JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

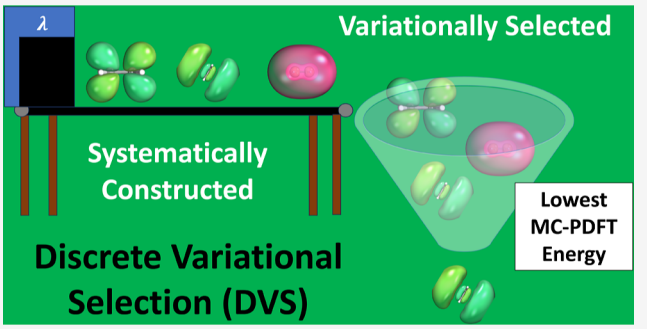
Variational Active Space Selection with Multiconfiguration Pair-Density Functional Theory

Daniel S. King, Donald G. Truhlar,* and Laura Gagliardi*

Cite This: J. Chem. Theory Comput. 2023, 19, 8118−8128 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The selection of an adequate set of active orbitals for modeling strongly correlated electronic states is difficult to automate because it is highly dependent on the states and molecule of interest. Although many approaches have shown some success, no single approach has worked well in all cases. In light of this, we present the "discrete variational selection" (DVS) approach to active space selection, in which one generates multiple trial wave functions from a diverse set of systematically constructed active spaces and then selects between these wave functions variationally. We apply this DVS approach to 207 vertical excitations of small-to-medium-sized organic and inorganic molecules (with 3 to 18 atoms) in the QUESTDB database by (i) constructing various sets of active space orbitals through diagonalization of parametrized operators and (ii) choosing the result with the lowest average energy among the states of interest. This approach proves ineffective when variationally selecting between wave functions using the density matrix renormalization group (DMRG) or complete active space self-consistent field (CASSCF) energy but is able to provide good results when variationally selecting between wave functions using the energy of the translated PBE (tPBE) functional from multiconfiguration pair-density functional theory (MC-PDFT). Applying this DVS-tPBE approach to selection among state-averaged DMRG wave functions, we obtain a mean unsigned error of only 0.17 eV using hybrid MC-PDFT. This result matches that of our previous benchmark without the need to filter out poor active spaces and with no further orbital optimization following active space selection of the SA-DMRG wave functions. Furthermore, we find that DVS-tPBE is able to robustly and effectively select between the new SA-DMRG wave functions and our previous SA-CASSCF results.



1. INTRODUCTION

The accurate treatment of excited electronic states of molecules is a long-standing and active area of research in computational chemistry.[1−18] It is especially difficult when a single-determinant ground state provides a poor reference for computing the excited states (e.g., double excitations[19,20] or strongly correlated systems[21−25]). A useful form of wave function for overcoming such difficulties is the complete active space configuration interaction (CASCI) trial function

$$|\Psi_{\mathrm{CASCI}}\rangle = |22\ldots2\rangle \wedge \sum_{n_1 n_2 \ldots n_L} C_{n_1 n_2 \ldots n_L} |n_1 n_2 \ldots n_L\rangle \tag{1}$$

in which $|22\cdots2\rangle$ is a single Slater determinant consisting of doubly occupied orbitals (called inactive orbitals in the CASCI context), the $C_{n_1 n_2 \ldots n_L}$ are coefficients, and the determinants $|n_1 n_2 \ldots n_L\rangle$ span the space of all possible configurations obtained by distributing a fixed number N_{elec} of active electrons among L active orbitals. Each determinant $|n_1 n_2 \ldots n_L\rangle$ is defined by its orbital occupation numbers $n_i \in \{0, \uparrow, \downarrow, 2\}$ of the active electrons in the active orbitals, and diagonalization of the Hamiltonian in this space (and in the mean field of $|22\cdots2\rangle$) is known as CASCI. However, because the size of the space scales exponentially with the number of orbitals L, this approach is only feasible up to active space sizes of about 20 electrons in 20 orbitals.[26]

Many methods exist to approximate the solution for the coefficients $C_{n_1 n_2 \ldots n_L}$ in eq 1.[27] Among the most successful approaches is the density matrix renormalization group (DMRG) configuration interaction method,[13,28−40] in which the coefficients of eq 1 are approximated by the matrix product[32]

$$C_{n_1 n_2 \ldots n_L} = \sum_{ij\ldots(L-1)} A_i^{n_1} A_{ij}^{n_2} A_{jk}^{n_3} \cdots A_{(L-1)}^{n_L} \tag{2}$$

In eq 2, each possible occupation of each active orbital n_i is given by its own matrix or vector A_i^n. The maximum inner dimension of these matrices is called the bond order M and is the number of states retained during the renormalization step. As M

Received: July 20, 2023
Revised: September 24, 2023
Accepted: October 19, 2023
Published: October 31, 2023

https://doi.org/10.1021/acs.jctc.3c00792
J. Chem. Theory Comput. 2023, 19, 8118−8128

$\rightarrow \infty$, results obtained with this method approach those obtained with full diagonalization (although useful results for well-chosen active spaces are generally obtainable with practical values of M). Using this approach, it is possible to describe active spaces with up to about 100 orbitals.[41]

The success of CAS-based methods relies heavily on the construction and selection of the orbitals defining the active space because this selection affects both the convergence of self-consistent-field iterations and the quality of the energetic results. Variationally optimizing the active-space orbitals is known as CAS self-consistent field (CASSCF)[42] when used with a full-configuration-interaction solver or as DMRG-SCF[43,44] when used with a density-matrix-renormalization-group solver. To try to obtain a consistent treatment of multiple states, one may optimize the state-averaged (SA) energy with respect to the active orbitals, yielding SA-CASSCF[45] or SA DMRG-SCF (SA-DMRG-SCF).[10,44,46] (If orbitals are predetermined rather than optimized, then one obtains SA-DMRG). We emphasize two difficulties with the conventional methods of optimizing orbitals: (i) the energetic optimization is prone to converging to local minima, and (ii) SA variational optimization of orbitals is not necessarily optimal for computing energy differences between states, especially when states with different characters are considered;[47] the latter of these difficulties is made worse by the fact that the CASSCF orbitals are generally optimized without regard for post-CAS correlation generally included in the computation of excitation energies. Furthermore, orbital optimization significantly increases the cost of the computation. Thus, although SCF generally helps improve the quality of the active space, it does not eliminate the need to develop good active space construction and selection schemes for excited-state calculations and comes at a computational cost.

Because of the above considerations, active-space construction and selection remains a vigorous area of research, and several approaches have been proposed to date.[18,41,48−78] The most commonly applied method involves chemical intuition with trial and error.[50,52] However, this approach is unsystematic and is difficult to apply in a high-throughput fashion. In recent years, there has been much interest in developing more systematic methodology for fashioning active spaces.[18,41,48,49,51,53−77] A tool called AVAS, developed by Sayfutyarova and co-workers,[58] allows one to semiautomatize the active-space construction by using molecular orbitals that overlap optimally with a user-selected set of atomic orbitals. Other approaches involve some preliminary calculations, such as the natural orbital occupancies of a unrestricted Hartree−Fock (UHF) calculation,[48,49,53,79,80] entanglement information from a large DMRG calculation,[41,55,81,82] the quantitative accuracy of some physical observable such as the dipole moment,[77] machine learning predictive models,[63,66] and physically motivated equations based on information such as Hartree−Fock (HF) matrix elements.[72,74] An assumption of all of these approaches is that the key physics necessary to construct and select the active space can be captured by a preliminary calculation (UHF, DMRG, etc.). However, as we will show, even when selecting large active spaces (e.g., with 40 orbitals) for small molecules, it is difficult to make even qualitatively accurate predictions of excitation energies for any given excitation with a single method.

In recent work,[72] we employed one such automated approach, approximate pair coefficient (APC) active space selection, on the extensive QUESTDB database[83] of accurate vertical excitation energies for small-to-medium-sized organic systems. Through this, we were able to carry out extensive benchmarking of post-SA-CASSCF methods such as n-electron valence perturbation theory[84,85] and multiconfiguration pair-density functional theory (MC-PDFT)[86] using the translated PBE (tPBE) functional with active spaces generated by the automated approach. However, to ensure accurate evaluation of the post-SA-CASSCF methods and distinguish errors arising from poor active spaces, we considered only the active spaces for which the SA-CASSCF result fell within 1.1 eV of the best estimate in the QUESTDB database.[72] This criterion was satisfied for 363−436 (68−82%) of the 532 excitations in the database, depending on the active-space size and basis set. Although the APC scheme for these cases proved to be competitive with active spaces selected by hand,[87−89] it was observed that the remaining active spaces (18−32% of the excitations) exhibited very high errors, sometimes exceeding 5 eV. Consequently, the predictive utility of these active spaces for our purposes was only partially acceptable.

In the present study, our objective is to develop a new framework for active space selection that is more broadly accurate for predicting vertical excitation energies. The key element of the new method is the premise that no single active-space-selection scheme will be successful in all cases. Therefore, we hypothesize that the important missing component of the current schemes is the lack of a way to effectively *choose* between active spaces generated by different methods and different parameters. To address this, we propose the "discrete variational selection" (DVS) approach to active space selection in which (i) one generates trial wave functions with a variety of active spaces constructed with different methods (e.g., any of those mentioned or cited above) or different parameters, and then (ii) one chooses between the generated active-space wave functions variationally.

In this work, we apply this DVS approach to the calculation of 207 vertical excitations of small-to-medium sized main-group molecules in the QUESTDB database.[83] These excitations provide a rich variety of different types of excitations (e.g., Rydberg and valence excitations of organic molecules, including both $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ excitations and excitations of inorganic molecules) and thus present a demanding challenge to the systematic prediction of their excitation energies. We find that the DVS scheme is unsuccessful when variationally selecting between results using the CASSCF/DMRG energy but performs well when applied using the tPBE energy from MC-PDFT. Applying this DVS-tPBE approach to selection among systematically constructed wave functions with SA-DMRG, we are able to obtain a mean unsigned error of only 0.17 eV with hybrid MC-PDFT. This result reproduces that of our previous benchmarks of hybrid MC-PDFT[74] without the need to filter out poor active spaces and with no further orbital optimization following the active space selection of the SA-DMRG wave functions. Furthermore, we find that DVS-tPBE is able to robustly select between the newly generated SA-DMRG wave functions and our previously generated SA-CASSCF results.[74]

2. THEORY AND METHODS

In this section, we provide an overview of MC-PDFT[86] and hybrid MC-PDFT[90] and provide a description of the approach used in this work to systematically construct the active spaces and SA-DMRG wave functions for DVS-tPBE. Finally we describe the QUESTDB data used to judge the performance of this method.

https://doi.org/10.1021/acs.jctc.3c00792
J. Chem. Theory Comput. 2023, 19, 8118−8128

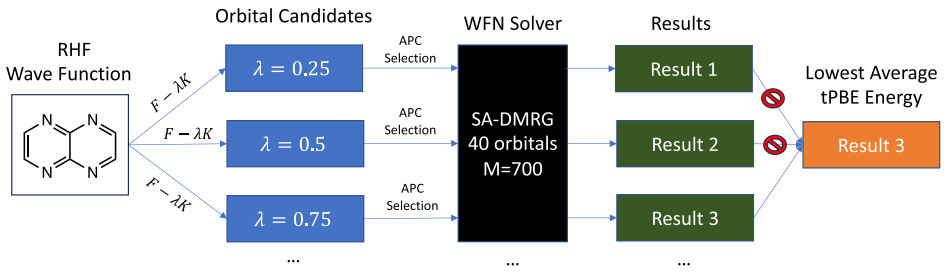


Figure 1. Schematic of the scheme used to systematically construct active spaces for DVS-tPBE. Starting from an RHF or ROHF wave function, different sets of orbitals are generated by diagonalizing $F - \lambda K$ in the space of virtual orbitals. Active spaces of 40 orbitals are then selected from these orbital candidates using APC selection.[72,74] Wave functions are then generated using these selected active spaces by SA-DMRG. The final step represents the DVS-tPBE approach, in which the final result is chosen as the one with the lowest sum of the tPBE energies between the two states of interest.

2.1. Multiconfiguration Pair-Density Functional Theory. The energy expression of MC-PDFT may be written as[86]

$$E_{\text{MC-PDFT}} = V_{nn} + \sum_{ij} h_{ij}\gamma_{ij} + \frac{1}{2}\sum_{ijkl} g_{ijkl}\gamma_{ij}\gamma_{kl} + E_{\text{NE}} \tag{3}$$

where V_{nn} is the nuclear repulsion, i, j, k, and l are orbital indices, h_{ij} is a one-electron integral, γ_{ij} is the one-electron density matrix, g_{ijkl} is a two-electron integral, and E_{NE} is the nonclassical-energy functional. In most of our work, E_{NE} is written as a function of the electron density ρ and the on-top pair density Π and is called an on-top density functional. Recently, we have begun to explore different types of nonclassical-energy functionals derived from machine learning[91] or the density coherence,[92] and we refer the reader to a recent review.[17] However, all practical applications so far have employed a translated version of the PBE[93] Kohn–Sham functional which is an on-top functional denoted as tPBE.

The on-top functional may also be combined with the wave function exchange–correlation energy to form hybrid MC-PDFT,[90] for which the energy expression becomes

$$E_{\text{HMC-PDFT}} = XE_{\text{SA-CASSCF}} + (1 - X)E_{\text{MC-PDFT}} \tag{4}$$

where $E_{\text{SA-CASSCF}}$ is the SA-CASSCF energy computed by wave function theory, and X is a parameter. We have often found good results using tPBE with $X = 0.25$,[74] which is called tPBE0.

2.2. Systematically Constructed Active Spaces for DVS-tBPE. The outline of the scheme used to systematically generate active spaces for DVS-tPBE is shown in Figure 1. The approach consists of (i) calculation of initial HF wave functions, (ii) virtual orbital construction via diagonalization of a parametrized operator, (iii) selection of active spaces with the approximate-pair-coefficient selection (APC) method,[72,74] and (iv) generation of SA-DMRG wave functions in these active spaces using Block2.[94] In the following, we explain these components step-by-step.

2.2.1. Hartree–Fock Calculations. HF orbitals were generated for closed-shell singlet ground states by restricted Hartree–Fock (RHF) theory[95] and for doublet ground states by restricted open-shell Hartree–Fock (ROHF) theory[96] using the aug-cc-pVTZ basis[97,98] as was used for the theoretical best estimates listed in the QUESTDB database. (Note that good results for these excitations are likely achievable with the smaller jun-cc-pVTZ[99] basis as was observed in our previous benchmark).[74]

The definition of the Fock operator in ROHF theory is ambiguous, but the choice must be specified as it affects the basis of orbitals from which we select to form the SA-DMRG wave functions as well as the inputs into the APC theory[72,74] used to select the active space from these orbitals. For the present article, we employ Roothaan's effective Fock operator,[96,100] which is the default choice in PySCF.[101,102]

2.2.2. Orbital Construction. Starting from the set of canonical orbitals obtained from the RHF or ROHF wave function, we index the doubly occupied orbitals with i and the virtual orbitals with a. We then generate multiple trial orbital sets for a calculation by diagonalizing the parametrized operator

$$G = F - \lambda K \tag{5}$$

in the space of the RHF or ROHF virtual orbitals, where F and K are the Fock and exchange matrices generated from the RHF or ROHF density matrix, and λ is a tuneable parameter used to generate different sets of orbitals. Each of these trial orbital sets serves as a set of candidate orbitals from which the active space will be selected for separate multiconfigurational wave function calculations; after their generation in this step, they remain unchanged. The next step serves to select 40 active orbitals and a set of inactive orbitals from each of these initial sets of orbitals.

2.2.3. APC Active Space Selection. The APC method is a method for estimating the one-orbital reduced density matrix entropies of candidate orbitals for the active space from HF matrix elements.[72] Using this approach, it is possible to efficiently estimate orbital importance for the active space (and thus rank the orbitals appropriately) as a higher orbital entropy is a measure of higher multireference character. For a doubly occupied orbital i and virtual orbital a, the APC between these two orbitals is defined by[72]

$$C_{ia} = \frac{0.5K_{aa}}{F_{aa} - F_{ii} + \sqrt{(0.5K_{aa})^2 + (F_{aa} - F_{ii})^2}} \tag{6}$$

where F and K are again the Fock operator and exchange operator generated from the HF density matrix. The entropies of doubly occupied orbitals and virtual orbitals are defined as

$$S_i = -\frac{1}{1+\sum_a C_{ia}^2}\ln\frac{1}{1+\sum_a C_{ia}^2} - \frac{\sum_a C_{ia}^2}{1+\sum_a C_{ia}^2}\ln\frac{\sum_a C_{ia}^2}{1+\sum_a C_{ia}^2} \tag{7}$$

and

$$S_a = -\frac{1}{1+\sum_i C_{ia}^2}\ln\frac{1}{1+\sum_i C_{ia}^2} - \frac{\sum_i C_{ia}^2}{1+\sum_i C_{ia}^2}\ln\frac{\sum_i C_{ia}^2}{1+\sum_i C_{ia}^2} \tag{8}$$

https://doi.org/10.1021/acs.jctc.3c00792
J. Chem. Theory Comput. 2023, 19, 8118–8128

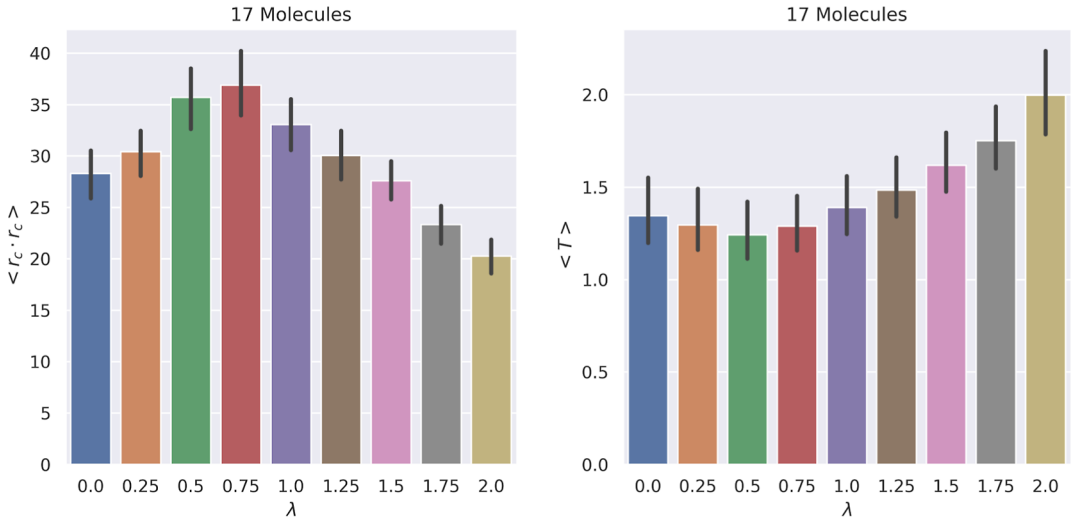


Figure 2. Left: averaged squared distance from the centroid $< \mathbf{r}_c \cdot \mathbf{r}_c >$ over the selected active orbitals for the 17 unique molecules of the 30-excitation test set. Right: averaged kinetic energy of the selected active orbitals for the 17 unique molecules of the 30-excitation test set.

where the sums over i includes all HF doubly occupied orbitals, and the sums over a initially includes all virtual orbitals generated in the orbital construction step. We will eventually select high-entropy orbitals for the active space, but in our previous work,[74] we have found the entropies calculated with the full sums to be overly biased toward doubly occupied orbitals, resulting in less-than-optimal active spaces. Therefore, we use a virtual-orbital removal step, in which the C_{ia} involving the highest-entropy virtual orbital is removed from the sums in eq 7, and the entropies are recalculated. After N such virtual-orbital removal steps are taken, the entropies of the removed virtual orbitals are set to the maximum entropy of the remaining orbitals plus some small value, decreasing in order of removal; we have found good results for small-to-medium-sized organic molecules with $N = 2$,[74] which is used in this work.

Having calculated orbital entropies for each trial set of doubly occupied orbitals i and virtual orbitals a (constructed in step (ii)), the 40 highest-entropy orbitals are selected as the active orbitals, and the other orbitals are dropped from the active space. Any dropped doubly occupied orbitals become inactive, whereas dropped virtual orbitals become secondary. As such, there are always 40 active orbitals in the active space of each subsequent SA-DMRG calculation, and the number of inactive orbitals is the number of doubly occupied orbitals dropped from the active space in the above selection stage. We note that although we do not exclude core orbitals from selection, they are highly biased against by the APC scheme (eq 6) and mostly harmless if added to the active space (generally when one has exhausted all other orbitals). The number of active electrons in each calculation is set to two times the number of doubly occupied orbitals remaining in the active set, and the number of inactive electrons is equal to two times the number of inactive orbitals.

2.2.4. Computation of SA-DMRG Wave Functions. Having selected a 40-orbital active space for each set of trial orbitals in step (iii) (one for each value of λ selected in step (ii)), density matrix renormalization group calculations were carried out without reoptimization of orbitals by using the SA-DMRG in Block2[94] as integrated into PySCF.[101,102] The maximum bond dimension of these calculations (i.e., the maximum number of renormalized states) was fixed at $M = 700$. The choice of selecting 40-orbital active spaces with a bond dimension of $M = 700$ was made because it provided good accuracy while still remaining computationally affordable (here defined as being able to run on 24 Intel Cascade Lake cores with 96 GB of memory in less than a few hours for all systems). As in our previous study,[74] excited-state wave functions were calculated in a SA fashion averaging over the ground state and the required number of excited states (for example, to approximate the 2 $^1A_2$ state, we would include the ground state of symmetry $^1A_1$ and two states of symmetry $^1A_2$).[74]

We then select among the multiple wave function results generated for each excitation via steps (i−iv) by variational selection with tPBE. In particular, we select the SA-DMRG wave function that yields the lowest sum of tPBE absolute energies between the ground state and the excited state of interest. Energies with tPBE were calculated using a version of PySCF that incorporates the MRH code[103] now available in PySCF-Forge.[104] Grid integration was carried out for evaluation of the on-top functional with fineness grids_level = 3, as judged to be sufficient in our previous benchmark study.[74]

2.3. Benchmarking Data. We investigate the approach described above on a subset of theoretical best estimates in the QUESTDB dataset[83] for vertical excitation energies of small-to-medium-sized main-group molecules. This set of excitations includes many of the most widely studied molecules of the quantum chemistry community (e.g., water, ethylene, and naphthalene) as well as a rich variety of different excited states (e.g., valence, Rydberg, $n \rightarrow \pi^*$, and $\pi \rightarrow \pi^*$ excitations in many organic molecules with as many as 18 atoms plus excitations in H_2S, HPO, HPS, HSiF, and HNO). Thus, it presents a difficult challenge for any active space selection scheme that aims to be predictive in its calculations. We form a subset of these excitations by applying the following constraints:

- Excitations must be labeled as "safe" in the original QUESTDB dataset (considered by the authors of that work as chemically accurate or within 0.05 eV of the FCI limit for the given geometry and basis set).[83]
- The full symmetry of the molecule must be supported in the CAS module of PySCF; this limits us to molecules with symmetries C_s, C_{2v}, C_{2h}, and D_{2h}.
- The symmetry of the states must be unambiguously specified with regard to the axis convention. This excludes

https://doi.org/10.1021/acs.jctc.3c00792
J. Chem. Theory Comput. 2023, 19, 8118−8128

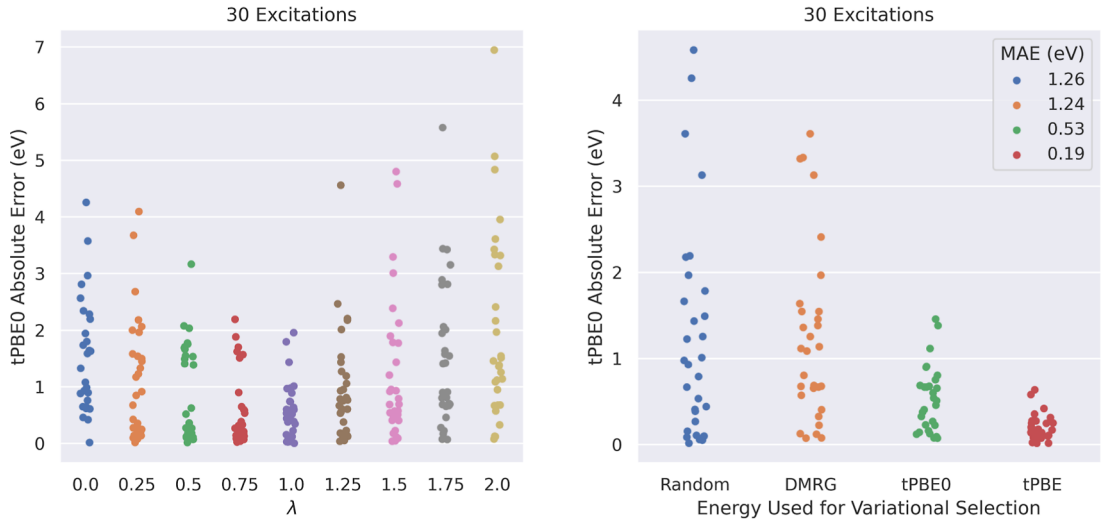


Figure 3. Left: tPBE0 absolute error of 30 difficult vertical excitations using active spaces selected with different values of λ. Right: errors of these excitations with different values of λ selected variationally by different energies: random selection, variational selection with DMRG, tPBE0, and tPBE.

excitations involving the irreps B_1 and B_2 in C_{2v} and B_{1g}, B_{2g}, B_{3g}, B_{1u}, B_{2u}, B_{3u} in D_{2h}.

These criteria exclude any possibility of the calculated excitations being inaccurate due to unavailable symmetry or mislabeled symmetry. After eliminating data according to these criteria, we are left with a set of 207 excitations for testing the present approach (199 excitations from singlet states and eight from doublet states).

2.3.1. 2^1A_g State of Ethylene. Special attention is given to the theoretical best estimate listed in QUESTDB for the 2^1A_g state of ethylene, which is characterized by Loos and co-workers as a valence $(\pi,\pi) \rightarrow (\pi^*,\pi^*)$ double excitation at roughly 12.15 eV, referencing a 2004 study by Barbatti et al.[20,105] However, in the comprehensive 2014 study on the excited states of ethylene carried out by Feller et al.,[106] the 2^1A_g state of ethylene is clearly characterized by both the experiment and theory as a single $(\pi,3p)$ Rydberg excitation at about 8.45 eV. Although we have been able to converge to the double excitation in the ^1A_g irrep described by Loos et al. with some active space selections, it is clear that our best estimates converge to the lower Rydberg excitation supported by the Feller et al.[106] study. Thus, we have changed the theoretical best estimate of this excitation in the QUESTDB database to the value of 8.45 eV reported by Feller et al.[106]

3. RESULTS

3.1. 30-Excitation Tests. We first show the robustness of the new active space selection approach by carrying out calculations for a set of 30 excitations for which APC selection in the aug-cc-pVTZ basis[74] had a SA-CASSCF tPBE0 error greater than 0.55 eV; these 30 excitations involve a set of 17 molecules. We generated nine active spaces for each excitation by the method explained in Section 2.2, using λ equal to 0, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, and 2.0 in step (ii).

Figure 2 shows trends in the orbital character of the selected active orbitals as λ is varied from 0 to 2 for the 17 molecules present in the 30-excitation test subset. The left side of Figure 2 shows the average squared distance from the centroid (with the centroid defined as the average coordinates of all of the nuclei in the molecule) over the selected active orbitals. The figure shows that different values of λ lead to significantly different averaged diffuse character of the selected orbitals, with the most diffuse character for λ = 0.75. The right side of Figure 2 shows the average kinetic energy of the selected orbitals and illustrates the well-known quantum mechanical relation by which average kinetic energy is inversely related to average spatial extent. Thus, modifications of λ provide an effective means to explore active spaces targeting different kinds of states, e.g., Rydberg vs valence excitations. This is demonstrated clearly in the calculation of the 2^1A_g state of ethylene, where λ = 0.25 selects an active space converging to the valence doubly excited ^1A_g of Loos and co-workers,[20] while λ = 0.75 converges to the lower-energy singly excited Rydberg state (Supporting Information).[106]

We next examined the accuracy of excitation energies calculated from the selected active spaces by tPBE0. To prevent confusion, we stress that although the DVS-tPBE selection scheme employs the tPBE functional for variational selection, our calculations of excitation energies are based on tPBE0. These choices simply reflect that tPBE performs better in the selection scheme (as discussed below), whereas tPBE0 gives more accurate excitation energies (as shown in previous work[74] and discussed below.).

The left side of Figure 3 shows the absolute error in the tPBE0 calculations of the excitation energies with active spaces generated by the nine values of λ. As can be seen, no single value of λ yields accurate results for all 30 cases. For each value of λ, several excitations have an error greater than 1 eV. Although the mean absolute error is lowest for λ = 1 (0.56 eV), this is much larger than the mean absolute error of our previous benchmark results (0.19 eV) when we excluded poor active spaces. However, for all 30 excitations, the new scheme produces at least one value of λ that gives an absolute error less than 0.55 eV (the threshold for qualitative accuracy found in our previous benchmark).[74] This motivated the use of a variational scheme to find the best value of λ for each case.

As mentioned above, the criterion we use in DVS-tPBE is to choose the active space that gives the lowest sum of the tPBE absolute energies for the ground state and the excited state under investigation. To show the effectiveness of this approach, we compare this selection rule to three other schemes: variational selection using the summed DMRG energy, variational selection using the summed tPBE0 energy, and random selection. The right side of Figure 3 compares these approaches in choosing

https://doi.org/10.1021/acs.jctc.3c00792
J. Chem. Theory Comput. 2023, 19, 8118−8128

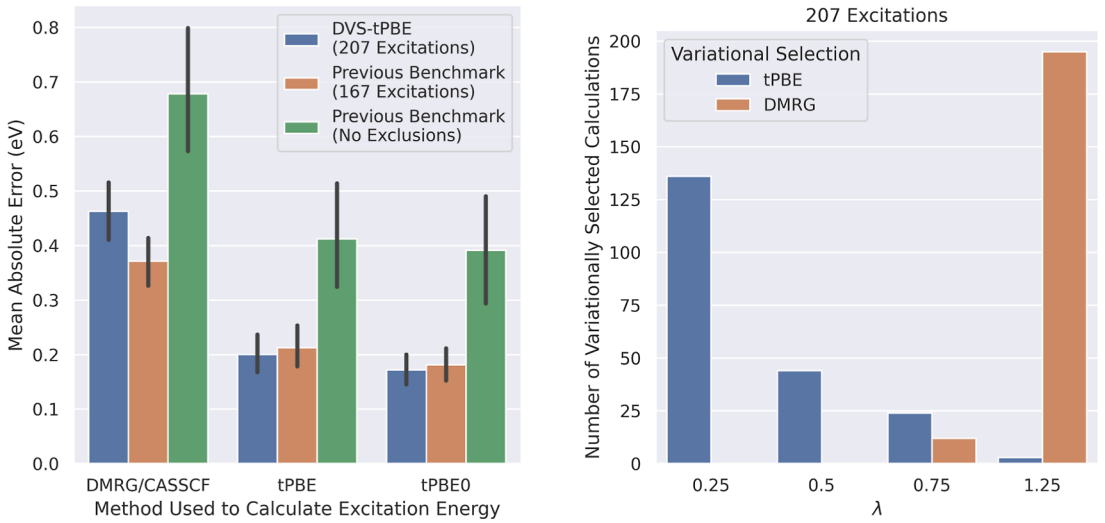


Figure 4. Left: mean absolute errors achieved by DMRG/CASSCF, tPBE, and tPBE0 on the 207-excitation test set with active spaces selected by DVS-tPBE compared to the active spaces used in our previous benchmark, both before (no exclusions) and after (167 excitations), eliminating the poor active spaces. Right: comparison of number of wave functions variationally selected with tPBE vs number of wave functions variationally selected with DMRG at each value of λ.

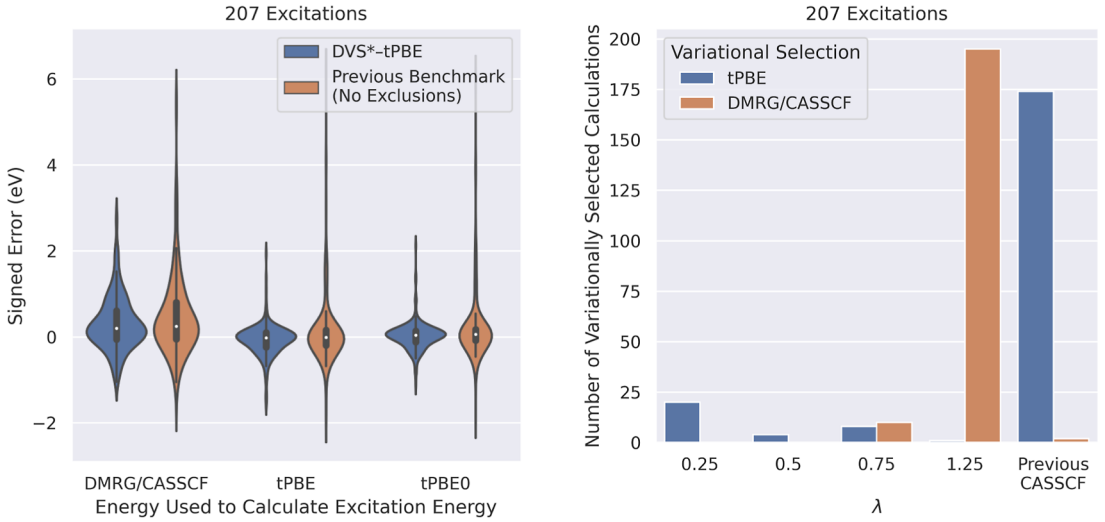


Figure 5. Left: violin plots comparing the distribution of errors for three kinds of energy calculations (DMRG/CASSCF, tPBE, and tPBE0) on the 207-excitation test set when using tPBE to select among the previous pair of SA-CASSCF wave functions and the four pairs of SA-DMRG wave functions generated in this work to the distribution of errors using just the SA-CASSCF wave functions on the same test set (not excluding poor active spaces). Right: number of wave functions variationally selected from among the five trial wave functions by using tPBE to select or DMRG/CASSCF to select. Note that the selection among five trial pairs of wave functions is labeled in the plot as DVS*-tPBE, and the previously generated SA-CASSCF wave functions are labeled as "Previous Benchmark".

among the active spaces generated with different values of λ. The figure shows that tPBE distinguishes robustly between qualitatively accurate and inaccurate complete-active-space wave functions (i.e., there are no very large errors), while DMRG does little better than random chance, which does very poorly. The mean absolute error of the tPBE0 excitation energies is 1.24 eV with DMRG used for selection as compared to 0.19 eV with tPBE used for selection. Furthermore, the maximum absolute error decreases from 3.61 eV with DMRG selection to 0.63 eV with tPBE selection. As a hybrid between tPBE and DMRG, selection with tPBE0 performs midway between these two approaches.

We also examined other ways to try to select the best active space from the trial set, but none worked as well as the tPBE selection. For example, using as the sum of orbital entropies in the active space[41] or the sum of occupation number deviations from zero or 2 are unable to select well between the different values of λ (see the Supporting Information for details of these tests).

3.2. 207-Excitation Tests. We next consider the performance of DVS-tPBE on the entire set of 207 excitations in the QUESTDB database that meet the selection criteria in Section 2.3. For this larger test, we used only four values of λ to generate active spaces: $\lambda = 0.25, 0.5, 0.75$, and 1.25. These values of λ were chosen based on their good performance on the 30-excitation tests (see the Supporting Information for more discussion of this point).

The left panel of Figure 4 shows the mean absolute errors achieved by DMRG, tPBE, and tPBE0 transition energy calculations with DVS-tPBE active-space selection for the full

https://doi.org/10.1021/acs.jctc.3c00792
J. Chem. Theory Comput. 2023, 19, 8118−8128

set of 207 excitations. These results are compared to our previous benchmark for the subset that excluded poor active spaces (those with SA-CASSCF errors greater than 1.1 eV). As can be seen, errors for all three of these methods are as good as or exceed the performance of the previous benchmark. The comparison of mean unsigned errors is as follows:

- DMRG/CASSCF: 0.46 eV presently vs 0.37 eV previously.
- tPBE: 0.20 eV presently vs 0.21 eV previously.
- tPBE0:0.17 eV presently vs 0.18 eV previously.

We note that the performance using the wave function energy (DMRG/CASSCF) is slightly worse, as might have been expected due to the bias of the previous benchmark in excluding SA-CASSCF errors larger than 1.1 eV. However, we stress that here we achieved this comparable performance without excluding any cases, whereas previously the errors were only for the better active spaces. The results of our previous benchmark without excluding poor active spaces are shown by the green bars in Figure 4; as can be seen, inclusion of these active spaces significantly diminishes the performance of the method and returns a tPBE0 mean absolute error of 0.39 eV.

The key to the success of tPBE selection compared to DMRG selection seems to be that it chooses lower values of λ. The right panel of Figure 4 shows the frequency with which each value of λ was chosen in the tPBE selection compared to that in the selection with DMRG. The figure shows that the frequency decreases quickly as a function of λ for tPBE selection. In contrast, this trend is reversed in the variational selection by DMRG, for which the selected values of λ are instead clustered heavily around $\lambda = 1.25$. The same trend toward preferring higher λ is also found in the tests on smaller 30-excitation data set where we explored λ values as high as $\lambda = 2$. In that case, we found that the selections by DMRG are clustered around $\lambda = 2$ (see the Supporting Information).

We next evaluated the usefulness of variational selection with tPBE for the problem of comparing active spaces of vastly different sizes. To do this, we use the publicly available SA-CASSCF wave functions of our previous benchmark study,[74] but here not excluding any wave functions due to poor active spaces. We then use tPBE to variationally select among five active space results: the previous SA-CASSCF results (with active spaces of about 12 active orbitals) and the four new SA-DMRG results generated with the four values of λ (large active spaces with 40 active orbitals and $M = 700$). We label this broader selection scheme as DVS*-tPBE, and the left panel of Figure 5 shows the results of this scheme compared with simply using the previous SA-CASSCF wave function results (again, not excluding any active spaces due to poor selection). Although the performance of DVS*-tPBE is slightly reduced compared to DVS-tPBE (0.20 eV tPBE0 error vs 0.17 eV), variational selection with tPBE is able to robustly discriminate against the outlier SA-CASSCF active spaces.

The right panel of Figure 5 again compares the distribution of wave functions variationally selected (among these five trial wave functions) by tPBE to those selected by DMRG/CASSCF. The figure shows that although the active-space wave functions selected by DVS*-tPBE have significantly smaller mean absolute errors (0.20 vs 0.39 eV), most of the wave functions variationally selected by tPBE come from the previous SA-CASSCF wave functions. Thus, variational selection with tPBE mainly improves the results by avoiding poor SA-CASSCF wave functions and replacing them with relatively good SA-DMRG wave functions. In contrast, variational selection with DMRG/CASSCF yields mainly wave functions generated with high values of λ and hardly any of the SA-CASSCF wave functions from our previous work.[74]

Although we have emphasized the excitation energies calculated by tPBE0, examination of the above results shows that tPBE excitation energies are, on average, only slightly worse. Another conclusion that can be drawn from the above comparisons is that tPBE and tPBE0 excitation energy calculations are not overly sensitive to the nature of the multiconfigurational wave functions. We obtain good results both with the selection among four active spaces for SA-DMRG calculations and with the selection among five trial active spaces, although in the latter case, a DMRG active space is not usually the one chosen. Therefore, for the great majority of the excitations, we get good results with MC-PDFT and HMC-PDFT with quite different kinds of multiconfigurational wave functions.

Finally, we performed some tests to evaluate the sensitivity of DVS-tPBE to the number of active orbitals chosen. Keeping the bond dimension (700) and number of active spaces (4) fixed, we find that increasing the number of orbitals to 40 is the point at which our tPBE results start to replicate the accuracy of our previous study;[74] selecting 30 orbitals significantly decreases performance (Supporting Information). Thus, the success of the approach in avoiding the expensive step of orbital optimization is largely enabled by the large active spaces afforded by SA-DMRG.

4. CONCLUDING REMARKS

The goal of this work was to develop an automated framework for selecting active spaces for calculating vertical excitation energies with a useful predictive accuracy. Toward this goal, we have presented the DVS approach to active space selection in which one generates multiple trial wave functions from a set of constructed active spaces and employs a variational selection scheme to choose the final result. To practically implement this approach for vertical excitation energies in the QUESTDB database, we have presented a scheme in which, for each excitation, (i) an RHF or ROHF wave function is calculated for the ground state, (ii) different sets of candidate orbitals are generated by diagonalization of a parametrized operator, (iii) 40-orbital trial active spaces are chosen from these sets using APC selection, (iv) ground and excited-state SA-DMRG wave functions are calculated for each of these active spaces with a bond dimension of M = 700, and (v) the final result is chosen from among the resulting wave functions as the one that gives the lowest sum of absolute energies of the ground state and the excited state under consideration. We find that this approach performs poorly when using the DMRG/CASSCF absolute energies to select between wave functions but robustly when using the absolute energy given by the translated tPBE functional of MC-PDFT (DVS-tPBE).

We have tested this method on 207 vertical excitations in the QUESTDB data set (199 excitations from singlet states and eight from doublet states). When choosing between only four trial active spaces with no further orbital optimization, we are able to obtain equally as accurate tPBE0 results as in our previous benchmark[74] but now for all systems without the need to filter out poor active spaces. The success of this approach in avoiding the costly step of orbital optimization is largely enabled by the large active spaces afforded by SA-DMRG, and it is consistent with the recent perspective that "CASCI is not merely

https://doi.org/10.1021/acs.jctc.3c00792
J. Chem. Theory Comput. 2023, 19, 8118−8128

an approximation to CASSCF, in that it can be designed to have important qualitative advantages over CASSCF."[69] While the results in the article proper show that this approach is successful for systems in QUESTDB,[83] we show in the Supporting Information that it can also have success in the transition metal system MnO_4^- with only minor modification (using larger N in the calculation of the APC entropies).[107]

Of course, application to different systems may require a greater number of orbitals and larger bond dimension or a different approach entirely to constructing the candidate active space wave functions. Toward this end, we have shown that DVS-tPBE remains effective even when choosing between the large SA-DMRG active spaces of this work and the smaller SA-CASSCF active spaces of our previous benchmark.[74] That is, if we enlarge the trial set of active spaces to include both those from the SA-CASSCF calculations with small active spaces and the new large active spaces (with 40 active orbitals) and choose among them with variational selection by tPBE, we again obtain good results, even though we are now comparing quite different kinds of wave functions. These results show that DVS-tPBE can choose robustly between active spaces of vastly different sizes. This flexibility provides the basis for the further development of DVS-tPBE for applications of more metal-containing systems, extended organic systems, and adiabatic excitations.

In summary, we have proposed an approach for automatically selecting between active spaces for vertical excitations variationally through use of the tPBE energy from MC-PDFT. We have practically implemented this approach for the QUESTDB database through use of a parametrized operator to generate different active spaces and large SA-DMRG wave functions. Our results show that such an approach can potentially enable the application of CAS-based approaches in a high-throughput and predictive fashion. Although one cannot guarantee that any single active-space selection method will always work well, DVS with tPBE (DVS-tBPE) appears robust.

Converged density matrices of all 40-orbital DMRG calculations for the singlet and triplet QUESTDB excitations are available on Zenodo.[108] The code used for the APC active space selection is now available in PySCF.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.3c00792.

Analysis of the 2^1A_g state of ethylene with varying λ; additional analysis of orbital trends with varying λ; performance of DVS-tPBE selecting from 30-orbital active spaces; performance of variational selection with methods besides tPBE; and for the singlet and triplet excitations studied from QUESTDB, the name of the molecule, the total number of atoms, the number of nonhydrogenic atoms, the transition, the M diagnostic[51] of the initial and final states, and the reference and calculated excitation energy (PDF)

Supporting_data (ZIP)

■ AUTHOR INFORMATION

Corresponding Authors

Donald G. Truhlar − *Department of Chemistry, Chemical Theory Group, and Supercomputing Institute, University of Minnesota, Minneapolis, Minnesota 55455, United States;* orcid.org/0000-0002-7742-7294; Email: truhlar@umn.edu

Laura Gagliardi − *Department of Chemistry, Pritzker School of Molecular Engineering, James Franck Institute, Chicago Center for Theoretical Chemistry, University of Chicago, Chicago, Illinois 60637, United States;* orcid.org/0000-0001-5227-1396; Email: lgagliardi@uchicago.edu

Author

Daniel S. King − *Department of Chemistry, University of Chicago, Chicago, Illinois 60637, United States;* orcid.org/0000-0003-0208-5274

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.3c00792

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Matthew Hennefarth and Matthew Hermes for helpful discussions. This work is supported by the National Science Foundation under grant CHE-2054723. We thank the Research Computing Center (RCC) at the University of Chicago for computational resources.

■ REFERENCES

(1) Piecuch, P.; Kowalski, K.; Pimienta, I. S. O.; Mcguire, M. J. Recent Advances in Electronic Structure Theory: Method of Moments of Coupled-Cluster Equations and Renormalized Coupled-Cluster Approaches. *Int. Rev. Phys. Chem.* **2002**, *21*, 527−655.

(2) Krylov, A. I. Spin-Flip Equation-Of-Motion Coupled-Cluster Electronic Structure Method for a Description of Excited States, Bond Breaking, Diradicals, and Triradicals. *Acc. Chem. Res.* **2006**, *39*, 83−91.

(3) González, L.; Escudero, D.; Serrano-Andrés, L. Progress and Challenges in the Calculation of Electronic Excited States. *ChemPhysChem* **2012**, *13*, 28−51.

(4) Sneskov, K.; Christiansen, O. Excited State Coupled Cluster Methods: Excited State Coupled Cluster Methods. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 566−584.

(5) Adamo, C.; Jacquemin, D. The Calculations of Excited-State Properties With Time-Dependent Density Functional Theory. *Chem. Soc. Rev.* **2013**, *42*, 845−856.

(6) Laurent, A. D.; Jacquemin, D. TD-DFT Benchmarks: A Review. *Int. J. Quantum Chem.* **2013**, *113*, 2019−2039.

(7) Faber, C.; Boulanger, P.; Attaccalite, C.; Duchemin, I.; Blase, X. Excited States Properties of Organic Molecules: From Density Functional Theory to the GW and Bethe-Salpeter Green's Function Formalisms. *Philos. Trans. R. Soc., A* **2014**, *372*, 20130271.

(8) Lischka, H.; Nachtigallová, D.; Aquino, A.; Szalay, P. G.; Plasser, F.; Machado, F. B. C.; Barbatti, M. Multireference Approaches for Excited States of Molecules. *Chem. Rev.* **2018**, *118*, 7293−7361.

(9) Ghosh, S.; Verma, P.; Cramer, C. J.; Gagliardi, L.; Truhlar, D. G. Combining Wave Function Methods With Density Functional Theory for Excited States. *Chem. Rev.* **2018**, *118*, 7249−7292.

(10) Roemelt, M.; Krewald, V.; Pantazis, D. A. Exchange Coupling Interactions From the Density Matrix Renormalization Group and N-Electron Valence Perturbation Theory: Application to a Biomimetic Mixed-Valence Manganese Complex. *J. Chem. Theory Comput.* **2018**, *14*, 166−179.

(11) Levine, B. G.; Esch, M. P.; Fales, B. S.; Hardwick, D. T.; Peng, W.-T.; Shu, Y. Conical Intersections at the Nanoscale: Molecular Ideas for Materials. *Annu. Rev. Phys. Chem.* **2019**, *70*, 21−43.

(12) Blase, X.; Duchemin, I.; Jacquemin, D.; Loos, P.-F. The Bethe−Salpeter Equation Formalism: From Physics to Chemistry. *J. Phys. Chem. Lett.* **2020**, *11*, 7371−7382.

(13) Izsák, R. Single-Reference Coupled Cluster Methods for Computing Excitation Energies in Large Molecules: The Efficiency and Accuracy of Approximations. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2020**, *10*, No. e1445.
(14) Westermayr, J.; Marquetand, P. Machine Learning for Electronically Excited States of Molecules. *Chem. Rev.* **2021**, *121*, 9873−9926.
(15) Eriksen, J. J. The Shape of Full Configuration Interaction to Come. *J. Phys. Chem. Lett.* **2021**, *12*, 418−432.
(16) Dral, P.; Barbatti, M. Molecular Excited States Through a Machine Learning Lens. *Nature Rev. Chem.* **2021**, *5*, 388−405.
(17) Zhou, C.; Hermes, M.; Wu, D.; Bao, J. J.; Pandharkar, R.; King, D. S.; Zhang, D.; Scott, T.; Lykhin, A.; Gagliardi, L.; Truhlar, D. G. Electronic Structure of Strongly Correlated Systems: Recent Developments in Multiconfiguration Pair-Density Functional Theory and Multiconfiguration Nonclassical-Energy Functional Theory. *Chem. Sci.* **2022**, *13*, 7685−7706.
(18) Casanova, D. Restricted Active Space Configuration Interaction Methods for Strong Correlation: Recent Developments. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2022**, *12*, No. e1561.
(19) Shu, Y.; Truhlar, D. G. Doubly Excited Character or Static Correlation of the Reference State in the Controversial 2^1A_g State of Trans-Butadiene? *J. Am. Chem. Soc.* **2017**, *139*, 13770−13778.
(20) Loos, P.-F.; Boggio-Pasqua, M.; Scemama, A.; Caffarel, M.; Jacquemin, D. Reference Energies for Double Excitations. *J. Chem. Theory Comput.* **2019**, *15*, 1939−1956.
(21) Vancoillie, S.; Zhao, H.; Tran, V. T.; Hendrickx, M. F. A.; Pierloot, K. Multiconfigurational Second-Order Perturbation Theory Restricted Active Space (RASPT2) Studies on Mononuclear First-Row Transition-Metal Systems. *J. Chem. Theory Comput.* **2011**, *7*, 3961−3977.
(22) Jacquemin, D.; Zhao, Y.; Valero, R.; Adamo, C.; Ciofini, I.; Truhlar, D. G. Verdict: Time-Dependent Density Functional Theory "Not Guilty" of Large Errors for Cyanines. *J. Chem. Theory Comput.* **2012**, *8*, 1255−1259.
(23) Zhou, C.; Gagliardi, L.; Truhlar, D. G. Multiconfiguration Pair-Density Functional Theory for Iron Porphyrin With CAS, RAS, and DMRG Active Spaces. *J. Phys. Chem. A* **2019**, *123*, 3389−3394.
(24) Blunt, N. S.; Mahajan, A.; Sharma, S. Efficient Multireference Perturbation Theory Without High-Order Reduced Density Matrices. *J. Chem. Phys.* **2020**, *153*, 164120.
(25) Beran, P.; Matoušek, M.; Hapka, M.; Pernal, K.; Veis, L. Density Matrix Renormalization Group With Dynamical Correlation via Adiabatic Connection. *J. Chem. Theory Comput.* **2021**, *17*, 7575−7585.
(26) Gaggioli, C. A.; Stoneburner, S. J.; Cramer, C. J.; Gagliardi, L. Beyond Density Functional Theory: The Multiconfigurational Approach To Model Heterogeneous Catalysis. *ACS Catal.* **2019**, *9*, 8481−8502.
(27) Eriksen, J. J. The Shape of Full Configuration Interaction to Come. *J. Phys. Chem. Lett.* **2021**, *12*, 418−432.
(28) White, S. R. Density Matrix Formulation for Quantum Renormalization Groups. *Phys. Rev. Lett.* **1992**, *69*, 2863−2866.
(29) White, S. R. Density-Matrix Algorithms for Quantum Renormalization Groups. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1993**, *48*, 10345−10356.
(30) White, S. R.; Martin, R. L. Ab Initio Quantum Chemistry Using the Density Matrix Renormalization Group. *J. Chem. Phys.* **1999**, *110*, 4127−4130.
(31) Chan, G. K.-L.; Sharma, S. The Density Matrix Renormalization Group in Quantum Chemistry. *Annu. Rev. Phys. Chem.* **2011**, *62*, 465−481.
(32) Schollwöck, U. The Density-Matrix Renormalization Group in the Age of Matrix Product States. *Ann. Phys.* **2011**, *326*, 96−192.
(33) Marti, K. H.; Reiher, M. New Electron Correlation Theories for Transition Metal Chemistry. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6750−6759.
(34) Wouters, S.; Van Neck, D. The Density Matrix Renormalization Group for Ab Initio Quantum Chemistry. *Eur. Phys. J. D* **2014**, *68*, 272.
(35) Olivares-Amaya, R.; Hu, W.; Nakatani, N.; Sharma, S.; Yang, J.; Chan, G. K.-L. The Ab-Initio Density Matrix Renormalization Group in Practice. *J. Chem. Phys.* **2015**, *142*, 034102.
(36) Szalay, S.; Pfeffer, M.; Murg, V.; Barcza, G.; Verstraete, F.; Schneider, R.; Legeza, Ö. Tensor Product Methods and Entanglement Optimization for Ab Initio Quantum Chemistry. *Int. J. Quantum Chem.* **2015**, *115*, 1342−1391.
(37) Yanai, T.; Kurashige, Y.; Mizukami, W.; Chalupský, J.; Lan, T. N.; Saitow, M. Density Matrix Renormalization Group for Ab Initio Calculations and Associated Dynamic Correlation Methods: A Review of Theory and Applications. *Int. J. Quantum Chem.* **2015**, *115*, 283−299.
(38) Chan, G. K.-L.; Keselman, A.; Nakatani, N.; Li, Z.; White, S. R. Matrix Product Operators, Matrix Product States, and Ab Initio Density Matrix Renormalization Group Algorithms. *J. Chem. Phys.* **2016**, *145*, 014102.
(39) Knecht, S.; Hedegård, E. D.; Keller, S.; Kovyrshin, A.; Ma, Y.; Muolo, A.; Stein, C. J.; Reiher, M. New Approaches for Ab Initio Calculations of Molecules With Strong Electron Correlation. *Chimia* **2016**, *70*, 244−251.
(40) Baiardi, A.; Reiher, M. The Density Matrix Renormalization Group in Chemistry and Molecular Physics: Recent Developments and New Challenges. *J. Chem. Phys.* **2020**, *152*, 040903.
(41) Stein, C. J.; Reiher, M. autoCAS: A Program for Fully Automated Multiconfigurational Calculations. *J. Comput. Chem.* **2019**, *40*, 2216−2226.
(42) Roos, B. O.; Taylor, P. R.; Sigbahn, P. E. A Complete Active Space SCF Method (CASSCF) Using a Density Matrix Formulated Super-Ci Approach. *Chem. Phys.* **1980**, *48*, 157−173.
(43) Zgid, D.; Nooijen, M. The Density Matrix Renormalization Group Self-Consistent Field Method: Orbital Optimization with the Density Matrix Renormalization Group Method in the Active Space. *J. Chem. Phys.* **2008**, *128*, 144116.
(44) Ghosh, D.; Hachmann, J.; Yanai, T.; Chan, G. K.-L. Orbital Optimization in the Density Matrix Renormalization Group, with Applications to Polyenes and β-Carotene. *J. Chem. Phys.* **2008**, *128*, 144117.
(45) Ruedenberg, K.; Cheung, L. M.; Elbert, S. T. MCSCF Optimization through Combined Use of Natural Orbitals and the Brillouin-Levy-Berthier Theorem. *Int. J. Quantum Chem.* **1979**, *16*, 1069−1101.
(46) Freitag, L.; Ma, Y.; Baiardi, A.; Knecht, S.; Reiher, M. Approximate Analytical Gradients and Nonadiabatic Couplings for the State-Average Density Matrix Renormalization Group Self-Consistent-Field Method. *J. Chem. Theory Comput.* **2019**, *15*, 6724−6737.
(47) Roos, B. O. Perspectives in Calculations on Excited State in Molecular Systems. In *Computational Photochemisty*; Olivucci, M., Ed.; Elsevier: Amsterdam, 2005; pp 317−348.
(48) Pulay, P.; Hamilton, T. P. UHF Natural Orbitals for Defining and Starting MC-SCF Calculations. *J. Chem. Phys.* **1988**, *88*, 4926−4933.
(49) Bofill, J. M.; Pulay, P. The Unrestricted Natural Orbital−complete Active Space (UNO−CAS) Method: An Inexpensive Alternative to the Complete Active Space−self-consistent-field (CAS−SCF) Method. *J. Chem. Phys.* **1989**, *90*, 3637−3646.
(50) Schmidt, M. W.; Gordon, M. S. The Construction and Interpretation of Mcscf Wavefunctions. *Annu. Rev. Phys. Chem.* **1998**, *49*, 233−266.
(51) Tishchenko, O.; Zheng, J.; Truhlar, D. G. Multireference Model Chemistries for Thermochemical Kinetics. *J. Chem. Theory Comput.* **2008**, *4*, 1208−1219.
(52) Veryazov, V.; Malmqvist, P. Å.; Roos, B. O. How to Select Active Space for Multiconfigurational Quantum Chemistry? *Int. J. Quantum Chem.* **2011**, *111*, 3329−3338.
(53) Keller, S.; Boguslawski, K.; Janowski, T.; Reiher, M.; Pulay, P. Selection of Active Spaces for Multiconfigurational Wavefunctions. *J. Chem. Phys.* **2015**, *142*, 244104.
(54) Bao, J. L.; Sand, A.; Gagliardi, L.; Truhlar, D. G. Correlated-Participating-Orbitals Pair-Density Functional Method and Applica-

https://doi.org/10.1021/acs.jctc.3c00792
J. Chem. Theory Comput. 2023, 19, 8118−8128

tion to Multiplet Energy Splittings of Main-Group Divalent Radicals. *J. Chem. Theory Comput.* **2016**, *12*, 4274−4283.
(55) Stein, C. J.; Reiher, M. Automated Selection of Active Orbital Spaces. *J. Chem. Theory Comput.* **2016**, *12*, 1760−1771.
(56) Bao, J. L.; Odoh, S. O.; Gagliardi, L.; Truhlar, D. G. 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.* **2017**, *13*, 616−626.
(57) Stein, C. J.; Reiher, M. Measuring Multi-Configurational Character by Orbital Entanglement. *Mol. Phys.* **2017**, *115*, 2110−2119.
(58) Sayfutyarova, E. R.; Sun, Q.; Chan, G. K.-L.; Knizia, G. Automated Construction of Molecular Active Spaces From Atomic Valence Orbitals. *J. Chem. Theory Comput.* **2017**, *13*, 4063−4078.
(59) Bao, J. J.; Dong, S. S.; Gagliardi, L.; Truhlar, D. G. Automatic Selection of an Active Space for Calculating Electronic Excitation Spectra by MS-CASPT2 or MC-PDFT. *J. Chem. Theory Comput.* **2018**, *14*, 2017−2025.
(60) Bao, J. J.; Truhlar, D. G. Automatic Active Space Selection for Calculating Electronic Excitation Energies Based on High-Spin Unrestricted Hartree−Fock Orbitals. *J. Chem. Theory Comput.* **2019**, *15*, 5308−5318.
(61) Khedkar, A.; Roemelt, M. Active Space Selection Based on Natural Orbital Occupation Numbers From N-Electron Valence Perturbation Theory. *J. Chem. Theory Comput.* **2019**, *15*, 3522−3536.
(62) Sayfutyarova, E. R.; Hammes-Schiffer, S. Constructing Molecular Π-Orbital Active Spaces for Multireference Calculations of Conjugated Systems. *J. Chem. Theory Comput.* **2019**, *15*, 1679−1689.
(63) Jeong, W.; Stoneburner, S. J.; King, D.; Li, R.; Walker, A.; Lindh, R.; Gagliardi, L. Automation of Active Space Selection for Multi-reference Methods via Machine Learning on Chemical Bond Dissociation. *J. Chem. Theory Comput.* **2020**, *16*, 2389−2399.
(64) Li, S. J.; Gagliardi, L.; Truhlar, D. G. Extended Separated-Pair Approximation for Transition Metal Potential Energy Curves. *J. Chem. Phys.* **2020**, *152*, 124118.
(65) Gieseking, R. L. M. A new release of MOPAC incorporating the INDO/S semiempirical model with CI excited states. *J. Comput. Chem.* **2021**, *42*, 365−378.
(66) Golub, P.; Antalik, A.; Veis, L.; Brabec, J. Machine Learning-Assisted Selection of Active Spaces for Strongly Correlated Transition Metal Systems. *J. Chem. Theory Comput.* **2021**, *17*, 6053−6072.
(67) Khedkar, A.; Roemelt, M. Modern Multireference Methods and Their Application in Transition Metal Chemistry. *Phys. Chem. Chem. Phys.* **2021**, *23*, 17097−17112.
(68) Lei, Y.; Suo, B.; Liu, W. iCAS: Imposed Automatic Selection and Localization of Complete Active Spaces. *J. Chem. Theory Comput.* **2021**, *17*, 4846−4859.
(69) Levine, B. G.; Durden, A. S.; Esch, M. P.; Liang, F.; Shu, Y. CAS Without SCF-Why to Use CASCI and Where to Get the Orbitals. *J. Chem. Phys.* **2021**, *154*, 090902.
(70) Oakley, M. S.; Gagliardi, L.; Truhlar, D. G. Multiconfiguration Pair-Density Functional Theory for Transition Metal Silicide Bond Dissociation Energies, Bond Lengths, and State Orderings. *Molecules* **2021**, *26*, 2881.
(71) Weser, O.; Guther, K.; Ghanem, K.; Li Manni, G. Stochastic Generalized Active Space Self-Consistent Field: Theory and Application. *J. Chem. Theory Comput.* **2022**, *18*, 251−272.
(72) King, D. S.; Gagliardi, L. A Ranked-Orbital Approach to Select Active Spaces for High-Throughput Multireference Computation. *J. Chem. Theory Comput.* **2021**, *17*, 2817−2831.
(73) Cheng, Y.; Xie, Z.; Ma, H. Post-Density Matrix Renormalization Group Methods for Describing Dynamic Electron Correlation With Large Active Spaces. *J. Phys. Chem. Lett.* **2022**, *13*, 904−915.
(74) King, D. S.; Hermes, M. R.; Truhlar, D. G.; Gagliardi, L. Large-Scale Benchmarking of Multireference Vertical-Excitation Calculations via Automated Active-Space Selection. *J. Chem. Theory Comput.* **2022**, *18*, 6065−6076.
(75) Bensberg, M.; Reiher, M. Corresponding Active Orbital Spaces Along Chemical Reaction Paths. *J. Phys. Chem. Lett.* **2023**, *14*, 2112−2118.
(76) Golub, P.; Antalik, A.; Beran, P.; Brabec, J. Mutual Information Prediction for Strongly Correlated Systems. *Chem. Phys. Lett.* **2023**, *813*, 140297.
(77) Kaufold, B. W.; Chintala, N.; Pandeya, P.; Dong, S. S. Automated Active Space Selection With Dipole Moments. *J. Chem. Theory Comput.* **2023**, *19*, 2469−2483.
(78) Serwatka, T.; Roy, P.-N. Optimized Basis Sets for DMRG Calculations of Quantum Chains of Rotating Water Molecules. *J. Chem. Phys.* **2023**, *158*, 214103.
(79) Tóth, Z.; Pulay, P. Finding Symmetry Breaking Hartree-Fock Solutions: The Case of Triplet Instability. *J. Chem. Phys.* **2016**, *145*, 164102.
(80) Tóth, Z.; Pulay, P. Comparison of Methods for Active Orbital Selection in Multiconfigurational Calculations. *J. Chem. Theory Comput.* **2020**, *16*, 7328.
(81) Stein, C. J.; von Burg, V.; Reiher, M. The Delicate Balance of Static and Dynamic Electron Correlation. *J. Chem. Theory Comput.* **2016**, *12*, 3764−3773.
(82) Stein, C.; Reiher, M. Automated Identification of Relevant Frontier Orbitals for Chemical Compounds and Processes. *Chimia* **2017**, *71*, 170−176.
(83) Véril, M.; Scemama, A.; Caffarel, M.; Lipparini, F.; Boggio-Pasqua, M.; Jacquemin, D.; Loos, P. QUESTDB: A Database of Highly Accurate Excitation Energies for the Electronic Structure Community. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2021**, *11*, No. e1517.
(84) Dyall, K. G. The Choice of a Zeroth-order Hamiltonian for Second-order Perturbation Theory With a Complete Active Space Self-consistent-field Reference Function. *J. Chem. Phys.* **1995**, *102*, 4909−4918.
(85) Angeli, C.; Cimiraglia, R.; Evangelisti, S.; Leininger, T.; Malrieu, J.-P. Introduction of n -Electron Valence States for Multireference Perturbation Theory. *J. Chem. Phys.* **2001**, *114*, 10252−10264.
(86) Li Manni, G.; Carlson, R. K.; Luo, S.; Ma, D.; Olsen, J.; Truhlar, D. G.; Gagliardi, L. Multiconfiguration Pair-Density Functional Theory. *J. Chem. Theory Comput.* **2014**, *10*, 3669−3680.
(87) Schapiro, I.; Sivalingam, K.; Neese, F. Assessment of n-Electron Valence State Perturbation Theory for Vertical Excitation Energies. *J. Chem. Theory Comput.* **2013**, *9*, 3567−3580.
(88) Hoyer, C. E.; Ghosh, S.; Truhlar, D. G.; Gagliardi, L. Multiconfiguration Pair-Density Functional Theory Is as Accurate as CASPT2 for Electronic Excitation. *J. Phys. Chem. Lett.* **2016**, *7*, 586−591.
(89) Sarkar, R.; Loos, P.-F.; Boggio-Pasqua, M.; Jacquemin, D. Assessing the Performances of CASPT2 and NEVPT2 for Vertical Excitation Energies. *J. Chem. Theory Comput.* **2022**, *18*, 2418−2436.
(90) Pandharkar, R.; Hermes, M. R.; Truhlar, D. G.; Gagliardi, L. A New Mixing of Nonlocal Exchange and Nonlocal Correlation With Multiconfiguration Pair-Density Functional Theory. *J. Phys. Chem. Lett.* **2020**, *11*, 10158−10163.
(91) King, D. S.; Truhlar, D. G.; Gagliardi, L. Machine-Learned Energy Functionals for Multiconfigurational Wave Functions. *J. Phys. Chem. Lett.* **2021**, *12*, 7761−7767.
(92) Zhang, D.; Hermes, M. R.; Gagliardi, L.; Truhlar, D. G. Multiconfiguration Density-Coherence Functional Theory. *J. Chem. Theory Comput.* **2021**, *17*, 2775−2782.
(93) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(94) Zhai, H.; Chan, G. K.-L. Low Communication High Performance Ab Initio Density Matrix Renormalization Group Algorithms. *J. Chem. Phys.* **2021**, *154*, 224116.
(95) Roothaan, C. New Developments in Molecular Orbital Theory. *Rev. Mod. Phys.* **1951**, *23*, 69−89.
(96) Roothaan, C. C. J. Self-Consistent Field Theory for Open Shells of Electronic Systems. *Rev. Mod. Phys.* **1960**, *32*, 179−185.

https://doi.org/10.1021/acs.jctc.3c00792
J. Chem. Theory Comput. 2023, 19, 8118−8128

(97) Kendall, R. A.; Dunning, T. H.; Harrison, R. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96*, 6796−6806.
(98) Woon, D. E.; Dunning, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. III. The Atoms Aluminum Through Argon. *J. Chem. Phys.* **1993**, *98*, 1358−1371.
(99) Papajak, E.; Zheng, J.; Xu, X.; Leverentz, H. R.; Truhlar, D. G. Perspectives on Basis Sets Beautiful: Seasonal Plantings of Diffuse Basis Functions. *J. Chem. Theory Comput.* **2011**, *7*, 3027−3034.
(100) Tsuchimochi, T.; Scuseria, G. E. Communication: ROHF Theory Made Simple. *J. Chem. Phys.* **2010**, *133*, 141102.
(101) Sun, Q.; Berkelbach, T. C.; Blunt, N. S.; Booth, G. H.; Guo, S.; Li, Z.; Liu, J.; McClain, J. D.; Sayfutyarova, E. R.; Sharma, S.; Wouters, S.; Chan, G. K.-L. PySCF: The Python-Based Simulations of Chemistry Framework. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2018**, *8*, No. e1340.
(102) Sun, Q.; Zhang, X.; Banerjee, S.; Bao, P.; Barbry, M.; Blunt, N. S.; Bogdanov, N. A.; Booth, G. H.; Chen, J.; Cui, Z.-H.; Eriksen, J. J.; Gao, Y.; Guo, S.; Hermann, J.; Hermes, M. R.; Koh, K.; Koval, P.; Lehtola, S.; Li, Z.; Liu, J.; Mardirossian, N.; McClain, J. D.; Motta, M.; Mussard, B.; Pham, H. Q.; Pulkin, A.; Purwanto, W.; Robinson, P. J.; Ronca, E.; Sayfutyarova, E. R.; Scheurer, M.; Schurkus, H. F.; Smith, J. E. T.; Sun, C.; Sun, S.-N.; Upadhyay, S.; Wagner, L. K.; Wang, X.; White, A.; Whitfield, J. D.; Williamson, M. J.; Wouters, S.; Yang, J.; Yu, J. M.; Zhu, T.; Berkelbach, T. C.; Sharma, S.; Sokolov, A. Y.; Chan, G. K.-L. Recent Developments in the PySCF Program Package. *J. Chem. Phys.* **2020**, *153*, 024109.
(103) GitHub. *MRH*. https://github.com/MatthewRHermes/mrh, (accessed Feb 1, 2023).
(104) GitHub. *Pyscf-Forge*. https://github.com/pyscf/pyscf-forge, (accessed Sept 15, 2023).
(105) Barbatti, M.; Paier, J.; Lischka, H. Photochemistry of Ethylene: A Multireference Configuration Interaction Investigation of the Excited-State Energy Surfaces. *J. Chem. Phys.* **2004**, *121*, 11614−11624.
(106) Feller, D.; Peterson, K. A.; Davidson, E. R. A Systematic Approach to Vertically Excited States of Ethylene Using Configuration Interaction and Coupled Cluster Techniques. *J. Chem. Phys.* **2014**, *141*, 104302.
(107) Sharma, P.; Truhlar, D. G.; Gagliardi, L. Multiconfiguration Pair-Density Functional Theory Investigation of the Electronic Spectrum of MnO_4^-. *J. Chem. Phys.* **2018**, *148*, 124305.
(108) King, D. S.; Truhlar, D. G.; Gagliardi, L. Variational Active Space Selection with Multiconfiguration Pair-Density Functional Theory. **2023**; [Online; accessed 30-October-2023].

https://doi.org/10.1021/acs.jctc.3c00792
J. Chem. Theory Comput. 2023, 19, 8118−8128