenerate singly occupied molecular orbitals (SOMOs)[20–22] and challenging closed-shell singlets with multiple significantly contributing configuration state functions (CSFs),[19,23] and the treatment of these states has to be balanced with the treatment of triplet states that have a single-reference nature. One can try to use conventional single-reference methods of coupled-cluster (CC) theory,[24–29] such as the CC singles and doubles (CCSD) approach[30] or the CC method with singles, doubles, and quasiperturbative connected triples [CCSD(T)],[31] or use Kohn-Sham (KS) density functional theory (DFT) with symmetry-broken solutions,[15] but these approaches can lead to spin-contaminated results and an erratic description of the multi-determinantal singlet states.[19,32,33] In this work, we turn to multi-reference methods and a new generation of particle non-conserving single-reference CC schemes that can address deficiencies of other quantum chemistry approaches in applications involving diradicals in a computationally manageable fashion.

The most widely used multi-reference methods are based on complete active space self-consistent field (CASSCF)[34–36] reference states. In the CASSCF method, the wave function is defined by partitioning MOs into three disjoint sets, namely, the inactive, active, and external orbitals. The inactive orbitals are kept doubly occupied and the external orbitals are kept empty during the calculations. The electrons in the active orbitals are allowed to distribute in all possible ways, generating a full configuration interaction (CI) state within the active space.[35,37] In order to obtain reliable results, the active orbitals should be chosen such that the CSFs included in the CASSCF calculation dominate the electronic states of interest, capturing the correlation effects due to electronic near-degeneracies. CASSCF should provide a good treatment of static correlation, but it neglects most of the dynamic correlation effects that originate from short-range electron-electron repulsion and long-range dispersion interactions.[38] In one approach employed in the present work, the missing dynamic correlations are added with the help of multi-reference perturbation theory[36] following the complete active space second order perturbation theory (CASPT2) model.[39,40]

CASPT2 allows one to handle electronic near-degeneracies and dynamic correlations in a reasonably accurate and balanced manner if adequate active spaces can be found and used. However, in analogy to full CI, the number of CSFs in the active space scales factorially with the numbers of active orbitals and electrons.[41] As a result, the CASSCF and CASPT2 approaches with active spaces larger than 16 electrons in 16 orbitals are unaffordable with current standard

a)Authors to whom correspondence should be addressed: piecuch@chemistry.msu.edu; truhlar@umn.edu; and gagliard@umn.edu

programs.[41] Thus, it is desirable to consider less expensive alternatives to CASSCF for generating reference wave functions for the subsequent multi-reference perturbation theory, CI, and CC calculations. One such alternative is offered by the restricted active space SCF (RASSCF) approach,[42] which decomposes the active orbital space into three subspaces, abbreviated as RAS1, RAS2, and RAS3, so that the numbers of CSFs used in the CI diagonalization steps are much smaller than those characterizing CASSCF calculations. In RAS1, all orbitals are doubly occupied except for electronic excitations up to a certain excitation rank (typically, two) into RAS2 and RAS3. The active orbitals in RAS3 are unoccupied except for electronic excitations up to a certain excitation rank (once again, typically, two) from RAS1 and RAS2. The remaining active electrons are distributed among the available RAS2 orbitals in all possible ways. RASSCF allows much larger active spaces than those that can presently be used in CASSCF computations, but the calculations can still become unaffordable as the system size increases, so finding ways to minimize the numbers of active electrons and active orbitals in multi-reference work remains an important objective.[43–46] In a typical application, the choice of active space is made by chemical intuition and trial and error. This makes the results of multi-reference calculations user-dependent and the choice of adequate active space can be labor intensive. Here, we instead consider a more systematic procedure, namely, the "correlated participating orbital" (CPO) scheme proposed in Ref. 47. Originally developed for reactions and barrier heights, the CPO scheme has recently been systematically and successfully applied to singlet–triplet splittings in divalent radicals.[48,49]

The main objective of the present study is to explore the usefulness of CPOs in CASPT2 and RASPT2 calculations of the singlet–triplet gaps in a series of organic diradical π-systems that were previously explored by Saito *et al.*[33] using the restricted and unrestricted CCSD and CCSD(T) methods, the state-specific multi-reference CCSD approach of Mukherjee and co-workers,[50] abbreviated as MkCCSD, and unrestricted KS-DFT approaches employing selected exchange–correlation functionals. We systematically examine three CPO-type active spaces and their subdivisions with the goal of finding active spaces that can provide a reliable description of the systems examined in this work and that can serve as the basis for a more general recipe, which might be used in CASPT2 and RASPT2 calculations for other diradical organic π-systems in the future.

The significant disagreements among the different methods employed by Saito *et al.*,[33] as well as in various other papers (e.g., Refs. 51–53), show that the systems examined by these authors and in the present study are computationally very challenging. In particular, the various single- and multi-reference CC results for the singlet–triplet gaps reported in Ref. 33 are not consistent enough to serve as reliable reference values to benchmark our CPO-based CASPT2 and RASPT2 methods. To address this concern, we performed in this work new benchmark calculations using methods based on the doubly electron-attached (DEA) equation-of-motion (EOM) CC formalism,[54–59] which belongs to a broader category of particle non-conserving EOMCC theories (see Refs. 60 and 61 for selected reviews). The DEA-EOMCC framework allows one to determine ground and excited states of systems, such as diradicals, that are formally obtained by attaching two electrons to closed shells. In addition to the usual features of the CC/EOMCC methodology, such as fast convergence toward the exact, full CI limit, size extensivity in describing the underlying ground states, and size intensivity of the excitation (in this case, electron attachment) energies, the DEA-EOMCC calculations produce wave functions that are automatically adapted to the spin symmetry, i.e., one avoids the spin-contamination issues that arise when the conventional single-reference CC and EOMCC approaches using the unrestricted Hartree-Fock (UHF) or restricted open-shell Hartree-Fock (ROHF) references are exploited. Because of our interest in providing reliable data for benchmarking the CPO-based CASPT2 and RASPT2 approaches, we focus on the DEA-EOMCC calculations with up to four-particle–two-hole ($4p$-$2h$) components in the corresponding electron-attachment operator, which, as shown in Refs. 57–59, provide a nearly exact description of the electronic spectra of diradicals. Since the full DEA-EOMCC($4p$-$2h$) calculations for systems with larger numbers of electrons are prohibitively expensive, we use the more practical active-space DEA-EOMCC models, in which one selects the leading $4p$-$2h$ or $4p$-$2h$ and $3p$-$1h$ contributions with the help of small subsets of active orbitals.[57–59] As shown in Refs. 57–59, the DEA-EOMCC approaches with an active-space treatment of $4p$-$2h$ or $4p$-$2h$ and $3p$-$1h$ components accurately reproduce the results of the full DEA-EOMCC($4p$-$2h$) calculations at the small fraction of the computational cost, so they are well suited for generating reliable data for benchmarking the CPO-based CASPT2 and RASPT2 methods in this work. Although, as shown in this work, the effect of the basis set on the calculated singlet–triplet values is small, we consider it as well by combining the highest-level DEA-EOMCC($4p$-$2h$)-type data obtained with a smaller basis set with the results of the larger-basis set DEA-EOMCC($3p$-$1h$) calculations, in which $4p$-$2h$ terms are neglected.

In summary, the main objective of this work is to test various choices of the CPO active spaces and their RAS subdivisions, so that we can find optimum spaces that predict singlet–triplet gaps that are in good agreement with the DEA-EOMCC benchmark data, while being small enough to be applicable to larger organic diradicals. By having access to the DEA-EOMCC information at the $3p$-$1h$ and $4p$-$2h$ levels, we can also comment on the importance of $4p$-$2h$ contributions in studies of diradicals.

The paper is organized as follows. In Sec. II, we describe the molecular systems under consideration and the computational methods. In Sec. III, we present the results and discuss them. Section IV contains conclusions.

II. COMPUTATIONAL DETAILS

A. Molecular systems examined in this study

The following diradical systems, shown in Fig. 1, are considered in this work: D_{4h}-symmetric form of cyclobutadiene (1), D_{5h}-symmetric cyclopentadienyl cation

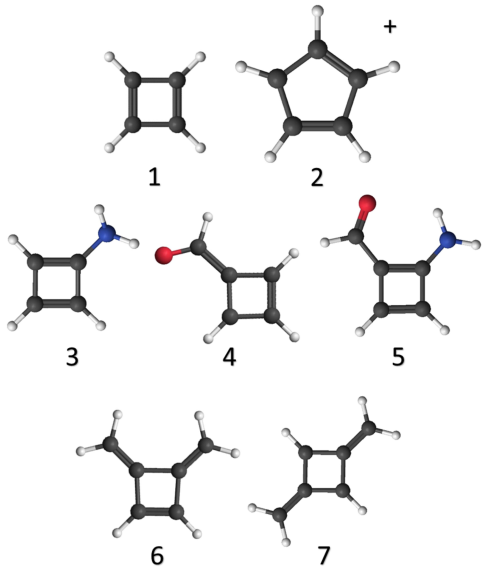

FIG. 1. Diradical systems under investigation. 1: C_4H_4, 2: $C_5H_5^+$, 3: $C_4H_3NH_2$, 4: C_4H_3CHO, 5: $C_4H_2NH_2CHO$, 6: C_4H_2-1,2-$(CH_2)_2$, 7: C_4H_2-1,3-$(CH_2)_2$.

(2), and five cyclobutadiene derivatives with polar substituents, including C_1-symmetric aminocyclobutadiene (3), C_1-symmetric formylcyclobutadiene (4), C_1-symmetric 1-amino-2-formyl-cyclobutadiene (5), C_{2v}-symmetric 1,2-bis(methylene)cyclobutadiene (6), and D_{2h}-symmetric 1,3-bis(methylene)cyclobutadiene (7). All geometries were taken from the work of Saito *et al.*[33] Notice that, even for the more symmetric systems, all calculations were performed in the C_1 point group. Each system features two degenerate (systems 1 and 2) or nearly degenerate (the remaining systems) singly occupied π orbitals centered primarily on the carbon rings. The singly occupied orbitals of system 1 are shown in Fig. 2. In each case, the lowest-energy singlet and triplet states differ by a spin-flip $\pi \rightarrow \pi^*$ transition and the corresponding energy gap is defined by

$$\Delta E_{ST} = E_{singlet} - E_{triplet}, \tag{1}$$

where a negative number indicates that the singlet is lower in energy. In the case of system 1, the singlet is $^1B_{1g}$ and the triplet is $^3A_{2g}$ (however, we run the calculations without imposing symmetry constraints).[62]

The orbitals involved in the singlet–triplet transition, whose occupancies change in the dominant CSFs, are always the frontier orbitals, which are the SOMOs except for system 5. These SOMOs are singly occupied π orbitals located primarily on the carbon ring, although the substituents are also involved for systems 6 and 7. The lowest singlet of system 5 has frontier orbitals with occupation numbers close to two or zero, so for the singlet state of system 5, the frontier orbitals are the highest occupied and lowest unoccupied MOs.

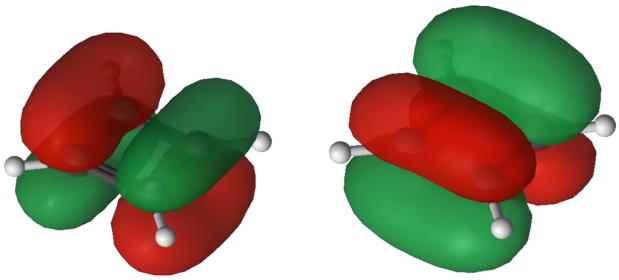

FIG. 2. Singly occupied π orbitals for system 1.

B. Benchmark DEA-EOMCC calculations

The DEA-EOMCC methods aim at the determination of ground and excited states of systems, such as diradicals, which can be obtained by attaching two electrons to the corresponding closed-shell cores. This is accomplished using the wave function ansatz $\left|\Psi_\mu^{(N)}\right\rangle = R_\mu^{(+2)}\left|\Psi_0^{(N-2)}\right\rangle$, where $\left|\Psi_\mu^{(N)}\right\rangle$ is the ground ($\mu = 0$) or excited ($\mu > 0$) state of the N-electron diradical of interest, $\left|\Psi_0^{(N-2)}\right\rangle = e^T\left|\Phi^{(N-2)}\right\rangle$ is the CC ground state of the $(N-2)$-electron closed-shell core (with T and $\left|\Phi^{(N-2)}\right\rangle$ representing the corresponding cluster operator and reference determinant), and $R_\mu^{(+2)} = R_{\mu,2p}+R_{\mu,3p-1h}+R_{\mu,4p-2h}+\cdots$ is the operator attaching two electrons to $\left|\Psi_0^{(N-2)}\right\rangle$ using the $2p$ component $R_{\mu,2p}$, while allowing the relaxation of the remaining electrons via its $3p$-$1h$ ($R_{\mu,3p-1h}$), $4p$-$2h$ ($R_{\mu,4p-2h}$), and other many-body components.

As shown in Refs. 57–59, the level of the DEA-EOMCC theory that provides a very accurate description of diradical electronic spectra, including energy gaps between the low-lying singlet and triplet states, is DEA-EOMCC($4p$-$2h$), where the electron-attaching operator $R_\mu^{(+2)}$ is truncated at the $4p$-$2h$ component $R_{\mu,4p-2h}$. Unfortunately, the most expensive steps of full DEA-EOMCC($4p$-$2h$) scale as $n_o{}^2n_u{}^6$, where n_o (n_u) is the number of orbitals occupied (unoccupied) in the underlying reference determinant $\left|\Phi^{(N-2)}\right\rangle$, limiting the DEA-EOMCC($4p$-$2h$) calculations to small systems. However, as demonstrated in Refs. 57–59, it is sufficient to use small subsets of orbitals unoccupied in $\left|\Phi^{(N-2)}\right\rangle$ to select the dominant $4p$-$2h$ terms, with virtually no loss in accuracy and at the small fraction of the cost of parent DEA-EOMCC($4p$-$2h$) computations. The resulting DEA-EOMCC($4p$-$2h$)$\{N_u\}$ approach, where $N_u << n_u$ designates the number of active unoccupied orbitals used to select the leading $4p$-$2h$ contributions, which belongs to a larger family of the active-space CC and EOMCC theories,[63] reduces the $n_o{}^2n_u{}^6$ steps of its full DEA-EOMCC($4p$-$2h$) parent to a more manageable $N_u{}^2n_o{}^2n_u{}^4$ level. One can use similar ideas to select the dominant $3p$-$1h$ contributions, either within the DEA-EOMCC($4p$-$2h$)$\{N_u\}$ scheme or within its lower-level DEA-EOMCC($3p$-$1h$) counterpart where $4p$-$2h$ terms are neglected, replacing the $n_on_u{}^5$ steps associated with $3p$-$1h$ contributions by the less expensive $N_un_on_u{}^4$ operations.[59] As shown in Ref. 59, the resulting DEA-EOMCC($3p$-$1h$,$4p$-$2h$)$\{N_u\}$ and DEA-EOMCC($3p$-$1h$)$\{N_u\}$ approaches accurately reproduce the corresponding DEA-EOMCC($4p$-$2h$) or DEA-EOMCC($4p$-$2h$)$\{N_u\}$ and DEA-EOMCC($3p$-$1h$) data at the small fraction of the computational costs.

The highest level of the DEA-EOMCC theory used in this work is DEA-EOMCC($4p$-$2h$)$\{N_u\}$. In carrying out the DEA-EOMCC($4p$-$2h$)$\{N_u\}$ calculations, we followed Saito *et al.*[33] and used the cc-pVDZ basis set.[64] In order to examine the dependence of our results on the basis set, we also

used the larger maug-cc-pVTZ basis.[65] For the larger systems considered in this study, namely, cyclobutadiene derivatives, the DEA-EOMCC(4*p*-2*h*){N_u}/maug-cc-pVTZ calculations using our present codes turned out to be quite expensive, so to estimate the DEA-EOMCC(4*p*-2*h*){N_u}/maug-cc-pVTZ results we adopted a simple extrapolation scheme, abbreviated as DEA-EOMCC[4*p*-2*h*], where we calculate the final energies as follows:

$$E[4p\text{-}2h] = E(4p\text{-}2h)\{N_u\}/\mathrm{DZ} + E(3p\text{-}1h)\{N_u\}/\mathrm{mTZ} - E(3p\text{-}1h)\{N_u\}/\mathrm{DZ}. \quad (2)$$

The first term on the right-hand side of Eq. (2) is the DEA-EOMCC(4*p*-2*h*){N_u}/cc-pVDZ energy. The effect of going from the cc-pVDZ basis set (abbreviated as DZ) to maug-cc-pVTZ (abbreviated as mTZ) is estimated by forming the difference of energies obtained in the DEA-EOMCC(3*p*-1*h*){N_u}/maug-cc-pVTZ and DEA-EOMCC(3*p*-1*h*){N_u}/cc-pVDZ calculations.

In addition to the calculations entering Eq. (2), we performed the full DEA-EOMCC(3*p*-1*h*) and active-space DEA-EOMCC(3*p*-1*h*,4*p*-2*h*){N_u} computations using the cc-pVDZ basis set (all seven systems) and the DEA-EOMCC(3*p*-1*h*,4*p*-2*h*){N_u}/maug-cc-pVTZ calculations for the smallest system 1. We carried out these extra computations to validate Eq. (2), especially the usefulness of the DEA-EOMCC(3*p*-1*h*){N_u} approach in estimating the effect of going from the cc-pVDZ basis set to maug-cc-pVTZ (see Sec. III A for a discussion). Following Refs. 57–59, in all of the DEA-EOMCC calculations performed in this work, the ground states of the underlying $(N-2)$-electron closed-shell cores were obtained using CCSD.

All of the DEA-EOMCC calculations reported in this work and the underlying CCSD computations were performed using the restricted Hartree-Fock (RHF) MOs corresponding to the $(N-2)$-electron closed-shell cores. In this way, we could maintain all of the relevant symmetries throughout the calculations. We tested the usage of other orbitals, such as the N-electron ROHF MOs obtained for the triplet states of diradicals examined in this work, but, in agreement with Refs. 57–59, the resulting singlet–triplet gaps, especially those obtained with the highest DEA-EOMCC(4*p*-2*h*)-type levels, turned out to be virtually independent of the type of MOs used in the calculations. As in Ref. 33, in all of the post-HF calculations, the core orbitals correlating with the 1*s* shells of the C, N, and O atoms were kept frozen and the spherical components of *d* and *f* basis functions were employed throughout.

In carrying out the various DEA-EOMCC computations, we followed the strategy employed in Ref. 33. Thus, we used the D_{4h} point group for system 1, the D_{5h} group for system 2, and C_1 for the remaining systems 3–7. In each case, the closed-shell $(N-2)$-electron reference system used to set up the DEA-EOMCC calculations was obtained by vacating the two valence partly occupied orbitals that define the singlet and triplet states of interest, which are exactly degenerate in systems 1 and 2 and nearly degenerate in systems 3–7. For example, the $(N-2)$-electron reference dication used in the DEA-EOMCC calculations for system 1 was obtained by vacating the two valence SOMOs of e_g symmetry. For system 2, we vacated the degenerate valence e_1'' shell, etc. Consistent with the structure of the valence π shells in systems 1–7, which consist of one doubly occupied, two partly occupied, and one unoccupied MOs in systems 1, 3, and 4 and one doubly occupied, two partly occupied, and two unoccupied MOs in systems 2 and 5–7, the active spaces needed to perform the DEA-EOMCC(3*p*-1*h*){N_u}, DEA-EOMCC(3*p*-1*h*,4*p*-2*h*){N_u}, and DEA-EOMCC(4*p*-2*h*){N_u} calculations were defined in the following manner. For systems 1, 3, and 4, we used the $N_u = 3$ MOs, which are the three lowest-energy unoccupied orbitals in the respective 1^{2+}, 3^{2+}, and 4^{2+} reference dications. For systems 2, 5, 6, and 7, we used the $N_u = 4$ orbitals, which are the four lowest-energy unoccupied MOs in the respective $(N-2)$-electron 2^{2+}, 5^{2+}, 6^{2+}, and 7^{2+} species. We verified the appropriateness of the above active orbital choices by comparing the full DEA-EOMCC(3*p*-1*h*)/cc-pVDZ and active-space DEA-EOMCC(3*p*-1*h*){N_u}/cc-pVDZ data (see Sec. III A for further discussion).

All of the DEA-EOMCC calculations reported in this work were performed using the codes developed in Refs. 57–59, interfaced with GAMESS[66] and taking advantage of the spin-free CCSD GAMESS routines[67] and the routines used in some of our earlier EOMCC studies.[68–70]

C. CASPT2 and RASPT2 calculations

CASPT2[39] and RASPT2[71] calculations, including the underlying reference state calculations by CASSCF[34] and RASSCF,[42] were performed using the maug-cc-pVTZ[72] and ANO-RCC-VTZP[73] basis sets with the Cholesky decomposition[74] using a developer version of *Molcas* 8.1.[41,75,76] Orbitals were visualized using *Luscus* 0.8.3.[77] All calculations were performed without symmetry restrictions, i.e., in C_1 symmetry.

For CASSCF calculations, the active space notation is (n,N), where n is the number of active electrons, and N is the number of orbitals in the active space. For RASSCF calculations, the active space notation is $(n,h,p;N_1,N_2,N_3)$, where n is the total number of active electrons, h is the maximum number of holes in RAS1, p is the maximum number of particles in RAS3, and N_i is the number of orbitals in RASi.

CASPT2 and RASPT2 calculations were performed with an imaginary shift of 0.1 hartree to alleviate intruder state problems. The default ionization potential-electron affinity (IPEA) shift of 0.25 hartree[78] was used to compensate for the systematic overestimation of correlation energy in CASPT2. (In the supplementary material, we give some comparison results obtained without an IPEA shift.)

1. CPO definitions

The CPO scheme is based on the idea that the active space should consist of "participating" orbitals, i.e., the orbitals most strongly involved in the process of interest, plus one correlating orbital for each participating orbital.[47] Participating orbitals are identified based on the orbital occupations from the dominant configurations, not the occupation numbers from the zeroth-order wave function. For all systems studied other than system 5, considering the occupation numbers from the zeroth-order wave function would erroneously suggest that there is no difference between the singlet and the triplet, as the frontier orbitals are singly occupied in both cases. However, although the triplet has a single dominant CSF, the singlet has two

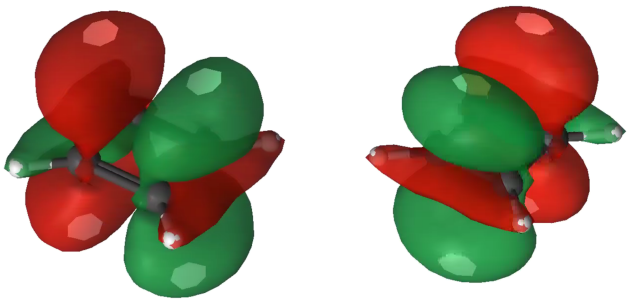

FIG. 3. Correlating π′ orbitals of the nCPO scheme for system 1.

dominant configurations: one in which one of the frontier orbitals is doubly occupied and one in which the other is doubly occupied. More specific information regarding the wave functions and dominant configurations is included in Sec. III B and in the supplementary material.

The original CPO scheme had three choices: nominal, moderate, and extended, abbreviated as nom-, mod-, and ext-CPO.[47] In the present article, we introduce a fourth option for π-systems that lies between nominal and moderate, referred to as "π-CPO." These four choices will be referred to for the rest of the paper as nCPO, πCPO, mCPO, and eCPO. In nCPO, active orbitals are the frontier orbitals and their correlating orbitals. See Figs. 2 and 3 for examples of frontier orbitals and their correlating orbitals, respectively, for system 1. In the other CPO options, as discussed next, we add additional orbitals on the atoms on which the frontier orbitals reside; these atoms are called "participating atoms." The substituent carbons of systems 6 and 7 are participating atoms, but the substituents of systems 3, 4, and 5 are not.

In the πCPO scheme, all valence π orbitals of participating atoms are active, and correlating orbitals are added as needed to ensure that each singly or doubly occupied orbital is paired with an unoccupied orbital. See Fig. 4 for examples of participating π orbitals in system 1.

In the mCPO scheme, all valence *p* orbitals of participating atoms are active, and correlating orbitals are added as needed to ensure that each singly or doubly occupied orbital is paired with an unoccupied orbital. Where there is significant *s*-*p* mixing, the *s* orbitals are taken to be those of lowest energy (one for each participating atom), and the rest are treated as *p* orbitals. For example, in system 1, there are four participating atoms, and the four lowest valence orbitals are considered to be the *s* orbitals. See Fig. 5 for examples of included *p* orbitals for system 1.

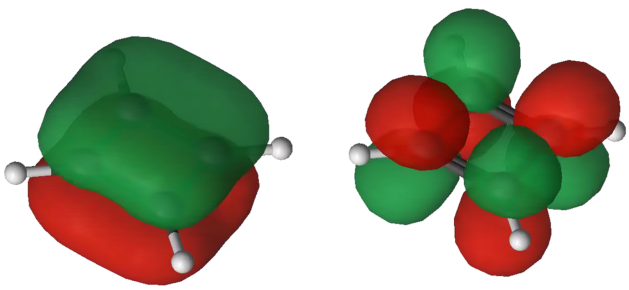

FIG. 4. Included π orbitals in the πCPO scheme for system 1.

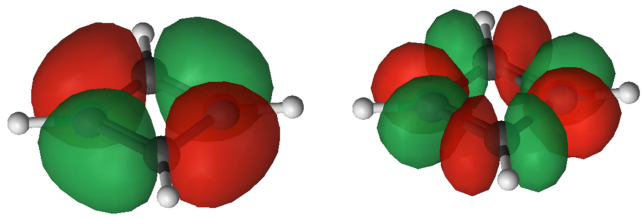

FIG. 5. Included *p* orbitals of the mCPO scheme for system 1.

In the eCPO scheme, all valence *p* and *s* orbitals of participating atoms are active, and correlating orbitals are added as needed to ensure that each singly or doubly occupied orbital is paired with an unoccupied orbital. See Fig. 6 for examples of included *s* orbitals for system 1.

2. RAS subdivisions

As discussed in the Introduction, active spaces larger than 16 electrons in 16 orbitals are unaffordable in current standard implementations of CASSCF and CASPT2. However, the active spaces chosen using eCPO would be larger than the (16,16) limit, even for systems as small as those studied here. For system 1, the smallest system under consideration, the active space chosen with eCPO would be (20,22), with 6×10^{10} CSFs for the singlet state. In order to employ such large active spaces, RASPT2 was employed. We also employed RASPT2 for some systems where CASPT2 is affordable because the goal is to test the accuracy of RASPT2 against the benchmarks that are available on small systems so we know whether it is expected to be accurate for large systems where CASPT2 is not affordable.

In RASSCF calculations in this work, two excitations were permitted from RAS1 and two excitations were permitted into RAS3. Just as CPO provides a scheme for choosing which orbitals are included in the active space, it is also helpful to have a systematic way of choosing how to divide the active space into the three RAS subspaces. For the systems studied here, we define two schemes: "valence-π" and "limited-π." In the "valence-π" scheme, RAS2 includes any valence π orbitals that are included in the given CPO level; any π orbitals that are higher in energy than the valence orbitals are placed in RAS3, along with any other unoccupied orbitals. Any doubly occupied orbitals that are not part of the π-system are placed in RAS1.

For eCPO with systems 6 and 7, the valence-π RAS division scheme results in active spaces of (30,2,2;12,6,14). The RASPT2 portion would have taken more than 22 days of computer time, so valence-π was not performed for systems 6 and 7. Instead, only the "limited-π" scheme was employed.

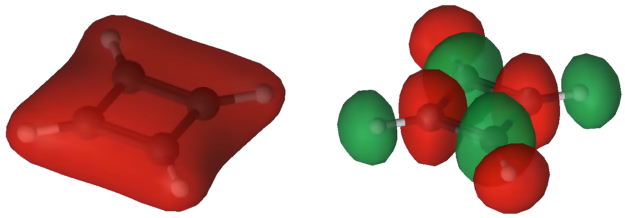

FIG. 6. Included *s* orbitals of the eCPO scheme for system 1.

Limited-π is similar to valence-π, but instead of all valence π orbitals being in RAS2, occupied π orbitals below the highest two occupied π orbitals are in RAS1 and their correlating orbitals are in RAS3. Systems 1 through 5 have only two occupied π orbitals, so the limited-π active spaces are identical to the valence-π active spaces for those systems.

For systems 6 and 7, convergence could not be achieved with limited-π RASSCF using an active space defined by πCPO, so instead restricted active space CI calculations were performed using orbitals from valence-π RASSCF with an active space defined by πCPO. Additionally, the CASSCF active space defined by mCPO for system 2 was prohibitively expensive to perform, and instead complete active space CI calculations were performed using orbitals from valence-π RASSCF with an active space defined by mCPO.

III. RESULTS AND DISCUSSION

A. DEA-EOMCC benchmark calculations

The results of our various DEA-EOMCC calculations for the singlet−triplet gaps in systems 1−7 are summarized in Table I. Our highest-level calculated DEA-EOMCC(4*p*-2*h*)$\{N_u\}$/cc-pVDZ data and their extrapolation to the larger maug-cc-pVTZ basis set using Eq. (2), abbreviated as DEA-EOMCC[4*p*-2*h*], which we treat in this work as best estimates of the ΔE_{ST} values of interest, indicate that systems 1 and 3−6 have singlet ground states, whereas the ground states of systems 2 and 7 are triplets. As shown in Table II, where we compare our extrapolated DEA-EOMCC[4*p*-2*h*] values with the singlet−triplet gaps resulting from the symmetry-broken, UHF-based, calculations using the single-reference CCSD(T) [UCCSD(T)] approach and its Brueckner-orbital analog, abbreviated as UBD(T),[79] and the multi-reference MkCCSD computations using the ROHF and CASSCF orbitals, our DEA-EOMCC(4*p*-2*h*)-level results agree in this regard with the findings of Saito *et al.*[33]

Before making further comparisons between the results of our DEA-EOMCC calculations and the ΔE_{ST} values reported in Ref. 33, we comment on the extrapolation procedure defined by Eq. (2), which is used in this work to provide reference data for benchmarking the CPO-based CASPT2 and RASPT2 schemes. We begin with the choice of active orbitals used to select the dominant 3*p*-1*h* and 4*p*-2*h* contributions in the DEA-EOMCC(3*p*-1*h*)$\{N_u\}$ and DEA-EOMCC(4*p*-2*h*)$\{N_u\}$ computations. A comparison of the results of the full DEA-EOMCC(3*p*-1*h*) and active-space DEA-EOMCC(3*p*-1*h*)$\{N_u\}$ calculations using the cc-pVDZ basis set demonstrates that our choice of active orbitals allowing us to select the dominant higher-than-2*p* contributions in the DEA-EOMCC wave function ansatz is appropriate. Indeed, as shown in Table I, the differences between the singlet−triplet gaps resulting from the DEA-EOMCC(3*p*-1*h*)/cc-pVDZ and DEA-EOMCC(3*p*-1*h*)$\{N_u\}$/cc-pVDZ calculations are very small, ranging from 0.05 kcal/mol for system 1 to 0.56 kcal/mol for system 5, where the DEA-EOMCC(3*p*-1*h*)/cc-pVDZ gap value is −7.46 kcal/mol. In fact, one observes similarly small differences when comparing the results of the higher-level DEA-EOMCC(4*p*-2*h*)$\{N_u\}$ calculations, in which 4*p*-2*h* terms are treated using active orbitals, but 3*p*-1*h* terms are treated fully, with the results obtained with the DEA-EOMCC(3*p*-1*h*,4*p*-2*h*)$\{N_u\}$ approach, in which both types of terms are treated using active orbitals. These observations are consistent with the well-known characteristic of the active-space CC and EOMCC methods, including the active-space DEA-EOMCC approaches employed in this study, which is their ability to reproduce the results of the parent CC/EOMCC calculations with small numbers of active-orbitals used in selecting higher-order excitations.[57–59,63] We can certainly conclude that the use of the active-space DEA-EOMCC(3*p*-1*h*)$\{N_u\}$ approach in Eq. (2), as a substitute for the considerably more expensive full DEA-EOMCC(3*p*-1*h*) parent in estimating the effect of going from the cc-pVDZ basis set to the maug-cc-pVTZ basis, is an appropriate procedure.

Equation (2) is also justified by the fact that the effect of going from the smaller cc-pVDZ basis to the larger maug-cc-pVTZ basis set on the calculated ΔE_{ST} values is generally rather small, implying that it is safe to estimate it using the lower-level DEA-EOMCC(3*p*-1*h*)$\{N_u\}$ method, as opposed to the significantly more expensive DEA-EOMCC(4*p*-2*h*)$\{N_u\}$ approach. Indeed, as shown in Table I, the differences between the singlet−triplet gaps resulting from the DEA-EOMCC(3*p*-1*h*)$\{N_u\}$/cc-pVDZ and DEA-EOMCC(3*p*-1*h*)$\{N_u\}$/maug-cc-pVTZ calculations range from 0.03 kcal/mol for system 2 to

TABLE I. The various DEA-EOMCC results for the singlet–triplet gaps ΔE_{ST} (in kcal/mol) in systems 1–7.

Method	1	2	3	4	5	6	7
cc-pVDZ							
(3*p*-1*h*)$\{N_u\}$	−1.37	16.38	0.44	−1.24	−6.90	−81.63	18.26
(3*p*-1*h*)	−1.42	16.06	0.34	−1.32	−7.46	−81.84	17.95
(3*p*-1*h*,4*p*-2*h*)$\{N_u\}$	−4.98	14.25	−3.22	−4.32	−4.82	−78.42	20.03
(4*p*-2*h*)$\{N_u\}$	−5.04	13.91	−3.30	−4.40	−5.49	−78.75	19.76
maug-cc-pVTZ							
(3*p*-1*h*)$\{N_u\}$	−0.53	16.35	1.09	−0.48	−7.09	−80.56	16.98
[4*p*-2*h*][a]	−4.20[b]	13.88	−2.65	−3.65	−5.68	−77.68	18.49
N_u	3	4	3	3	4	4	4

[a]Best estimate defined by the extrapolation formula given by Eq. (2).
[b]The DEA-EOMCC(3*p*-1*h*,4*p*-2*h*){3}/maug-cc-pVTZ calculation gives −4.08 kcal/mol.

TABLE II. A comparison of the ΔE_{ST} values (in kcal/mol) characterizing systems 1–7 obtained with the DEA-EOMCC[4*p*-2*h*] extrapolation defined by Eq. (2) and in the DEA-EOMCC(4*p*-2*h*)$\{N_u\}$/cc-pVDZ calculations with the UCCSD(T), UBD(T), ROHF-MkCCSD, and CASSCF-MkCCSD results reported by Saito *et al.*[33]

	DEA-EOMCC	Saito *et al.*[33]			
Molecule	[4*p*-2*h*]/(4p-2h)$\{N_u\}$	UCCSD(T)	UBD(T)	ROHF-MkCCSD	CASSCF-MkCCSD
1	−4.2/−5.0	−4.8	−5.1	−8.6	−8.1
2	13.9/13.9	14.8	14.0	13.5	9.4
3	−2.7/−3.3	−3.2	−3.6	−6.5	−7.3
4	−3.6/−4.4	−4.5	−4.5	−7.1	−6.9
5	−5.7/−5.5	−0.6	−0.9	−2.7	−4.5
6	−77.7/−78.8	−82.7	−79.8	−82.7	−84.2
7	18.5/19.8	15.0	17.1	20.0	19.5
MUE[a]	0.0/0.7	2.4	1.6	3.1	3.6

[a]Mean unsigned errors relative to the extrapolated DEA-EOMCC[4*p*-2*h*] results using Eq. (2).

1.28 kcal/mol for system 7, where the DEA-EOMCC(3*p*-1*h*)/cc-pVDZ gap value is 18.26 kcal/mol, for an average of 0.69 kcal/mol. Furthermore, although we were unable to perform the DEA-EOMCC/maug-cc-pVTZ calculations with 3*p*-1*h* and 4*p*-2*h* terms in the electron-attaching $R_\mu^{(+2)}$ operator using our existing codes for all of the systems examined in this work, we managed to obtain the DEA-EOMCC(3*p*-1*h*,4*p*-2*h*)$\{3\}$/maug-cc-pVTZ value for the singlet−triplet gap in system 1, obtaining −4.08 kcal/mol (see Table I). Our extrapolation of the DEA-EOMCC(4*p*-2*h*)/maug-cc-pVTZ-level result based on Eq. (2) gives −4.20 kcal/mol, in virtually perfect agreement with the DEA-EOMCC(3*p*-1*h*,4*p*-2*h*)$\{3\}$/maug-cc-pVTZ calculation. This means that Eq. (2) works well, allowing us to capture the effect of high-order 4*p*-2*h* correlations and the effect of going from the cc-pVDZ basis set to maug-cc-pVTZ in an accurate and computationally manageable manner.

Having established the validity of Eq. (2), which, given the above analysis and previous extensive studies of the DEA-EOMCC approaches with up to 4*p*-2*h* excitations,[57–59] is expected to produce singlet−triplet gap values in systems 1–7 to within 1 kcal/mol or better, we comment on our best DEA-EOMCC[4*p*-2*h*] [and the corresponding DEA-EOMCC(4*p*-2*h*)$\{N_u\}$/cc-pVDZ] ΔE_{ST} values. First, it is important to note that although bulk of the correlation effects is captured at the DEA-EOMCC(3*p*-1*h*) level, the high-order 4*p*-2*h* effects can be quite substantial. When we compare the extrapolated DEA-EOMCC[4*p*-2*h*] and calculated DEA-EOMCC(3*p*-1*h*)$\{N_u\}$/maug-cc-pVTZ gap values or, equivalently, the DEA-EOMCC(4*p*-2*h*)$\{N_u\}$/cc-pVDZ and DEA-EOMCC(3*p*-1*h*)$\{N_u\}$/cc-pVDZ data, the 4*p*-2*h* effects range, in absolute value, from 1.4 kcal/mol in system 5 to 3.7 kcal/mol in systems 1 and 3. Although they typically reduce the total electronic energies of the individual states, their net effect on the calculated singlet−triplet gaps can go either way. Indeed, we may encounter lowering of the signed ΔE_{ST} values due to 4*p*-2*h* correlations, as in systems 1–4, or we can find cases where the signed singlet−triplet gaps defined by Eq. (1) increase, as in systems 5–7. In some cases, the 4*p*-2*h* effects can change a singlet−triplet gap near zero to a considerably larger absolute value, as in systems 1 and 4, but there also are situations, such as system 3, where 4*p*-2*h* correlations change state ordering and the sign of ΔE_{ST}. It is quite clear from the results shown in Table I that one has to account for the high-order 4*p*-2*h* effects within the DEA-EOMCC framework to obtain reasonably converged values of the singlet−triplet gaps in diradicals. This is consistent with our earlier DEA-EOMCC studies reported in Refs. 57–59.

High accuracy of our extrapolated DEA-EOMCC[4p-2h] data and the underlying DEA-EOMCC(4*p*-2*h*)$\{N_u\}$/cc-pVDZ calculations, which include sophisticated 4*p*-2*h* terms, in addition to their lower-rank 2*p* and 3*p*-1*h* counterparts, on top of CCSD, implies that we should be able to judge other methods. Before discussing our assessment of the various CPO-based CASPT2 and RASPT2 calculations in Sec. III B, we comment on the UCCSD(T), UBD(T), ROHF-MkCCSD, and CASSCF-MkCCSD computations reported by Saito *et al.*[33] As already pointed out, all of these methods agree in predicting correct state ordering. Unfortunately, as shown in Table II, they disagree, sometimes rather significantly, in quantitative predictions. In the case of systems 1–4, there is a great deal of consistency among the singlet−triplet gap values provided by UCCSD(T) and UBD(T) and those obtained in our DEA-EOMCC(4*p*-2*h*)$\{N_u\}$/cc-pVDZ and DEA-EOMCC[4*p*-2*h*] calculations, which agree to within ∼1 kcal/mol, but one cannot say the same about the MkCCSD data, which seem to have rather large errors, on the order of 3–4 kcal/mol, displaying a significant dependence of the resulting ΔE_{ST} values on the type of orbitals used in the calculations in the case of system 2. The poor performance of MkCCSD for system 1 is reinforced by the results of the multi-reference averaged quadratic CC calculations,[80] reported in Ref. 33 as well, which give ΔE_{ST} of −5.5 kcal/mol, in good agreement with our highest-level DEA-EOMCC[4p-2h] and DEA-EOMCC(4*p*-2*h*)$\{N_u\}$/cc-pVDZ calculations and the UCCSD(T) and UBD(T) data, but in sharp disagreement with the ROHF- and CASSCF-based MkCCSD values. Based on the results for systems 1–4 and the mean-unsigned error (MUE) values relative to DEA-EOMCC[4*p*-2*h*] reported in Table II, one might recommend the use of the symmetry-broken UCCSD(T) and UBD(T) methods in the calculations of singlet−triplet gaps in diradicals, but the results for system 5, where errors relative to DEA-EOMCC(4*p*-2*h*)$\{N_u\}$/cc-pVDZ

and DEA-EOMCC[4*p*-2*h*] in the UCCSD(T) and UBD(T) ΔE_{ST} values are on the order of 5 kcal/mol, show that this would be misleading. The agreement among the UCCSD(T), UBD(T), ROHF-MkCCSD, and CASSCF-MkCCSD ΔE_{ST} values improves, when systems 6 and 7 are examined, but one still observes substantial differences among the results obtained with these four methods, on the order of 4-5 kcal/mol, which do not allow us to use them to benchmark our CPO-based CASPT2 and RASPT2 approaches. Our extrapolated DEA-EOMCC[4p-2h] data and the underlying DEA-EOMCC(4*p*-2*h*){N_u}/cc-pVDZ calculations are considerably more reliable in this regard.

B. CASPT2 and RASPT2 calculations

CASPT2 and RASPT2 results using the maug-cc-pVTZ basis set are presented in Table III. The various active spaces and their sizes are presented in Table IV. Results with the ANO-RCC-VTZP basis set are similar and are presented in the supplementary material.

The MUEs relative to the benchmark DEA-EOMCC[4*p*-2*h*] data shown in Table III are significantly larger for nCPO than for the other CPO choices. The MUEs of πCPO, mCPO, and eCPO are all under 1.0 kcal/mol, while the MUEs of nCPO are between 6 and 8 kcal/mol. Moreover, the maximum error of nCPO is above 16 kcal/mol. These poor results indicate that the nCPO calculations do not properly reflect the multi-reference character of the singlets. As explained in Sec. II C 1, the triplet states are dominated by a single configuration, while the singlet states have two dominant configurations; the weights of these two configurations vary depending on system and active space, but they are roughly equal except for all systems other than system 5. With nCPO, however, one configuration frequently outweighs the other, resulting in an inaccurate description of the wave function.

Although nCPO performs poorly, it most closely corresponds to the (2,2) active spaces used in the multi-reference CC calculations of Saito *et al.*[33] For both the singlet and triplet states, the π orbital directly below the nominal participating orbitals has an occupation number between 1.90 and 1.94, and

TABLE III. CASPT2/RASPT2 results. All values are ΔE_{ST} (kcal/mol), where a negative number indicates that the singlet is lower.

System	Active space	CASPT2	RASPT2 valence-π	RASPT2 limited-π	DEA-EOMCC [4*p*-2*h*]
1: C_4H_4	eCPO	[a]	−3.8	−3.8	−4.2
	mCPO	−4.3	−4.0	−4.0	
	πCPO	−4.4	−4.4	−4.4	
	nCPO	−12.0	−12.0	−12.0	
2: $C_5H_5^+$	eCPO	[a]	14.5	14.5	13.9
	mCPO	13.5[b]	13.7	13.7	
	πCPO	14.9	15.0	15.0	
	nCPO	21.9	20.5	20.5	
3: $C_4H_3NH_2$	eCPO	[a]	−2.2	−2.2	−2.7
	mCPO	−2.7	−2.8	−2.8	
	πCPO	−2.5	−2.5	−2.5	
	nCPO	8.0	13.5	13.5	
4: C_4H_3CHO	eCPO	[a]	−3.5	−3.5	−3.6
	mCPO	−4.0	−4.0	−4.0	
	πCPO	−3.9	−3.6	−3.6	
	nCPO	8.6	7.7	7.7	
5: $C_4H_2NH_2CHO$	eCPO	[a]	−7.2	−7.2	−5.7
	mCPO	−6.3	−6.3	−6.3	
	πCPO	−6.7	−7.4	−7.4	
	nCPO	−3.1	3.6	3.6	
6: C_4H_2-1,2-$(CH_2)_2$	eCPO	[a]	[a]	−75.9	−77.7
	mCPO	[a]	−75.1	−77.0	
	πCPO	−75.5	−75.3	−75.4[b]	
	nCPO	−81.9	−81.9	−81.9	
7: C_4H_2-1,3-$(CH_2)_2$	eCPO	[a]	[a]	18.7	18.5
	mCPO	[a]	18.8	18.4	
	πCPO	18.6	18.4	19.0[b]	
	nCPO	18.3	18.3	18.3	
MUE[c]	eCPO	...	0.6	0.7	0.0
	mCPO	0.3	0.6	0.3	
	πCPO	0.7	0.8	0.9	
	nCPO	6.5	7.9	7.9	

[a]Not available.
[b]CI only rather than CASSCF or RASSCF.
[c]MUEs exclude absent data.

TABLE IV. CASSCF/RASSCF active space sizes and numbers of CSFs, including cases that were too large to attempt. Exact numbers, including RASSCF limited-π, are presented in the supplementary material.

System	Active space	CASSCF			RASSCF valence-π		
		size	CSFs sing.	CSFs trip.	Size	CSFs sing.	CSFs trip.
1: C_4H_4	eCPO	(20,22)[a]	6×10^{10}[a]	1×10^{11}[a]	(20,2,2; 8,4,10)	2×10^5	3×10^5
	mCPO	(12,14)	2×10^7	4×10^7	(12,2,2; 4,4,6)	2×10^4	3×10^4
	πCPO	(4,6)	105	105	(4,0,2; 0,4,2)	96	97
	nCPO	(2,4)	10	6	(2,0,2; 0,2,2)	10	6
2: $C_5H_5^+$	eCPO	(24,26)[a]	1×10^{13}[a]	3×10^{13}[a]	(24,2,2; 10,5,11)	9×10^5	2×10^6
	mCPO	(14,16)[b]	3×10^8[b]	5×10^8[b]	(14,2,2; 5,5,6)	8×10^4	1×10^5
	πCPO	(4,6)	105	105	(4,0,2; 0,5,1)	105	105
	nCPO	(2,4)	10	6	(2,0,2; 0,2,2)	10	6
3: $C_4H_3NH_2$	eCPO	(20,22)[a]	6×10^{10}[a]	1×10^{11}[a]	(20,2,2; 8,4,10)	2×10^5	3×10^5
	mCPO	(12,14)	2×10^7	4×10^7	(12,2,2; 4,4,6)	2×10^4	3×10^4
	πCPO	(4,6)	105	105	(4,0,2; 0,4,2)	96	97
	nCPO	(2,4)	10	6	(2,0,2; 0,2,2)	10	6
4: C_4H_3CHO	eCPO	(20,22)[a]	6×10^{10}[a]	1×10^{11}[a]	(20,2,2; 8,4,10)	2×10^5	3×10^5
	mCPO	(12,14)	2×10^7	4×10^7	(12,2,2; 4,4,6)	2×10^4	3×10^4
	πCPO	(4,6)	105	105	(4,0,2; 0,4,2)	96	97
	nCPO	(2,4)	10	6	(2,0,2; 0,2,2)	10	6
5: $C_4H_2NH_2CHO$	eCPO	(20,22)[a]	6×10^{10}[a]	1×10^{11}[a]	(20,2,2; 8,4,10)	2×10^5	3×10^5
	mCPO	(12,14)	2×10^7	4×10^7	(12,2,2; 4,4,6)	2×10^4	3×10^4
	πCPO	(4,6)	105	105	(4,0,2; 0,4,2)	96	97
	nCPO	(2,4)	10	6	(2,0,2; 0,2,2)	10	6
6: C_4H_2-1,2- $(CH_2)_2$	eCPO	(30,32)[a]	4×10^{16}[a]	9×10^{16}[a]	(30,2,2; 12,6,14)[c]	7×10^6[c]	7×10^6[c]
	mCPO	(18,20)[a]	5×10^9[a]	1×10^{10}[a]	(18,2,2; 6,6,8)	2×10^5	1×10^6
	πCPO	(6,8)	1176	1512	(6,0,2; 0,6,2)	1015	1317
	nCPO	(2,4)	10	6	(2,0,2; 0,2,2)	10	6
7: C_4H_2-1,3- $(CH_2)_2$	eCPO	(30,32)[a]	4×10^{16}[a]	9×10^{16}[a]	(30,2,2; 12,6,14)[c]	7×10^6[c]	7×10^6[c]
	mCPO	(18,20)[a]	5×10^9[a]	1×10^{10}[a]	(18,2,2; 6,6,8)	2×10^5	1×10^6
	πCPO	(6,8)	1176	1512	(6,0,2; 0,6,2)	1015	1317
	nCPO	(2,4)	10	6	(2,0,2; 0,2,2)	10	6

[a]Not possible due to size. Number of CSFs calculated with Weyl's formula.[44]
[b]CI only rather than CASSCF or RASSCF.
[c]Not attempted due to excessive time required.

the π orbital directly above the nominal participating orbitals has an occupation number between 0.06 and 0.10. See Fig. 4 for examples of these orbitals for system 1. Active spaces chosen with nCPO and the (2,2) active space used in the work of Saito *et al.*[33] force these orbitals to have occupation numbers of 2.00 and 0.00, respectively. In contrast, the DEA-EOMCC calculations used for our reference values permit holes in the occupied π orbital and excitations into the unoccupied π orbital (among other possible holes and excitations that are more case-dependent), as do all CASPT2 and RASPT2 calculations presented here other than those using nCPO active spaces.

For πCPO, mCPO, and eCPO, there is little difference among the MUEs, as all MUEs are under 1.0 kcal/mol. The maximum errors vary depending on how the active space is divided, but in all cases they are 2.6 kcal/mol or below. We conclude that little is to be gained by using the eCPO active space, which roughly corresponds to a full-valence active space for participating atoms and is unaffordable for CASPT2, even with the small systems studied here. The mCPO active space is also unaffordable for systems 6 and 7, which are still very small systems, as the CASSCF active space would be (18,20) and have 5×10^9 CSFs for the singlet. Overall, πCPO offers the best balance between affordability and accuracy, as it provides a comparable level of accuracy with mCPO and eCPO, especially since Table IV shows that it requires active spaces many orders of magnitude smaller than the order of 10^7 to 10^{11} CSFs required by mCPO and eCPO. Good accuracy is achieved with πCPO because only π orbitals have occupation numbers less than 1.97 or more than 0.03, regardless of whether additional orbitals are included in the active space. Therefore, πCPO allows for a sufficient description of the multi-reference character of these systems, and mCPO and eCPO add considerable expense for no significant benefit.

Valence-π RASPT2 has very similar MUEs to the corresponding CASPT2 calculations. All of the orbitals in RAS1 or RAS3 in valence-π RASSCF have occupation numbers between 2.00 and 1.97 or between 0.03 and 0.00, just as for CASSCF. Encouragingly, limited-π RASPT2 also enjoys

similar accuracy to CASPT2, even though some orbitals placed in RAS1 or RAS3 have more intermediate occupation numbers associated with multi-reference character. This suggests that for larger systems featuring many more π orbitals, it may be possible to use limited-π RASPT2 with πCPO to keep computational costs low by placing most of the active orbitals in RAS1 or RAS3.

IV. CONCLUSIONS

Singlet–triplet gaps in several diradical organic π-systems, including cyclobutadiene and its derivatives and cyclopentadienyl cation, were calculated using the CPO-based CASPT2 and RASPT2 approaches benchmarked against high-level DEA-EOMCC data including up to 4*p*-2*h* excitations. The goal was to develop a systematic way to choose and subdivide active spaces within the CPO framework and find active spaces that can accurately reproduce the DEA-EOMCC(4*p*-2*h*)-level data, while being small enough to be applicable to larger organic diradicals.

To generate benchmark data for assessing the accuracy of the CPO-based CASPT2 and RASPT2 approaches, we performed a large number of DEA-EOMCC calculations, including full DEA-EOMCC(3*p*-1*h*) and active-space DEA-EOMCC(3*p*-1*h*)$\{N_u\}$ computations, in which the high-order 4*p*-2*h* terms are neglected, and active-space DEA-EOMCC(3*p*-1*h*,4*p*-2*h*)$\{N_u\}$ and DEA-EOMCC(4*p*-2*h*)$\{N_u\}$ calculations, in which 4*p*-2*h* effects are accounted for. We then developed a useful extrapolation scheme that allowed us to capture 3*p*-1*h* and 4*p*-2*h* correlations and the effect of going from the cc-pVDZ basis set to its larger maug-cc-pVTZ counterpart in an accurate and computationally manageable manner. While generating the benchmark DEA-EOMCC information, we investigated the role of high-order 4*p*-2*h* effects, showing that they can be quite important in obtaining accurate singlet–triplet gaps in diradicals, confirming the earlier findings in this regard.[57–59]

We find that the CPO scheme is quite successful for these systems; eCPO and mCPO are highly accurate, with MUEs relative to the DEA-EOMCC(4*p*-2*h*)-level data of 0.3–0.7 kcal/mol, but would usually be cost-prohibitive for systems of practical interest. At the other end of the quality spectrum, nCPO is insufficient, with MUEs of 6.5 and 7.9 kcal/mol. However πCPO has MUEs of 0.7-0.9 kcal/mol almost as good as mCPO and eCPO, and it is much more affordable and thus shows promise for calculation of π-system excitations in larger systems.

Examination of occupation numbers demonstrated that π orbitals, in general, not merely the two nominal participating orbitals, are important contributors to the multi-reference character, but orbitals outside of the π system are effectively either doubly occupied or unoccupied. These observations explain why eCPO and mCPO do not show improvements in accuracy over πCPO, but nCPO is inaccurate. This observation also explains why methods using only a (2,2) active space have not been able to achieve consistent and accurate results. Based on this data, πCPO is recommended for these sorts of systems, providing significant cost savings over full-valence approaches to selecting active spaces. Even greater savings can be obtained by using RASSCF to further reduce the cost of the CI expansion, especially in light of the fact that the RASPT2 MUEs are only 0.1–0.2 kcal/mol higher than CASPT2 MUEs for πCPO.

SUPPLEMENTARY MATERIAL

See supplementary material for active spaces (including visualized orbitals and occupation numbers) and absolute energies for all CASPT2 and RASPT2 calculations presented, exact numbers of CSFs, including for limited-π RASSCF, and ANO-RCC-VTZP results, including a select comparison with a zero IPEA shift.

ACKNOWLEDGMENTS

The authors thank Molly Andersen for helpful assistance. This work was supported in part (S.J.S., D.G.T., and L.G.) by the NSF Grant No. CHE-1464536 for the multi-reference calculations (CASSCF, CASPT2, and RASPT2) and in part (P.P., J.S., and A.O.A.) by the Chemical Sciences, Geosciences and Biosciences Division, Office of Basic Energy Sciences, Office of Science, U.S. Department of Energy under Award No. DE-FG02-01ER15228 for the coupled-cluster calculations.

[1] S. Pedersen, J. L. Herek, and A. H. Zewail, "The validity of the 'diradical' hypothesis: Direct femtosecond studies of the transition-state structures," Science **266**, 1359 (1994).
[2] C.-C. Ko and V. W.-W. Yam, in *Photochromic Materials: Preparation, Properties and Applications*, edited by H. Tian and J. Zhang (Wiley, Hoboken, NJ, 2016), p. 47.
[3] D. Cho, K. C. Ko, and J. Y. Lee, "Quantum chemical approaches for controlling and evaluating intramolecular magnetic interactions in organic diradicals," Int. J. Quantum Chem. **116**, 578 (2016).
[4] R. M. Davis, A. L. Sowers, W. DeGraff, M. Bernardo, A. Thetford, M. C. Krishna, and J. B. Mitchell, "A novel nitroxide is an effective brain redox imaging contrast agent and *in vivo* radioprotector," Free Radical Biol. Med. **51**, 780 (2011).
[5] T. Sugawara, H. Komatsu, and K. Suzuki, "Interplay between magnetism and conductivity derived from spin-polarized donor radicals," Chem. Soc. Rev. **40**, 3105 (2011).
[6] S. Sanvito, "Molecular spintronics," Chem. Soc. Rev. **40**, 3336 (2011).
[7] M. Nakano and B. Champagne, "Nonlinear optical properties in open-shell molecular systems," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **6**, 198 (2016).
[8] T. Minami and M. Nakano, "Diradical character view of singlet fission," J. Phys. Chem. Lett. **3**, 145 (2012).
[9] J. S. S. de Melo, H. D. Burrows, and J. Pina, in *Photochemistry*, edited by A. Albini and E. Fani (Royal Society of Chemistry, Cambridge, 2016), Vol. 43, p. 83.
[10] X. Li and M. L. Tang, "Triplet transport in thin films: Fundamentals and applications," Chem. Commun. **53**, 4429 (2017).
[11] G. J. Hedley, A. Ruseck, and I. D. W. Samuel, "Light harvesting for organic photovoltaics," Chem. Rev. **117**, 796 (2017).
[12] J. Niklas and O. G. Poluektov, "Charge transfer processes in OPV materials as revealed by EPR spectroscopy," Adv. Energy Mater. **7**, 1602226 (2017).
[13] N. M. Gallagher, A. Olankitwanit, and A. Rajca, "High-spin organic molecules," J. Org. Chem. **80**, 1291 (2015).
[14] C. U. Ibeji and D. Ghosh, "Singlet-triplet gaps in polyacenes: A delicate balance between dynamic and static correlations investigated by spin-flip methods," Phys. Chem. Chem. Phys. **17**, 9849 (2015).
[15] D. H. Ess, E. R. Johnson, X. Hu, and W. Yang, "Singlet-triplet energy gaps for diradicals from fractional-spin density-functional theory," J. Phys. Chem. A **115**, 76 (2011).
[16] A. J. Garza, C. A. Jimenez-Hoyos, and G. E. Scuseria, "Electronic correlation without double counting via a combination of spin projected Hartree-Fock and density functional theories," J. Chem. Phys. **140**, 244102 (2014).

[17] O. Demel, K. R. Shamasundar, L. Kong, and M. Nooijen, "Application of double ionization state-specific equation of motion coupled cluster method to organic diradicals," J. Phys. Chem. A **112**, 11895 (2008).
[18] M. Abe, "Diradicals," Chem. Rev. **113**, 7011 (2013).
[19] J. Shen and P. Piecuch, "Merging active-space and renormalized coupled-cluster methods via the CC(*P;Q*) formalism, with benchmark calculations for singlet–triplet gaps in biradical systems," J. Chem. Theory Comput. **8**, 4968 (2012).
[20] Z. B. Zeng, X. Shi, C. Chi, J. T. López Navarrete, J. Casado, and J. Wu, "Pro-aromatic and anti-aromatic π-conjugated molecules: An irresistible wish to be diradicals," Chem. Soc. Rev. **44**, 6578 (2015).
[21] Z. Varga and D. G Truhlar, "$B_2N_2O_4$: Prediction of a magnetic ground state for a light main-group molecule," Inorg. Chem. **54**, 8552 (2015).
[22] Z. Varga and D. G. Truhlar, "Singlet-triplet competition in the low–lying energy states of $C_4O_{4-n}S_n$ (n = 1–3) molecules," Struct. Chem. **26**, 1229 (2015).
[23] J. C. Sancho-García, J. Pittner, P. Čársky, and I. Hubač, "Multireference coupled-cluster calculations on the energy of activation in the automerization of cyclobutadiene: Assessment of the state-specific multireference Brillouin–Wigner theory," J. Chem. Phys. **112**, 8785 (2000).
[24] F. Coester, "Bound states of a many-particle system," Nucl. Phys. **7**, 421 (1958).
[25] F. Coester and H. Kümmel, "Short-range correlations in nuclear wave functions," Nucl. Phys. **17**, 477 (1960).
[26] J. Čížek, "On the correlation problem in atomic and molecular systems. Calculation of wavefunction components in Ursell-type expansion using quantum-field theoretical methods," J. Chem. Phys. **45**, 4256 (1966).
[27] J. Čížek, "On the use of the cluster expansion and the technique of diagrams in calculations of correlation effects in atoms and molecules," in *Advances in Chemical Physics: Correlation Effects in Atoms and Molecules*, Vol. 14, edited by R. LeFebvre and C. Moser (John Wiley & Sons, 1969), pp. 35–89.
[28] J. Čížek and J. Paldus, "Correlation problems in atomic and molecular systems. III. Rederivation of the coupled-pair many-electron theory using the traditional quantum chemical methods," Int. J. Quantum Chem. **5**, 359 (1971).
[29] J. Paldus, J. Čížek, and I. Shavitt, "Correlation problems in atomic and molecular systems. IV. Extended coupled-pair many-electron theory and its application to the BH_3 molecule," Phys. Rev. A **5**, 50 (1972).
[30] G. D. Purvis III and R. J. Bartlett, "A full coupled-cluster singles and doubles model: The inclusion of disconnected triples," J. Chem. Phys. **76**, 1910 (1982).
[31] K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, "A fifth-order perturbation comparison of electron correlation theories," Chem. Phys. Lett. **157**, 479 (1989).
[32] S. Ghosh, C. J. Cramer, D. G. Truhlar, and L. Gagliardi, "Generalized-active-space pair-density functional theory: An efficient method to study large, strongly correlated, conjugated systems," Chem. Sci. **8**, 2741 (2017).
[33] T. Saito, S. Nishihara, S. Yamanaka, Y. Kitagawa, T. Kawakami, S. Yamada, H. Isobe, M. Okumura, and K. Yamaguchi, "Symmetry and broken symmetry in molecular orbital description of unstable molecules IV: Comparison between single- and multi-reference computational results for antiaromtic molecules," Theor. Chem. Acc. **130**, 749 (2011).
[34] B. O. Roos, P. R. Taylor, and P. E. M. Siegbahn, "A complete active space SCF method (CASSCF) using a density matrix formulated super-CI approach," Chem. Phys. **48**, 157 (1980).
[35] B. O. Roos, in *Advances in Chemical Physics: Ab Initio Methods in Quantum Chemistry Part 2*, edited by K. P. Lawley (John Wiley & Sons, Hoboken, NJ, USA, 1987), Vol. 69, p. 399.
[36] P. G. Szalay, T. Müller, G. Gidofalvi, H. Lischka, and R. Shepard, "Multiconfiguration self-consistent field and multireference configuration interaction methods and applications," Chem. Rev. **112**, 108 (2012).
[37] K. Ruedenberg, M. W. Schmidt, M. M. Gilbert, and S. T. Elbert, "Are atoms intrinsic to molecular electronic wavefunctions? I. The FORS model," Chem. Phys. **71**, 41 (1982).
[38] L. Gagliardi, D. G. Truhlar, G. Li Manni, R. K. Carlson, C. E. Hoyer, and J. L. Bao, "Multiconfiguration pair-density functional theory: A new way to treat strongly correlated systems," Acc. Chem. Res. **50**, 66 (2017).
[39] K. Andersson, P.-Å. Malmqvist, B. O. Roos, A. J. Sadlej, and K. Wolinski, "Second-order perturbation theory with a CASSCF reference function," J. Phys. Chem. **94**, 5483 (1990).
[40] K. Andersson, P.-Å. Malmqvist, and B. O. Roos, "Second-order perturbation theory with a complete active space self-consistent field reference function," J. Chem. Phys. **96**, 1218 (1992).
[41] F. Aquilante, J. Autschbach, R. K. Carlson, L. F. Chibotaru, M. G. Delcey, L. De Vico, I. Fdez. Galván, N. Ferré, L. M. Frutos, L. Gagliardi, M. Garavelli, A. Giussani, C. E. Hoyer, G. Li Manni, H. Lischka, D. Ma, P.-Å. Malmqvist, T. Müller, A. Nenov, M. Olivucci, T. B. Pedersen, D. Peng, F. Plasser, B. Pritchard, M. Reiher, I. Rivalta, I. Schapiro, J. Segarra-Martí, M. Stenrup, D. G. Truhlar, L. Ungur, A. Valentini, S. Vancoillie, V. Veryazov, V. P. Vysotskiy, O. Weingart, F. Zapata, and R. Lindh, "Molcas 8: New capabilities for multiconfigurational quantum chemical calculations across the periodic table," J. Comput. Chem. **37**, 506 (2016).
[42] P.-Å. Malmqvist, A. Rendell, and B. O. Roos, "The restricted active space self-consistent-field method, implemented with a split graph unitary group approach," J. Phys. Chem. **94**, 5477 (1990).
[43] M. W. Schmidt and M. S. Gordon, "The construction and interpretation of MCSCF wavefunctions," Annu. Rev. Phys. Chem. **49**, 233 (1998).
[44] V. Veryazov, P.-Å. Malmqvist, and B. O. Roos, "How to select active space for multiconfigurational quantum chemistry?," Int. J. Quantum Chem. **111**, 3329 (2011).
[45] S. Keller, K. Boguslawski, T. Janowski, M. Reiher, and P. Pulay, "Selection of active spaces for multiconfigurational wavefunctions," J. Chem. Phys. **142**, 244104 (2015).
[46] C. J. Stein and M. Reiher, "Automated selection of active orbital spaces," J. Chem. Theory Comput. **12**, 1760 (2016).
[47] O. Tishchenko, J. J. Zheng, and D. G. Truhlar, "Multireference model chemistries for thermochemical kinetics," J. Chem. Theory Comput. **4**, 1208 (2008).
[48] J. L. Bao, A. Sand, L. Gagliardi, and D. G. Truhlar, "Correlated-participating-orbitals pair-density functional method and application to multiplet energy splittings of main-group divalent radicals," J. Chem. Theory Comput. **12**, 4274 (2016).
[49] J. L. Bao, S. O. Odoh, L. Gagliardi, and D. G. Truhlar, "Predicting bond dissociation energies of transition metal compounds by multiconfiguration pair-density functional theory and second-order perturbation theory based on correlated participating orbitals and separated pairs," J. Chem. Theory Comput. **13**, 616 (2017).
[50] U. S. Mahapatra, B. Datta, and D. Mukherjee, "A size-consistent state-specific multireference coupled cluster theory: Formal developments and molecular applications," J. Chem. Phys. **110**, 6171 (1999).
[51] A. Balková and R. J. Bartlett, "A multireference coupled-cluster study of the ground state and lowest excited states of cyclobutadiene," J. Chem. Phys. **101**, 8972 (1994).
[52] M. Eckert-Maksić, M. Vazdar, M. Barbatti, H. Lischka, and Z. B. Maksić, "Automerization reaction of cyclobutadiene and its barrier height: An *ab initio* benchmark multireference average-quadratic coupled cluster study," J. Chem. Phys. **125**, 064310 (2006).
[53] M. Vazdar, M. Eckert-Maksić, and H. Lischka, "The antiferromagnetic spin coupling in non-Kekulé acenes—Impressive polyradical character revealed by high-level multireference methods," ChemPhysChem **17**, 2013 (2016).
[54] M. Musiał, S. A. Kucharski, and R. J. Bartlett, "Multireference double electron attached coupled cluster method with full inclusion of the connected triple excitations: MR-DA-CCSDT," J. Chem. Theory Comput. **7**, 3088 (2011).
[55] M. Musiał, "Multireference Fock space coupled cluster method in the effective and intermediate Hamiltonian formulation for the (2,0) sector," J. Chem. Phys. **136**, 134111 (2012).
[56] M. Musiał, K. Kowalska-Szojda, D. I. Lyakh, and R. J. Bartlett, "Potential energy curves via double electron-attachment calculations: Dissociation of alkali metal dimers," J. Chem. Phys. **138**, 194103 (2013).
[57] J. Shen and P. Piecuch, "Doubly electron-attached and doubly ionized equation-of-motion coupled-cluster methods with 4-particle–2-hole and 4-hole–2-particle excitations and their active-space extensions," J. Chem. Phys. **138**, 194102 (2013).
[58] J. Shen and P. Piecuch, "Doubly electron-attached and doubly ionised equation-of-motion coupled-cluster methods with full and active-space treatments of 4-particle–2-hole and 4-hole–2-particle excitations: The role of orbital choices," Mol. Phys. **112**, 868 (2014).
[59] A. O. Ajala, J. Shen, and P. Piecuch, "Economical doubly electron-attached equation-of-motion coupled-cluster methods with an active-space treatment of three-particle–one-hole and four-particle–two-hole excitations," J. Phys. Chem. A **121**, 3469 (2017).
[60] P. Piecuch and R. J. Bartlett, "EOMXCC: A new coupled-cluster method for electronic excited states," Adv. Quantum Chem. **34**, 295 (1999).
[61] R. J. Bartlett and M. Musiał, "Coupled-cluster theory in quantum chemistry," Rev. Mod. Phys. **79**, 291 (2007).

[62]S. V. Levchenko and A. I. Krylov, "Equation-of-motion spin-flip coupled-cluster model with single and double substitutions: Theory and application to cyclobutadiene," J. Chem. Phys. **120**, 175 (2004).
[63]P. Piecuch, "Active-space coupled-cluster methods," Mol. Phys. **108**, 2987 (2010).
[64]T. H. Dunning, "Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen," J. Chem. Phys. **90**, 1007 (1989).
[65]E. Papajak and D. G. Truhlar, "Efficient diffuse basis sets for density functional theory," J. Chem. Theory Comput. **6**, 597 (2010).
[66]M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, Jr., "General atomic and molecular electronic structure system," J. Comput. Chem. **14**, 1347 (1993).
[67]P. Piecuch, S. A. Kucharski, K. Kowalski, and M. Musiał, "Efficient computer implementation of the renormalized coupled-cluster methods: The R-CCSD[T], R-CCSD(T), CR-CCSD[T], and CR-CCSD(T) approaches," Comput. Phys. Commun. **149**, 71 (2002).
[68]K. Kowalski and P. Piecuch, "New coupled-cluster methods with singles, doubles, and noniterative triples for high accuracy calculations of excited electronic states," J. Chem. Phys. **120**, 1715 (2004).
[69]M. Włoch, J. R. Gour, K. Kowalski, and P. Piecuch, "Extension of renormalized coupled-cluster methods including triple excitations to excited electronic states of open-shell molecules," J. Chem. Phys. **122**, 214107 (2005).
[70]P. Piecuch, J. R. Gour, and M. Włoch, "Left-eigenstate completely renormalized equation-of-motion coupled-cluster methods: Review of key concepts, extension to excited states of open-shell systems, and comparison with electron-attached and ionized approaches," Int. J. Quantum Chem. **109**, 3268 (2009).
[71]P.-Å. Malmqvist, K. Pierloot, A. R. M. Shahi, C. J. Cramer, and L. Gagliardi, "The restricted active space followed by second-order perturbation theory method: Theory and application to the study of CuO_2 and Cu_2O_2 systems," J. Chem. Phys. **128**, 204109 (2008).
[72]E. Papajak, H. R. Leverentz, J. Zheng, and D. G. Truhlar, "Efficient diffuse basis sets: cc-pVxZ+ and maug-cc-pVxZ," J. Chem. Theory Comput. **5**, 1197 (2009).
[73]P. O. Widmark, P.-Å. Malmqvist, and B. O. Roos, "Density matrix averaged atomic natural orbital (ANO) basis sets for correlated molecular wave functions," Theor. Chim. Acta **77**, 291 (1990).
[74]F. Aquilante, T. B. Pedersen, A. Sánchez de Merás, and H. J. Koch, "Fast noniterative orbital localization for large molecules," J. Chem. Phys. **125**, 174101 (2006).
[75]F. Aquilante, L. De Vico, N. Ferré, G. Ghigo, P.-Å. Malmqvist, P. Neogrády, T. B. Pedersen, M. Pitoňák, M. Reiher, B. O. Roos, L. Serrano-Andrés, M. Urban, V. Veryazov, and R. Lindh, "Software news and update MOLCAS 7: The next generation," J. Comput. Chem. **31**, 224 (2010).
[76]V. Veryazov, P. O. Widmark, L. Serrano-Andrés, R. Lindh, and B. O. Roos, "2MOLCAS as a development platform for quantum chemistry software," Int. J. Quantum Chem. **100**, 626 (2004).
[77]G. Kovacevic and V. Veryazov, "Luscus: Molecular viewer and editor for MOLCAS," J. Cheminf. **7**, 16 (2015).
[78]G. Ghigo, B. O. Roos, and P.-Å. Malmqvist, "A modified definition of the zeroth-order Hamiltonian in multiconfigurational perturbation theory (CASPT2)," Chem. Phys. Lett. **396**, 142 (2004).
[79]K. Raghavachari, J. A. Pople, E. S. Replogle, M. Head-Gordon, and N. C. Handy, "Size-consistent Brueckner theory limited to double and triple substitutions," Chem. Phys. Lett. **167**, 115 (1990).
[80]P. G. Szalay and R. J. Bartlett, "Multi-reference averaged quadratic coupled-cluster method: A size-extensive modification of multireference CI," Chem. Phys. Lett. **214**, 481 (1993).