

Automated Identification of Relevant Frontier Orbitals for Chemical Compounds and Processes

Christopher J. Stein[§] and Markus Reiher*

[§]SCS-Metrohm Award for best oral presentation in Computational Chemistry

Abstract: Quantum-chemical multi-configurational methods are required for a proper description of static electron correlation, a phenomenon inherent to the electronic structure of molecules with multiple (near-)degenerate frontier orbitals. Here, we review how a property of these frontier orbitals, namely the entanglement entropy is related to static electron correlation. A subset of orbitals, the so-called active orbital space is an essential ingredient for all multi-configurational methods. We proposed an automated selection of this active orbital space, that would otherwise be a tedious and error prone manual procedure, based on entanglement measures. Here, we extend this scheme to demonstrate its capability for the selection of consistent active spaces for several excited states and along reaction coordinates.

Keywords: Chemical reactions · Orbital entanglement · Quantum chemistry · Strong correlation



Christopher J. Stein studied chemistry at the Georg-August-Universität Göttingen from 2009 to 2014 and was awarded the Otto-Wallach-Prize for his Bachelor's degree and the Gustav-Tammann-Prize for his Master's degree. Already during his studies, he joined the group of the late Prof. Peter Botschwina, where he studied the spectroscopy of astrochemical molecules with highly accurate theoretical calculations. In 2015, he joined the Theoretical Chemistry Group of Prof. Markus Reiher at ETH Zurich to pursue his PhD supported by a Kekulé fellowship of the Fonds der Chemischen Industrie. The focus of his research is on the development of new methods and tools to describe strong electron correlation.

1. Introduction

Quantum chemistry is primarily concerned with the calculation of electronic energies of molecules and materials, including their properties which can be expressed as derivatives of the energy. Such an energy is calculated from the electronic wave function or the electron density. A plethora of methods has been developed that allows chemists to determine the electronic wave function, the electron density, and the corresponding electronic energy. These methods typically scale unfavorably with system size, so that approaches which yield remarkably accurate results for small molecules^[1–3] are not feasible for large molecules. It is this trade-off between accuracy and feasibility that makes the selection of an appropriate method a central task in every quantum chemical investigation and stimulates the development of new methods. In addition to the size of a molecule, other factors make an electronic structure calculation a potentially difficult task.

One of these potential difficulties is known as *static electron correlation*, emerging from dense-lying frontier orbitals. It can be illustrated at the simple example of the oxygen molecule in its lowest singlet spin state. The electronic wave function is usually expanded in antisymmetrized products of molecular orbitals or symmetry-adapted linear combinations thereof. In Fig. 1, we show the frontier molecular orbitals of dioxygen. Six of the eight electrons occupy the bonding σ - and π -orbitals according to the Aufbau principle.

The remaining electron pair must then occupy one of the degenerate antibonding π^* -orbitals, but none can be preferred over the other. Any possible distribution of the electrons over the orbitals is referred to as an electronic configuration. For singlet dioxygen in its ${}^1\Delta_g$ state there exist two electronic configurations of which one cannot be preferred over the other. Note that the electronic configurations where two electrons of opposite spin singly occupy the degenerate π^* -orbitals give rise to a ${}^1\Sigma_g^+$ state of higher energy.

To describe the electronic structure of the ${}^1\Delta_g$ state qualitatively correct, both configurations have to be included in a qualitatively correct approximation of the wave function. The necessity to include more than one configuration represents the static-correlation problem and is encountered when molecular frontier orbitals are degenerate or near-degenerate. It is therefore prototypical for the electronic structure of many transition-metal complexes and clusters, molecules with extended π -systems, and for chemical bonds that are broken or formed.

Many of the commonly applied quantum-chemical methods such as Kohn-Sham density functional theory (DFT)^[4] and coupled-cluster methods^[5] are based on reference wave functions comprising a single configuration only. This deficiency of the reference wave function can often be corrected for multi-configurational systems. Where appropriate, high-excitation coupled-cluster expansions or even broken-spin-symmetry approaches may work. However, the only natural way to deal with

*Correspondence: Prof. Dr. M. Reiher
ETH Zurich
Laboratorium für Physikalische Chemie
Vladimir-Prelog-Weg 2, CH-8093 Zurich
E-mail: markus.reiher@phys.chem.ethz.ch

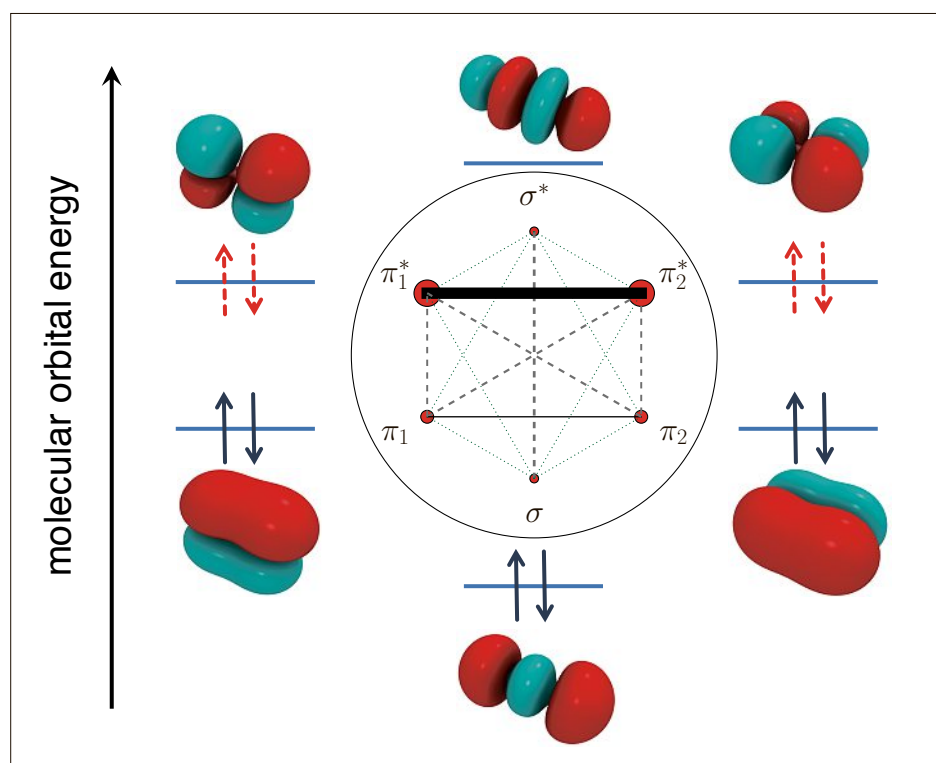


Fig. 1. Part of a molecular orbital diagram for the frontier orbitals of the oxygen molecule in its lowest singlet state (${}^1\Delta_g$) consisting of eight electrons in six orbitals. The dark blue arrows denote electrons that are distributed over the orbitals according to the Aufbau principle. The fact that the electron pair is equally likely to be in either of the π^* -orbitals is indicated by the red dashed arrows. The corresponding entanglement diagram (see text) is included in the center. Here, every orbital is assigned a red circle whose area is proportional to its single-orbital entropy. The connecting lines mirror the value of the mutual information with the green dashed lines indicating a low, the gray dashed lines an intermediate, and the black solid lines a high value of the mutual information.

the static correlation problem is to apply an intrinsically multi-configurational method.

The *de facto* standard method for the calculation of strongly correlated molecules in quantum chemistry is the complete active space self-consistent field (CASSCF) method.^[6–9] It implements the concept of a fully optimized reaction space^[10–13] and includes all configurations that can arise from a selected subset of so-called *active* orbitals L and electrons N , while the occupations of the low-energy *inactive* orbitals and the virtual orbitals are fixed to either doubly occupied or unoccupied, respectively. The *active space*, often denoted as CAS(N,L), is usually manually selected from frontier orbitals that are likely to be highly statically correlated (in our example the π^* -orbitals of dioxygen in its ${}^1\Delta_g$ state). This manual selection, however, is an error prone and tedious task,^[14–16] especially for molecules with a large number of frontier orbitals such as polynuclear transition-metal clusters. In addition, the scaling of these methods is exponential with the size of the active space such that active spaces with more than 18 electrons in 18 orbitals are not accessible to traditional CASSCF.^[17]

New methods overcome this unfavorable scaling of the traditional approach. The density matrix renormalization group (DMRG)^[18–32] and full configuration-interaction quantum Monte-Carlo^[33,34] are two examples. Although these approaches allow for active spaces of 50–100 orbitals, the need for an active space selection remains important for two reasons:

1) The inclusion of only weakly statically correlated orbitals blows up the calculation and ultimately renders calculations on large molecules with many frontier orbitals impossible.

2) Methods that capture the missing part of electron correlation (so-called *dynamical* correlation) may fail for unbalanced active spaces or whenever an undefined but considerably large fraction of this dynamical correlation is already represented in the active space.

The selection of a meaningful active space requires experience with both multi-configurational methods and the system under investigation. A scheme where the selection of the active orbitals is based on unbiased physical quantities is therefore highly desirable and ultimately enables the full automation of the active space selec-

tion. The natural orbital occupation numbers (NOONs) of an unrestricted Hartree–Fock calculation^[35–37] or a Moller–Plesset perturbation theory calculation to second order (MP2)^[38] are suitable quantities for this purpose. However, the small deviations from 0.00 or 2.00 that are typical for NOONs, especially for systems with a large number of close-lying frontier orbitals, make the definition of a suitable and general selection threshold a difficult task. Therefore, we introduced an automated active orbital space selection protocol^[39,40] based on orbital entanglement measures that are closely related to static correlation effects.^[41] Very recently, further work along these lines was presented.^[42]

In this work, we briefly review the orbital entanglement quantities which are the basis of our protocol and discuss them at the dioxygen example in its ${}^1\Delta_g$ state. For ethylene we then show how this approach is applied to describe several excited states of different spin symmetry. Finally, we investigate the Diels–Alder reaction of ethylene with 1,3-butadiene to demonstrate how our protocol selects a consistent active space along a reaction coordinate.

2. Orbital Entanglement and Review of Previous Work

For a spatial orbital four *pure states* can be defined: doubly occupied, spin up, spin down, and unoccupied. A single electron configuration assigns each orbital one of these pure states unambiguously so that, in the light of the definition of static correlation, each orbital contributes an (almost) pure sub-state in the absence of static correlation (*i.e.* one configuration is sufficient to approximate the wave function).

By contrast, when static correlation is pronounced as in the example of the ${}^1\Delta_g$ state of dioxygen (see Fig. 1), the occupation of some of the orbitals will deviate substantially from a pure state. In this example, the occupation of each of the π^* -orbitals is mainly a superposition of a doubly occupied and an empty single-orbital sub-state. The deviation from such a pure state is precisely what necessitates the inclusion of several configurations in the reference wave function (equivalently, it is the presence of several configurations with large weights in the qualitatively correct wave function that produces the deviation from a pure state). A quantity that measures this deviation from a pure state and that can be obtained at low cost may then guide the selection of active orbitals.

The information entropy of a bipartite system is such a measure.^[43,44] Defined for a single orbital i it is referred to as the single-orbital entropy $s_i(1)$ and can be written as Eqn. (1).^[43,45]

$$s_i(1) = - \sum_{\alpha=1}^4 w_{\alpha,i} \ln w_{\alpha,i} \quad (1)$$

where α runs over the four possible orbital occupations and $w_{\alpha,i}$ are the eigenvalues of the one-orbital reduced density matrix (1o-RDM) of orbital i , in which all environment sub-states defined on all other orbitals but i were traced out. It can be derived from the one- and two-particle reduced density matrices (1e- or 2e-RDM)^[44,46] or as expectation values over strings of excitation and annihilation operators.^[47] Analogously, for a pair of orbitals the two-orbital entropy $s_{ij}(2)$ reads as in Eqn. (2)

$$s_{ij}(2) = - \sum_{\alpha=1}^{16} w_{\alpha,ij} \ln w_{\alpha,ij} \quad (2)$$

where α includes now all combinations of the four one-orbital sub-states for each orbital and $w_{\alpha,ij}$ are the eigenvalues of the two-orbital reduced density matrix (2o-RDM) which now can be derived from RDM elements up to fourth order (4e-RDM). Subtracting the individual single-orbital entropies from $s_{ij}(2)$ yields the *mutual information* I_{ij} and is a measure of the entanglement between sub-states defined on these two orbitals as in Eqn. (3).

$$I_{ij} = \frac{1}{2} [s_i(1) + s_j(1) - s_{ij}(2)] (1 - \delta_{ij}) \quad (3)$$

where prefactor and signs correspond to those implemented in our DMRG program QCMaquis^[48–50] and we note that other definitions exist.^[41,45]

Entanglement diagrams that collect these measures in one figure for a given calculation provide intuitive insight into static correlation. One example is shown in the center of Fig. 1. Each orbital is represented by a red circle whose area is proportional to its single-orbital entropy. The thickness of the connecting lines is proportional to the mutual information defined for a pair of orbitals. In the singlet dioxygen example, the two π^* -orbitals have the highest single-orbital entropy and a strong mutual information as we would expect from our difficulties in applying the Aufbau principle. The mutual information further reveals a significant entanglement of the two π -orbitals as well as π - π^* - and σ - σ^* -entanglement.

These entanglement measures can be evaluated from partially converged, and therefore comparatively cheap, DMRG calculations that include a large number of possibly statically correlated frontier orbitals (ideally the whole valence orbital space for a reliable selection).^[39] Based on such

calculations, groups of orbitals with a high single-orbital entropy are automatically identified in our active orbital selection protocol after scaling to the highest $s_i(1)$ value found in the calculation under consideration. The final and fully converged DMRG or CASSCF calculation is then carried out with these orbitals only. For details concerning the selection protocol, we refer to the original publication in ref. [39]. We showed that suitable active orbital spaces were selected for many molecules that are known to be challenging and for which reliable active spaces were established in the literature.

At the example of homolytic dissociation energies of 3d-metallocenes, we further demonstrated^[40] that active spaces selected by our protocol are well-suited for the subsequent inclusion of dynamical correlation, which was in this case included by means of CASSCF perturbation theory to second order.^[51] In addition, we defined a diagnostic for the multi-configurational character of a wave function $Z_{s(1)}$.^[52]

3. Methodology

In this work, we extend the automated active space selection to the calculation of transition energies for several excited singlet and triplet states of ethylene and to the Diels–Alder reaction of ethylene and *cis*-butadiene. The workflow is schematically shown in Fig. 2. The calculation of transition energies for several electronic states with an active space method requires that each state is calculated with the same orbital basis. A state-specific orbital optimization in the final step might lead to slightly different shapes of the orbitals but the orbital character must remain the same for each orbital among all states to guarantee size-consistency of all calculations. Therefore we first generate an orbital basis that is then applied in initial DMRG calculations with large active spaces for all states of interest (panel A of Fig. 2). This orbital basis can be Hartree–Fock or CASSCF orbitals of the ground state but other choices are possible, too. With the automated active space selection protocol we select the most entangled orbitals in each state and construct the final active space from the union of these orbitals. In a final step, the energies of all states are calculated from this active space and dynamical correlation is included.

When we aim to select a consistent active orbital space for several structures along a trajectory or reaction coordinate, we apply the workflow in panel B of Fig. 2. For a given initial structure we generate an initial orbital basis from which we extract the most entangled orbitals with our automated active space selection protocol and optimize these orbitals with state-

averaged or state-specific CASSCF (or DMRG-SCF). These orbitals are the input orbitals for the next structure, for which we optimize the orbitals again, select the highly entangled orbitals according to our protocol, and reoptimize the orbitals if the active space changed. In this way the size of the active space can only *increase* along the reaction coordinate because active orbitals with little entropy for some structures cannot be excluded from the calculation due to their importance for other structures. After the procedure is applied for all structures and a final active space is identified, we go back to the structure, where the final active space was identified for the first time. In order to obtain size-consistent energies and wave functions we now revert the direction and perform the final calculations with the final active space for each structure until we reach the first structure. As the orbital optimization produces new orbitals for every structure, it is possible that the direction in which we traverse the reaction coordinate (*i.e.* reactant to product or vice versa) affects the final active space. We show, however, in the following that a) the neglect of orbitals with little entanglement may be leveled out by the inclusion of dynamical correlation and b) that a slight bias with respect to the entropies introduced by the orbital optimization of already selected active orbitals does not prevent the protocol from picking up highly entangled orbitals that are important to describe the reaction.

All Hartree–Fock, CASSCF^[6,7] and CASPT2^[51] calculations were performed with Molcas 8,^[17] while all DMRG calculations were carried out with our DMRG program QCMaquis.^[31,49,50] We chose the ANO-RCC basis set^[53,54] in its double-zeta contraction for the study of the Diels–Alder reaction, whereas the standard contraction was applied for the calculation of the excited states of ethylene. The active space selection was carried out with a newly developed graphical user interface that combines the automated active space selection protocol,^[39] with the protocols of Fig. 2 and additional analysis functionality such as the calculation of the multi-reference diagnostic $Z_{s(1)}$.^[52]

4. Singlet and Triplet Excited States of Ethylene

The electronically excited states of ethylene were the topic of more than 60 theoretical articles and we refer to the reviews in refs. [55–58] and a recent article^[59] for a comprehensive overview. Because of the size and symmetry of the molecule highly accurate approaches can be applied. The surprisingly diffuse character of the ${}^1B_{1u}$ state is only captured by the most sophis-

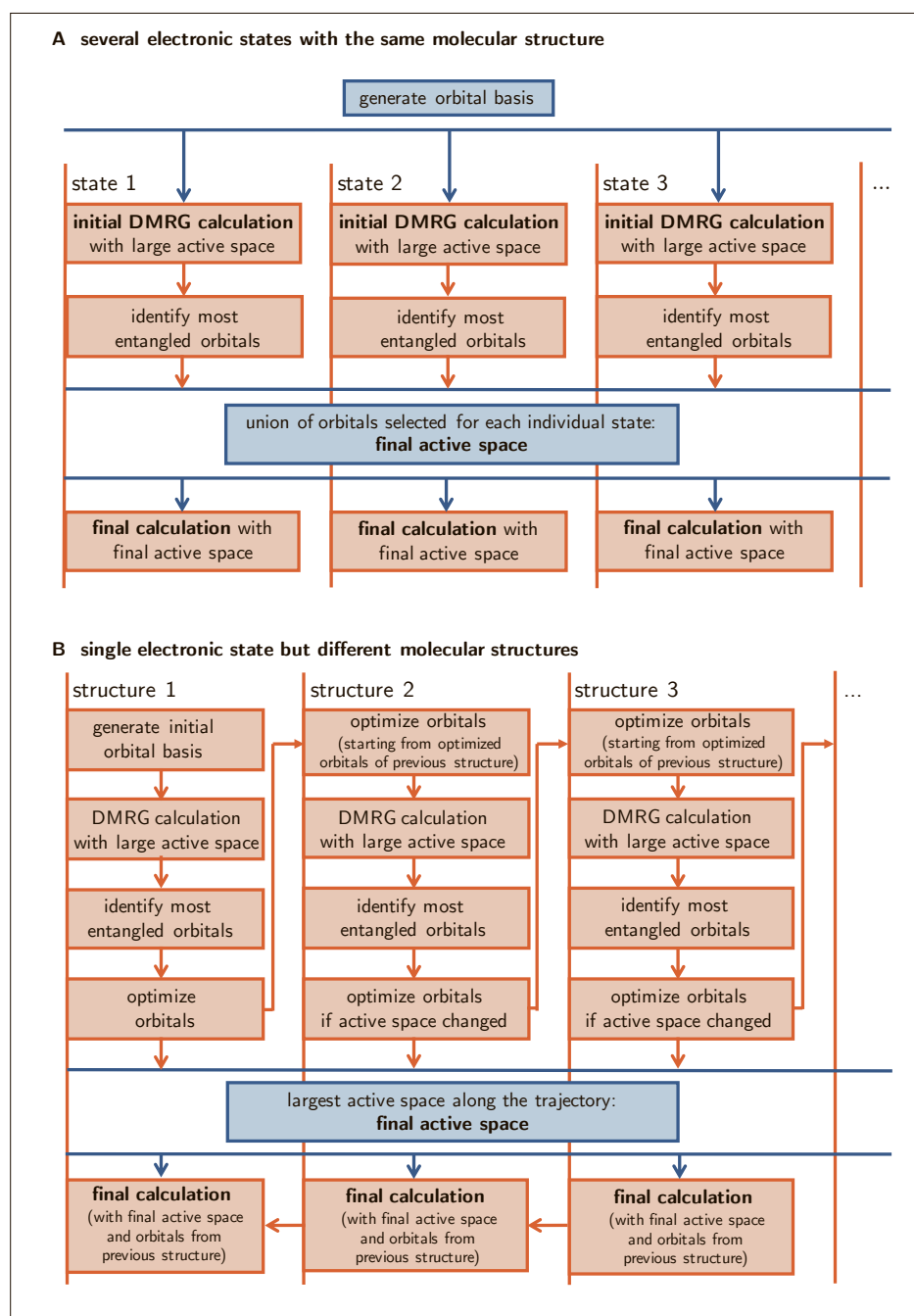


Fig. 2. Flowchart of the automated active space selection for: **A** several electronic states for a given molecular structure and **B** one electronic state but several molecular structures (e.g., along a dynamics trajectory or reaction coordinate). Steps in orange are carried out for an individual state/structure, while steps in blue combine information obtained from all states/structures or distribute information to all states/structures. Note that the direction along a reaction coordinate is in general not unique (it is possible to choose a coordinate from educts to products or vice versa) and the automated active space selection might suggest different final active spaces depending on the direction chosen. These differences, however, concern only orbitals with little entanglement entropy and their effect is levelled out by subsequent inclusion of dynamical correlation (see text). This flowchart is an extension of the automated active space selection protocol proposed in refs. [39,52].

ticated methods.^[59] Here, we illustrate benefits and possible drawbacks of our automated orbital selection approach. We followed the protocol in panel **A** of Fig. 2 to calculate vertical transition energies of four excited singlet and triplet states of ethylene. The initial set of orbitals for the DMRG calculations for all states included 24 (12 valence orbitals and 12 additional

Rydberg orbitals) of the 1A_g ground state Hartree–Fock orbitals.

Table 1 lists vertical transition energies T_v and the expectation value $\langle x^2 \rangle$ for the ground-state and four excited singlet and triplet states of ethylene, where $\langle x^2 \rangle$ is a measure of the diffuseness of a state along the out-of-plane axis (the x axis). All entries of Table 1 were calculated with

CASPT2 with state-specific CASSCF orbitals for three different active spaces. The first, and smallest, CAS(10,11) is the one selected by our automated procedure, whereas the two larger ones are taken from ref. [59], the latter being the largest and recommended active space in that work. A $\langle x^2 \rangle$ value of around 11 a_0^2 is typical for a pure valence state, one around 60 a_0^2 or higher is characteristic of a Rydberg state, and values in between correspond to mixed valence/Rydberg states.^[59] A comparison with the best values from ref. [59] is not meaningful because the authors constructed a hand-tailored atomic orbital basis set with additional diffuse functions. Moreover, their internally contracted state-average CAS configuration interaction method is more advanced than the CASPT2 approach we apply here for demonstration purposes. We therefore compare our CASPT2 results with those obtained from a CASPT2 calculation with the CAS(12,19) that was recommended by the authors of ref. [59]. It is noteworthy that the number of configurations is around 8000 in our case and around 18,000,000 for the large active space recommended in ref. [59]. Nevertheless, our results are in very close agreement to those obtained with the large active space for both T_v and $\langle x^2 \rangle$ for all states but the problematic ${}^1B_{1u}$ state. Hence, our method is capable of selecting the most entangled orbitals and therefore accurately accounts for the static correlation present in several electronic states.

An obvious explanation for the less accurate results that we obtained for the ${}^1B_{1u}$ state might be that the selection is based on orbitals that are optimized for the ground state and therefore differ substantially from the optimized orbitals for that particular state. However, our automated protocol gives identical results when we apply the ${}^1B_{1u}$ CAS(12,19)SCF orbitals as the orbital basis for which we calculate the entropy measures. This is not unexpected because this state is particularly sensitive to the method with which the dynamical correlation is calculated.^[59] Even the highly-accurate internally contracted state-averaged CAS configuration interaction composite approach of ref. [59] had to be adapted for this specific state to give satisfactory results. The results we obtained for this state are therefore not to be mistaken for a failure of our automated active orbital space selection because the static part of the electron correlation is adequately described.

5. Concerted Diels–Alder Reaction of Ethylene with *cis*-Butadiene

The Diels–Alder reaction of ethylene and *cis*-butadiene is a prototype for a [4+2]

Table 1. Vertical CASPT2 transition energies T_v (in eV) and $\langle x^2 \rangle$ values (in a_0^2) for active spaces of three different sizes. Values in parentheses correspond to differences with respect to the calculation with the largest active space.

state	CASPT2 with orbitals from						exp. ^b	
	auto-CAS(10,11)		CAS(12,16)		CAS(12,19)			icCAS-CI ^a
	T_v	$\langle x^2 \rangle$	T_v	$\langle x^2 \rangle$	T_v	$\langle x^2 \rangle$	T_v	
1A_g	0.00	11.49	0.00	11.56	0.00	11.56	0.00	0.00
$^1B_{3u}$	7.52 (0.06)	20.98	7.46	21.05	7.46	20.96	7.39	7.11
$^1B_{1u}$	8.31 (0.30)	23.65	8.11	20.76	8.01	18.89	8.05	7.68
$^1B_{1g}$	8.18 (0.04)	18.53	8.15	18.83	8.14	18.75	7.98	7.8
$^1B_{2g}$	8.20 (0.03)	19.74	8.27	19.60	8.17	19.51	8.05	7.9
$^3B_{3u}$	7.42 (0.07)	20.94	7.35	20.98	7.35	20.95	7.24	6.98
$^3B_{1u}$	4.53 (-0.01)	11.45	4.55	11.64	4.54	11.64	4.54	4.36
$^3B_{1g}$	8.14 (0.05)	18.21	8.11	18.53	8.10	18.44	7.94	7.79
$^2B_{2g}$	8.14 (0.04)	19.32	8.20	19.32	8.10	19.00	7.97	–

^aBest estimate from ref. [59] internally contracted state-average CAS configuration interaction calculations with core-valence correction and extrapolation to the basis set limit of hand-tailored atomic orbital bases with diffuse functions; ^bValues taken from ref. [59] with data from ref. [60].

cycloaddition and was extensively studied with multi-configurational methods.^[61–65] For this example we show that a consistent selection of active orbitals (and therefore a smooth potential energy surface) is possible with our automated approach (see panel **B** of Fig. 2). We chose a reaction coordinate consisting of 20 standard non-self-consistent density-functional tight-binding (DFTB)^[66,67] structures taken from ref. [68]. These structures result from a constrained optimization along a C_s -symmetric reaction coordinate. The large initial active orbital space for which we calculate the entanglement entropy contains all 34 valence orbitals (2s- and 2p-orbitals on carbon and 1s-orbitals on hydrogen) from a restricted Hartree–Fock calculation in case of the first structure. The automatically selected active orbitals are then optimized with CASSCF and for all subsequent structures along the reaction coordinate, the optimized orbitals of the previous structure serve as starting orbitals for the CASSCF calculation.

Results of these calculations are shown in Fig. 3. When we define the reaction coordinate to start from separated molecules, the automatically selected active space is a CAS(6,6) for all structures along this coordinate (upper diagram of Fig. 3). These active orbitals change their shape significantly during the reaction and resemble those displayed in the top part of Fig. 4. While they describe the π -system of the separated molecules at early stages of the reaction, they transform to the two σ_{cc} -bonds and the remaining π/π^* -orbitals of the product. A transition state can be observed close to a C_1 – C_2 distance of 2.25 Å (see the upper part of Fig. 3 for a definition of the reaction coordinate). The

most stable structure is found close to a C_1 – C_2 distance of 1.55 Å. For even shorter C_1 – C_2 distances, the electronic energy

rises again and the C_2 – C_3 distance increases significantly.

The energy diagram in the middle of Fig. 3 shows the results of CASSCF calculations for automatically selected active spaces for the inverse definition of the reaction coordinate from right to left in Figure 3. Interestingly, the long C_2 – C_3 distance of the initial structure is already indicating bond breaking as monitored by high orbital entropies for the σ - and σ^* -orbitals of these bonds in a Hartree–Fock basis. Hence, the initially selected CAS(6,6) (gray circles in the central diagram of Fig. 3) contains those four orbitals as well as the π/π^* orbitals of the cycloadduct. When the C_1 – C_2 distance reaches about 2.0 Å, four additional orbitals are selected by the automated procedure (red circles in the central diagram of Fig. 3). These orbitals are σ - and σ^* -orbitals of the bonds that are about to break. This CAS(10,10) is also the largest active space selected along the reaction coordinate and is therefore our final active space. Following the protocol of panel **B** of Fig. 2, we apply this final active space now also for the remaining structures of the reaction coordinate (for which we

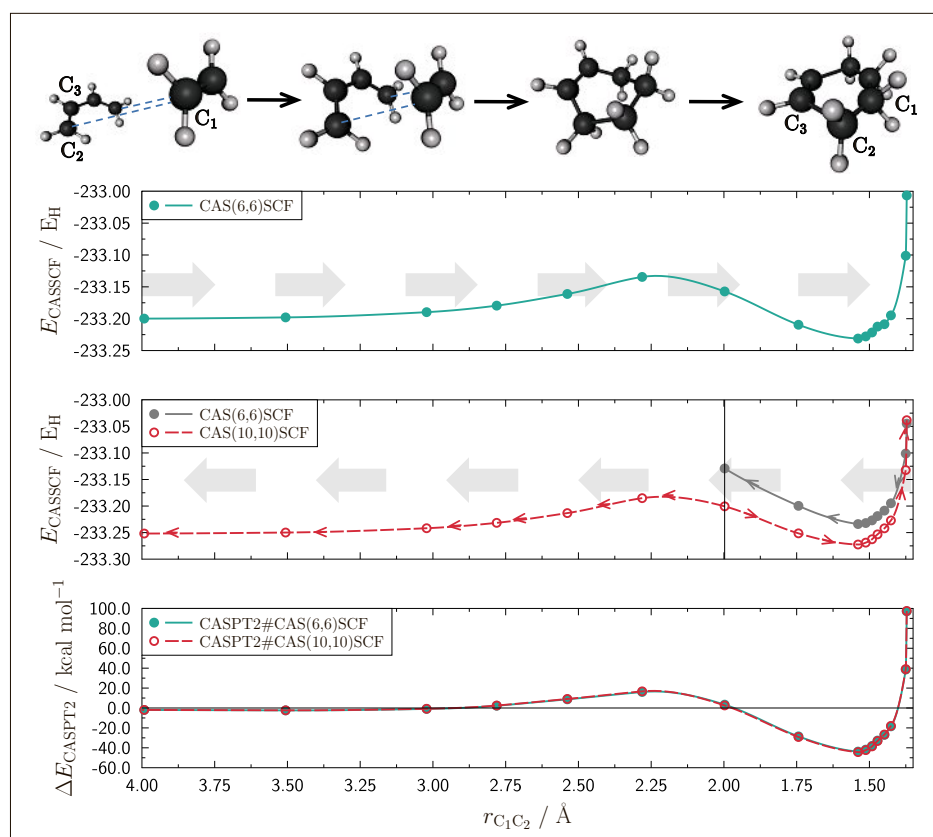


Fig. 3. Electronic energy profiles of the concerted Diels–Alder reaction of ethylene and *cis*-butadiene. The reaction path is shown on the top. The distance between the carbon atoms C_1 and C_2 is chosen as the reaction coordinate in a C_s -symmetric constrained optimization. The upper energy diagram shows the CASSCF energy of a CAS(6,6) that is consistently selected by the automated active space selection for all structures considered when starting at the educts. Defining the reaction coordinate in the opposite direction (middle panel) results in a CAS(6,6) (gray) until a C_1 – C_2 distance of about 2.0 Å is reached, where four additional orbitals are selected to yield a CAS(10,10) (red). CASPT2 energies based on both CAS(6,6)SCF (turquoise) and CAS(10,10)SCF (red) calculations are shown in the lower panel and are virtually indistinguishable.

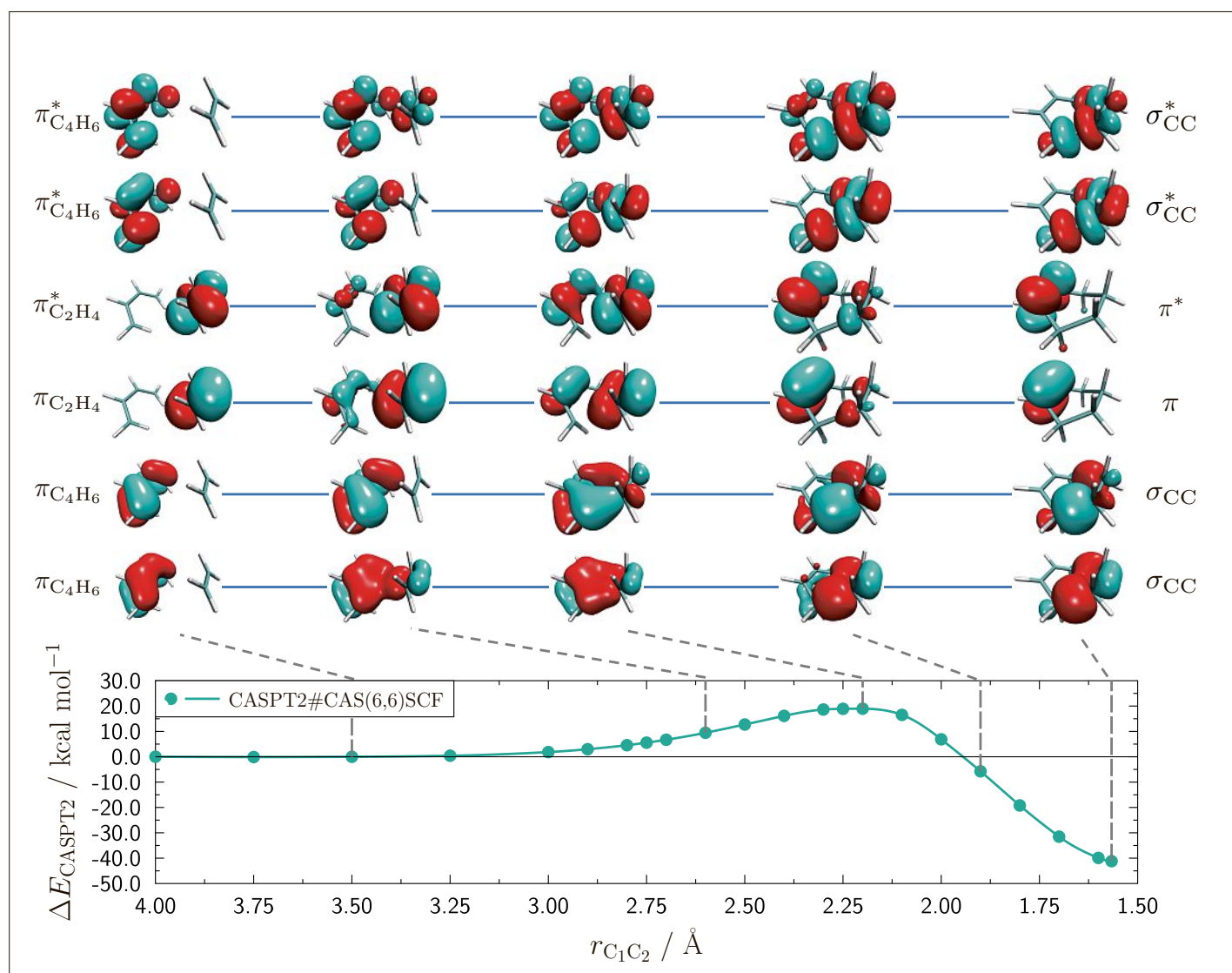


Fig. 4. Contour plots of the six active orbitals for five structures from a constrained CAS(6,6)SCF structure optimization along the $r_{C_1C_2}$ reaction coordinate of the concerted Diels–Alder reaction between ethylene and *cis*-butadiene. The lower part shows CASPT2 energies for the CAS(6,6)SCF structures and reference wave functions. The energies are taken relative to that of both molecules at infinite distance.

had applied a smaller active space before) to obtain a smooth potential energy curve.

The discrepancy in the selected active spaces depending on the definition of the reaction coordinate may seem odd at first glance but has virtually no effect in actual calculations. The first reason is that the thresholds in our automated procedure are designed to rather include orbitals with little entanglement than omit important orbitals such that the resulting active spaces are guaranteed to include all static correlation effects. This difficulty follows from the smooth transition between static and dynamic correlation. The second reason is that in chemistry we are rarely interested in total energies and often aim to calculate relative energies.

In the lower diagram of Fig. 3, we present CASPT2 energies relative to the energy of the separated molecules for both active spaces. Obviously, the energies are indistinguishable for both active spaces such that the missing correlation in

the CAS(6,6)SCF calculation is fully recovered by the CASPT2 calculation. This is in line with our previous work on the dissociation energy of metallocenes,^[52] where missing size-consistency stemming from different sizes of the combined active spaces of the dissociated fragments and the complex was counterbalanced by the inclusion of dynamical correlation as calculated with CASPT2. Therefore, our automated procedure will give consistent results as long as it guarantees that all static correlation effects are included in the reference wave function and hence that all *strongly* entangled orbitals are included. This is regulated by a single threshold in the automated protocol that is tuned accordingly.

For the final calculation of a reaction coordinate we focused on the region between the minimum in each of the diagrams of Fig. 3 and the dissociated molecule. In this region, only the six orbitals leading to the CAS(6,6) showed strong entanglement and

were selected for CAS(6,6)SCF structure optimizations. The minimum CAS(6,6) SCF structure was optimized without any constraint starting from the DFTB structure with $r_{C_1C_2} = 1.539 \text{ Å}$ with a resulting C_1 – C_2 distance of 1.56643 Å . All other structures were calculated by constraining the C_1 – C_2 distance (and C_1' – C_2' distance because of the C_s symmetry) to 22 values between 1.6 and 4.0 Å . Corresponding CASPT2 energies relative to the energy of the dissociated molecule are displayed in Fig. 4 along with contour plots of the six active orbitals for five structures along the reaction coordinate. The potential energy curve is almost indistinguishable from the one obtained with the DFTB structures. We note here that the change of the orbital character along this reaction coordinate might cause problems in the active space selection procedure if the changes between two subsequent structures are too pronounced so that large orbital rotations from the inactive or virtual space to the active orbital

space occur during orbital optimization. In that case, kinks in the CASSCF potential energy curve might occur and hence are a way to identify this possible obstacle.

6. Conclusions and Outlook

The selection of active orbital spaces is a tedious task that is essential to most multi-configurational calculations. We reviewed our automated protocol that rates orbitals according to entanglement entropy measures. This orbital entanglement is closely connected to the static correlation effects that are to be described by these multi-configurational calculations.^[41] While our previous work focused on static and ground-state properties, the emphasis here was on the consistent description of several excited states for a given molecular structure and the selection of an active space that is uniform along a reaction coordinate. In both cases, the automated procedure identifies the union of orbitals that are selected for the individual states or structures as the final active space although slightly different protocols are applied.

We emphasize that our automated active space selection will not be possible whenever the large preliminary DMRG calculation is either unfeasible or not converged with respect to the CAS size.

The calculation of the orbital entanglement entropy facilitates the full automation of the automated active orbital selection procedure and ultimately enables non-experts to easily carry out multi-configurational calculations.

Acknowledgments

This work was supported by the Schweizerischer Nationalfonds (No. 20020_169120). C.J.S. gratefully acknowledges a Kekulé fellowship from the Fonds der Chemischen Industrie.

Received: February 2, 2017

- [1] A. Karton, E. Rabinovich, J. M. L. Martin, B. Ruscic, *J. Chem. Phys.* **2006**, *125*, 144108.
- [2] W. Klopper, R. A. Bachorz, D. P. Tew, C. Hättig, *Phys. Rev. A* **2010**, *81*, 022503.
- [3] D. Feller, K. A. Peterson, D. A. Dixon, *Mol. Phys.* **2012**, *110*, 2381.
- [4] R. Parr, W. Yang, 'Density-Functional Theory of Atoms and Molecules', Oxford University Press, USA, **1994**.
- [5] T. Helgaker, P. Jørgensen, J. Olsen, 'Molecular Electronic-Structure Theory', 2nd ed., Wiley, Chichester, **2014**.
- [6] B. O. Roos, P. R. Taylor, P. E. Siegbahn, *Chem. Phys.* **1980**, *48*, 157.
- [7] H.-J. Werner, P. J. Knowles, *J. Chem. Phys.* **1985**, *82*, 5053.
- [8] B. O. Roos, *Adv. Chem. Phys.* **1987**, *69*, 399.
- [9] R. Shepard, *Adv. Chem. Phys.* **1987**, *69*, 63.
- [10] K. Ruedenberg, K. R. Sundberg, 'Quantum science, methods, and structure', Plenum Press, New York, **1976**.
- [11] K. Ruedenberg, M. W. Schmidt, M. M. Gilbert, S. Elbert, *Chem. Phys.* **1982**, *71*, 41.
- [12] K. Ruedenberg, M. W. Schmidt, M. M. Gilbert, *Chem. Phys.* **1982**, *71*, 51.
- [13] K. Ruedenberg, M. W. Schmidt, M. M. Gilbert, S. Elbert, *Chem. Phys.* **1982**, *71*, 65.
- [14] R. D. Bach, 'The chemistry of peroxides', Wiley, Chichester, **2006**.
- [15] P.-Å. Malmqvist, K. Pierloot, A. R. M. Shahi, C. J. Cramer, L. Gagliardi, *J. Chem. Phys.* **2008**, *128*, 204109.
- [16] V. Veryazov, P.-Å. Malmqvist, B. O. Roos, *Int. J. Quantum Chem.* **2011**, *111*, 3329.
- [17] 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-Mart, M. Stenrup, D. G. Truhlar, L. Ungur, A. Valentini, S. Vancoillie, V. Veryazov, V. P. Vysotskiy, O. Weingart, F. Zapata, R. Lindh, *J. Comput. Chem.* **2016**, *37*, 506.
- [18] S. R. White, *Phys. Rev. Lett.* **1992**, *69*, 2863.
- [19] S. R. White, *Phys. Rev. B* **1993**, *48*, 10345.
- [20] Ö. Legeza, R. Noack, J. Sólyom, L. Tincani, *Lect. Notes Phys.* **2008**, *739*, 653.
- [21] G. K.-L. Chan, J. J. Dorando, D. Ghosh, J. Hachmann, E. Neuscamman, H. Wang, T. Yanai, *Prog. Theor. Chem. Phys.* **2008**, *18*, 49.
- [22] G. K.-L. Chan, D. Zgid, *Annu. Rep. Comput. Chem.* **2009**, *5*, 149.
- [23] K. H. Marti, M. Reiher, *Z. Phys. Chem.* **2010**, *224*, 583.
- [24] K. H. Marti, M. Reiher, *Phys. Chem. Chem. Phys.* **2011**, *13*, 6750.
- [25] G. K.-L. Chan, S. Sharma, *Ann. Rev. Phys. Chem.* **2011**, *62*, 465.
- [26] U. Schollwöck, *Annals of Physics* **2011**, *326*, 96.
- [27] Y. Kurashige, *Mol. Phys.* **2014**, *112*, 1485.
- [28] S. Wouters, D. Van Neck, *Eur. Phys. J. D* **2014**, *68*, 272.
- [29] T. Yanai, Y. Kurashige, W. Mizukami, J. Chalupsky, T. N. Lan, M. Saitow, *Int. J. Quantum Chem.* **2015**, *115*, 283.
- [30] S. Szalay, M. Pfeiffer, V. Murg, G. Barcza, F. Verstraete, R. Schneider, Ö. Legeza, *Int. J. Quantum Chem.* **2015**, *115*, 1342.
- [31] S. Knecht, E. D. Hedegård, S. Keller, A. Kovyrshin, Y. Ma, A. Muolo, C. J. Stein, M. Reiher, *Chimia* **2016**, *70*, 244.
- [32] G. K.-L. Chan, A. Keselman, N. Nakatani, Z. Li, S. R. White, *J. Chem. Phys.* **2016**, *145*, 014102.
- [33] G. H. Booth, A. J. W. Thom, A. Alavi, *J. Chem. Phys.* **2009**, *131*, 054106.
- [34] G. H. Booth and A. Alavi, *J. Chem. Phys.* **2010**, *132*, 174104.
- [35] P. Pulay, T. P. Hamilton, *J. Chem. Phys.* **1988**, *88*, 4926.
- [36] J. M. Boffill, P. Pulay, *J. Chem. Phys.* **1989**, *90*, 4926.
- [37] S. Keller, K. Boguslawski, T. Janowski, M. Reiher, P. Pulay, *J. Chem. Phys.* **2015**, *142*, 244104.
- [38] H. J. A. Jensen, P. Jørgensen, H. Ågren, J. Olsen, *J. Chem. Phys.* **1988**, *88*, 3834.
- [39] C. J. Stein, M. Reiher, *J. Chem. Theory Comput.* **2016**, *12*, 1760.
- [40] C. J. Stein, V. von Burg, M. Reiher, *J. Chem. Theory Comput.* **2016**, *12*, 3764. [41] K. Boguslawski, P. Tecmer, Ö. Legeza, M. Reiher, *J. Phys. Chem. Lett.* **2012**, *3*, 3129.
- [42] E. R. Sayfutyarova, Q. Sun, G. K.-L. Chan, G. Knizia, arXiv:1701.07862 **2017**.
- [43] Ö. Legeza and J. Sólyom, *Phys. Rev. B* **2003**, *68*, 195116.
- [44] J. Rissler, R. M. Noack, S. R. White, *Chem. Phys.* **2006**, *323*, 519.
- [45] Ö. Legeza, J. Sólyom, *Phys. Rev. Lett.* **2006**, *96*, 116401.
- [46] K. Boguslawski, P. Tecmer, *Int. J. Quantum Chem.* **2015**, *115*, 1289.
- [47] K. Boguslawski, P. Tecmer, G. Barcza, Ö. Legeza, M. Reiher, *J. Chem. Theory Comput.* **2013**, *9*, 2959.
- [48] M. Dolfi, B. Bauer, S. Keller, A. Kosenkov, T. Ewart, A. Kantian, T. Giamarchi, M. Troyer, *Comput. Phys. Commun.* **2014**, *185*, 3430.
- [49] S. Keller, M. Dolfi, M. Troyer, M. Reiher, *J. Chem. Phys.* **2015**, *143*, 244118.
- [50] S. Keller, M. Reiher, *J. Chem. Phys.* **2016**, *144*, 134101.
- [51] K. Andersson, P.-Å. Malmqvist, B. O. Roos, *J. Chem. Phys.* **1992**, *96*, 1218.
- [52] C. J. Stein, M. Reiher, *Mol. Phys.* **2016**, DOI:10.1080/00268976.2017.1288934.
- [53] K. Pierloot, B. Dumez, P.-O. Widmark, B. O. Roos, *Theor. Chim. Acta* **1995**, *90*, 87.
- [54] B. O. Roos, R. Lindh, P.-Å. Malmqvist, V. Veryazov, P.-O. Widmark, *J. Phys. Chem. A* **2004**, *108*, 2851.
- [55] S. D. Peyerimhoff, R. J. Buenker, *Adv. Quant. Chem.* **1975**, *9*, 69.
- [56] W. T. Bonden, E. R. Davidson, *Acc. Chem. Res.* **1996**, *29*, 67.
- [57] R. J. Cave, in 'Modern Electronic Structure Theory and Applications in Organic Chemistry', Ed. E. R. Davidson, World Scientific, **1997** p. 197.
- [58] R. J. Buenker, G. Hirsch, Y. Liu, in 'Understanding Chemical Reactivity', Vol. 20, Ed. C. Sándorfy, Kluwer Academic Publishers, **1999** p. 57.
- [59] D. Feller, K. A. Peterson, E. R. Davidson, *J. Chem. Phys.* **2014**, *141*, 104302.
- [60] M. Robin, 'Higher excited states of polyatomic molecules', Vol. 3, Elsevier, Amsterdam, **2012**.
- [61] F. Bernardi, A. Bottoni, M. A. Robb, M. J. Field, I. H. Hillier, M. F. Guest, *J. Chem. Soc., Chem. Commun.* **1985**, 1051.
- [62] Y. Li, K. N. Houk, *J. Am. Chem. Soc.* **1993**, *115*, 7478.
- [63] S. Sakai, *J. Phys. Chem. A* **2000**, *104*, 922.
- [64] H. Lischka, E. Ventura, M. Dallos, *ChemPhysChem* **2004**, *5*, 1365.
- [65] Z. Chen, X. Chen, F. Ying, J. Gu, H. Zhang, W. Wu, *J. Chem. Phys.* **2014**, *141*, 134118.
- [66] D. Porezag, T. Frauenheim, T. Köhler, G. Seifert, R. Kaschner, *Phys. Rev. B* **1995**, *51*, 12947.
- [67] G. Seifert, D. Porezag, T. Frauenheim, *Int. J. Quantum Chem.* **1996**, *58*, 185.
- [68] M. P. Haag, A. C. Vaucher, M. Bosson, S. Redon, M. Reiher, *ChemPhysChem* **2014**, *15*, 3301.