

Automated construction of molecular active spaces from atomic valence orbitals

Elvira R. Sayfutyarova, Qiming Sun, Garnet Kin-Lic Chan, and Gerald Knizia

J. Chem. Theory Comput., **Just Accepted Manuscript** • DOI: 10.1021/acs.jctc.7b00128 • Publication Date (Web): 21 Jul 2017

Downloaded from <http://pubs.acs.org> on July 24, 2017

Just Accepted

“Just Accepted” manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides “Just Accepted” as a free service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. “Just Accepted” manuscripts appear in full in PDF format accompanied by an HTML abstract. “Just Accepted” manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are accessible to all readers and citable by the Digital Object Identifier (DOI®). “Just Accepted” is an optional service offered to authors. Therefore, the “Just Accepted” Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the “Just Accepted” Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these “Just Accepted” manuscripts.

Automated construction of molecular active spaces from atomic valence orbitals

Elvira R. Sayfutyarova,[†] Qiming Sun,[‡] Garnet Kin-Lic Chan,[‡] and Gerald Knizia^{*,¶}

[†]*Department of Chemistry, Princeton University, Princeton, NJ 08540, USA*

[‡]*Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125, USA*

[¶]*Department of Chemistry, The Pennsylvania State University, University Park, PA 16802, USA*

E-mail: knizia@psu.edu

Abstract

We introduce the atomic valence active space (AVAS), a simple and well-defined automated technique for constructing active orbital spaces for use in multi-configuration and multi-reference (MR) electronic structure calculations. Concretely, the technique constructs active molecular orbitals capable of describing *all* relevant electronic configurations emerging from a targeted set of atomic valence orbitals (e.g., the metal *d* orbitals in a coordination complex). This is achieved via a linear transformation of the occupied and unoccupied orbital spaces from an easily obtainable single-reference wavefunction (such as from a Hartree-Fock or Kohn-Sham calculations) based on projectors to targeted atomic valence orbitals. We discuss the premises, theory, and implementation of the idea, and several of its variations are tested. To investigate the performance and accuracy, we calculate the excitation energies for various transition metal complexes in typical application scenarios. Additionally, we follow the homolytic bond breaking process of a Fenton reaction along its reaction coordinate. While the described AVAS technique is not an universal solution to the active space problem, its premises are fulfilled in many application scenarios of transition metal chemistry and bond

dissociation processes. In these cases the technique makes MR calculations easier to execute, easier to reproduce by any user, and simplifies the determination of the appropriate size of the active space required for accurate results.

1 Introduction

Multiconfigurational and multireference (MR) methods remain indispensable in the treatment of challenging electronic structure problems. Transition metal complexes provide a rich source of examples, as they often feature strongly correlated electronic degrees of freedom, which render density functional theory (DFT) calculations unreliable. Unfortunately, MR methods require an a priori choice of a suitable set of active molecular orbitals (an *active space*), which critically determines the quality of the results. The choice of active space is non-trivial and often represents a major challenge in practical computations.

The standard way to choose an active space is as follows: First, a full set of molecular orbitals of the molecule is computed with a simple electronic structure method, such as Hartree-Fock (HF) or Kohn-Sham DFT (KS-DFT). Second, these molecular orbitals, both occupied and unoccupied, are visually inspected, and based on their shape, energy, occupation numbers, etc., one selects into the active space the set of orbitals which are expected to be chemically most relevant (e.g., with significant transition metal *d*-electron character or ligand character in the case of transition metal complexes—regions which are empirically known to be important). But despite the existence of general advice on how to select a good set of starting orbitals,^{1–3} there are some problems with this approach. First, the selection of molecular orbitals for the active space is performed by the user, normally based on personal experience. This makes MR methods hard to apply, and gives results that are complex to reproduce and hard to judge in terms of quality. This stands in contrast to single-reference calculations which do not require active spaces and therefore do not have this level of arbitrariness. Second, the molecular orbitals mix together valence orbitals of different character; for example, typically a metal's *d* atomic valence orbitals contribute to a very large number of

1
2
3
4 molecular orbitals, and it is often not easy to truncate these to a small active subset in such a way
5 that they retain the capability of describing all the right physics in complicated systems. In the
6 case of large complexes, especially with multiple metal ions, this procedure commonly becomes a
7 matter of trial and error.
8
9

10
11 Techniques to aid in the construction of high quality active spaces are therefore highly desir-
12 able, particularly as more powerful electronic structure methods are becoming available which are
13 capable of treating increasingly large active spaces efficiently.⁴⁻²⁵ There have been a number of
14 contributions in this area, and the general procedures can loosely be summarized as based on esti-
15 mating the (correlated) occupation numbers of the orbitals (including the full single orbital density
16 matrix in Ref. 26,27), followed by selecting the active orbitals based on partial occupancy accord-
17 ing to an input threshold. The various techniques differ primarily in how the occupation number
18 information is obtained. For example, correlated occupation numbers can be estimated from un-
19 restricted Hartree-Fock^{28,29} or Kohn-Sham calculations or from correlated calculations, such as
20 MP2^{30,31} or approximate DMRG calculations.^{27,32-34}
21
22
23
24
25
26
27
28
29
30
31

32 Nonetheless, while these existing automated approaches are advancements from typical ad hoc
33 active space constructions, there is still room for further improvement. For example, an obvious
34 drawback of the above procedures is that they all require a non-trivial preliminary calculation:
35 either a correlated calculation must be performed, or a suitable broken symmetry solution must
36 be found. This is not always possible: for example, there may not always be a broken symmetry
37 solution involving the region of interest, or the preliminary correlated calculation may simply be
38 too costly. A perhaps less obvious drawback, but one which one encounters in practice, is that there
39 is no guarantee that the active orbitals found in these automated procedures are actually spatially
40 located in the region of chemical interest. For example, in models of enzymatic binding sites,
41 particularly those with charged ligands, the unpaired electrons may lie in functional groups which
42 are spatially far from and irrelevant to the chemistry of the metal center. In this case, an additional
43 inspection is once again required to choose the subset of active orbitals of chemical interest.
44
45
46
47
48
49
50
51
52
53
54
55

56 Here we propose an alternative approach to construct molecular active spaces for multirefer-
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

ence problems automatically and systematically. This approach does not suffer from the drawbacks mentioned above, and is particularly well-suited for practical calculations involving transition metal complexes. In its simplest formulation, the procedure requires only an easily obtained single-determinant reference function, together with a choice of target atomic valence orbitals. The technique is based on the following idea: It is empirically known that one can typically identify a small set of atomic valence orbitals which give rise to the strong correlation effects (for example, *d* atomic valence orbitals in transition metal complexes). We therefore aim to construct a set of active molecular orbitals by defining them in terms of these atomic valence orbitals. Concretely, using simple linear algebra, we can define mathematical rotations of the occupied and virtual molecular orbitals of the single-determinant reference function which maximizes their given atomic valence character (e.g. *3d* character). After this rotation, the relevant molecular orbitals to include into the active space can be selected automatically. The ideas employed here are closely related to earlier work of Iwata³⁵ and Schmidt and coworkers³⁶ regarding the construction of approximate valence spaces; however these earlier works focused on constructing molecular orbitals capable of spanning the *entire* valence space (not only the valence space of selected atomic orbitals, with the goal of identifying active spaces for processes involving said atomic orbitals), and the latter work also markedly differs in intent. Similar mathematical techniques were also used in the construction of molecule-intrinsic minimal basis sets capable of spanning a set of given occupied orbital spaces.^{37–48}

Sec. 2 explains the motivation, background, theory, and implementation of the construction proposed here. Sec. 4 then describes criteria for judging the quality of the constructed active spaces, and discusses its application to a large number of prototypical MR calculations of metal complexes. Sec. 6 describes conclusions and possible implication for future work.

2 Theory

2.1 What is an active space?

Many naturally occurring stable molecules have an electronic structure which is qualitatively well described by a single-determinant self-consistent field (SCF) wave function, such as used in Kohn-Sham Density Functional Theory (KS-DFT) or Hartree-Fock (HF). In this case, the entire space of molecular orbitals (MO) is strictly divided into fully occupied and fully unoccupied molecular (spin-) orbitals. However, there are important classes of chemical systems in which this picture breaks down and where a superposition of multiple N -electron determinants is required to describe the electronic structure even qualitatively. This phenomenon is sometimes called *strong correlation*; prototypical cases in which it occurs are (a) the process of homolytic bond breaking, and (b) various kinds of transition metal complexes—particularly when the complex is in an overall low-spin state generated by coupling of the metal to other metals or redox non-innocent ligands, as frequently encountered in catalysis and in bio-inorganic systems.

Both of these prototypical cases share the same root cause: The occurrence of valence atomic orbitals with energy levels similar to other valence orbitals, but with poor orbital overlap, giving rise to small energy splittings between bonding and anti-bonding linear combinations. In these cases quantum resonances between both bonding and anti-bonding orbitals must be explicitly considered to qualitatively describe the electronic structure—and single-determinant wave functions are incapable of doing so, because in these each MO either *is* occupied or *is not*, but not both. The core idea developed in this manuscript is that in these two most important cases, the emergence of strong correlation is tightly linked to *specific valence atomic orbitals, which are easy to identify*. In the case of transition metals, these are the compact d orbitals (and possibly some specific ligand orbitals), and in the case of bond dissociation, these are the valence atomic orbitals of the dissociating atom(s). Thus, in the following we will assume that a small number of specific valence atomic orbitals are explicitly selected by the user (e.g., the d orbitals of metal centers), and that the goal is to construct an active space suitable for describing all highly relevant determinants that

1
2
3 they give rise to. We stress that this initial selection is easy in practice—the core problem is how
4 to use the information to build a suitable active space.
5
6

7 We thus briefly consider what an active space *is*. An active space wavefunction is one where
8 the superpositions of determinants are restricted so that varied occupations are found only within
9 the active orbitals $\{\varphi_1 \dots \varphi_m\}$. Technically, this means that the wavefunction can be written as a
10 second quantized product
11
12
13
14

$$|\Psi\rangle = \{\varphi_1 \dots \varphi_m\}|\text{core}\rangle \quad (1)$$

15
16
17 where $\{\varphi_1 \dots \varphi_m\}$ denotes a general occupancy wavefunction within the active orbitals, and $|\text{core}\rangle$
18 denotes a single determinant. It is clear that the active orbitals must span the space of our chosen
19 specific valence atomic orbitals. However, Eq. (1) additionally implies that the rest of the molecule
20 must be well described by the single core determinant. To achieve this, the active space must
21 contain orbitals additional to our set of specific valence atomic orbitals, and to remain compact,
22 we need to define the *minimal* additional set.
23
24
25
26
27
28
29
30
31
32

33 To this end, we here employ the basic observation in density matrix embedding theory (DMET),^{49–56}
34 which describes how to construct such an active space explicitly. In particular, DMET tells us that
35 the active space with the above properties is at most twice the size of the initial set of chosen
36 valence atomic orbitals. However, in this work we modify the presentation and practice of the
37 DMET procedure to make it more natural in the active space setting. In particular, in contrast to
38 the original presentation in references 49,50 (but as described in the appendix of Ref. 51) we do
39 not introduce separate “fragment” and “bath” orbitals, but rather retain the occupied and virtual
40 character of the constructed active orbitals. This has the important benefit that it leads to a natural
41 truncation procedure, which allows us to further reduce the size of the active space in the most
42 chemically meaningful way. In Sec. 2.2 we describe the isolation of entangled orbitals for the
43 active space construction, and Sec. 2.3 will provide technical details and discuss various practical
44 aspects relevant in the active space case. Finally, Sec. 5 outlines how the presented approach
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 may be extended to more complex cases, such as double-shell effects or complex metal/ligand
4 interactions.
5
6

2.2 Isolating target-overlapping orbitals for the active space

7
8
9
10
11 Let $A = \{|p\rangle\}$ denote the (small) set of chosen target valence atomic orbitals (not necessarily
12 orthonormal), which we expect to be responsible for strong correlations (e.g. the five $3d$ orbitals
13 of a third row transition metal atom of a complex, details on their selection and representation
14 will follow). Let $|\Phi\rangle$ denote a Slater determinant, which represents the electronic structure of our
15 system of interest at the SCF level (HF or KS-DFT).
16
17
18
19
20
21

22 Our method for active space construction is built around the following physical assumptions,
23 which are those used in DMET:^{49,50,52,57} (a) a SCF wave function $|\Phi\rangle$ may be unable to describe
24 precisely how our target atomic orbitals are bonded with the rest of the molecule; however, it will
25 generally describe the *rest* of the molecule reasonably well (experience in DMET suggest that this
26 works even in cases where the SCF wave function as a whole is qualitatively *very* wrong⁵⁰), and
27 (b) we can isolate the part of $|\Phi\rangle$ which involves our target AOs from the part which does not, and
28 employ the former part as the active space (therefore allowing it to be replaced by a more powerful
29 wave function description), and retain the simple determinantal description of rest.
30
31
32
33
34
35
36
37
38

39 To this end, we employ a rotation within the set of occupied molecular orbitals of $|\Phi\rangle$ which
40 splits them into two groups: one group which has overlap with our target AOs, and one group
41 which does not. A simple dimensional counting argument will show that for a set of $|A|$ selected
42 target AOs, there is a rotation of the occupied molecular orbitals such that at most $|A|$ of them have
43 non-zero weight on the target AOs. We similarly split the set of virtual molecular orbitals into
44 one group of at most $|A|$ virtual orbitals which have weight on the target atomic valence orbitals,
45 and the other group which does not. The idea is now to explicitly construct these rotated orbital
46 groups, and then employ the at most $2|A|$ combined occupied and virtual orbitals with target over-
47 lap as active orbitals, while leaving the other occupied and virtual orbitals without target overlap as
48 inactive (closed-shell) or virtual orbitals, respectively, in the following multi-configuration treat-
49
50
51
52
53
54
55
56
57
58
59
60

ment. As all of the selected AOs in A then lie within the span of this active space, the resulting multi-configurational wave function is then capable of representing arbitrary quantum resonances involving the target AOs.

We first discuss the occupied case. Let $i = 1 \dots N_{\text{occ}}$ denote N_{occ} occupied molecular orbitals (MOs) of $|\Phi\rangle$. The projector \hat{P} onto the space of atomic orbitals in A , is given by

$$\hat{P} = \sum_{p,q \in A} |p\rangle [\sigma^{-1}]_{pq} \langle q|. \quad (2)$$

Here σ denotes the $|A| \times |A|$ target AO overlap matrix with elements $[\sigma]_{pq} = \langle p|q\rangle$, and σ^{-1} its matrix inverse.

Employing these projectors, we construct the first set of active orbitals by rotating $|\Phi\rangle$'s occupied MOs $\{|i\rangle\}$ as follows. First, we calculate the $N_{\text{occ}} \times N_{\text{occ}}$ overlap matrix of occupied orbitals projected onto $\text{span}(A)$, the space of selected target atomic orbitals:

$$[\mathbf{S}^A]_{ij} = \langle i|\hat{P}|j\rangle, \quad (3)$$

where i, j are occupied orbital indices. Next we compute the $N_{\text{occ}} \times N_{\text{occ}}$ (unitary) matrix of eigenvectors $[\mathbf{U}]_{ij}$ of \mathbf{S}^A , such that

$$\mathbf{S}^A \mathbf{U} = \mathbf{U} \text{diag}(\sigma_1, \dots, \sigma_{N_{\text{occ}}}) \quad (4)$$

(where $\text{diag}(\dots)$ denotes a diagonal matrix of the given elements), or, written in component form,

$$\forall i, j: \sum_k [\mathbf{S}^A]_{ik} [\mathbf{U}]_{kj} = [\mathbf{U}]_{ij} \sigma_j. \quad (5)$$

There are at most $|A|$ non-zero eigenvalues $\{\sigma_i\}$, because $\text{span}(A)$ is a $|A|$ -dimensional space, and $[\mathbf{S}^A]_{ij}$ involves a projection onto it. Furthermore, the eigenvectors $[\mathbf{U}]_{ij}$ of \mathbf{S}^A define a rotation on

the occupied orbitals:

$$|i\rangle \mapsto |\tilde{i}\rangle = \sum_k |k\rangle [\mathbf{U}]_{ki}, \quad (6)$$

which clearly separates them into two groups: The at most $|A|$ rotated occupied orbitals $|\tilde{i}\rangle$ with $\sigma_i \neq 0$, which have non-vanishing overlap with our target atomic orbitals (and which therefore should go into the active space), and the remaining $|\tilde{i}\rangle$ with $\sigma_i = 0$ which have *no* overlap with our target atomic orbitals, and therefore can stay as inactive (inner closed shell) orbitals in the subsequent multiconfigurational methods.

Note that the rotated occupied orbitals $\{|\tilde{i}\rangle\}$ in (6) are obtained as a unitary transformation of $|\Phi\rangle$'s original occupied orbitals $\{|i\rangle\}$. Consequently, a determinantal wave function $|\tilde{\Phi}\rangle$ built from the $\{|\tilde{i}\rangle\}$ is physically equivalent (differs by at most a phase factor) from the original determinant $|\Phi\rangle$. That is, so far we have done nothing to $|\Phi\rangle$ except for splitting its occupied orbitals into a convenient set of at most $|A|$ orbitals related to our $|A|$ target AOs and the remaining set we can treat as inactive.

We then proceed similarly for the virtual orbitals $\{|a\rangle, 1 = 1 \dots, N_{\text{vir}}\}$ of $|\Phi\rangle$: We form the $N_{\text{vir}} \times N_{\text{vir}}$ projected overlap matrix

$$[\bar{\mathbf{S}}]_{ab}^A = \langle a | \hat{P} | b \rangle, \quad (7)$$

where a, b are virtual orbital indices, then find its unitary matrix of eigenvectors $\bar{\mathbf{U}}$ such that

$$\bar{\mathbf{S}}^A \bar{\mathbf{U}} = \bar{\mathbf{U}} \text{diag}(\sigma_1, \dots, \sigma_{N_{\text{vir}}}) \Leftrightarrow \forall a, b : \sum_c [\bar{\mathbf{S}}^A]_{ac} [\bar{\mathbf{U}}]_{cb} = [\bar{\mathbf{U}}]_{ab} \sigma_b, \quad (8)$$

and use $\bar{\mathbf{U}}$ to rotate the virtual orbitals via

$$|a\rangle \mapsto |\tilde{a}\rangle = \sum_c |c\rangle [\bar{\mathbf{U}}]_{ca}. \quad (9)$$

Again, the at most $|A|$ of the new virtual orbitals $\{|\tilde{a}\rangle\}$ with eigenvalues $\sigma_a \neq 0$ are selected for the active space, while the remaining orbitals will stay unoccupied in the subsequent multi-configuration treatment.

Finally, having active orbitals with overlap with A from both sides, occupied and unoccupied orbitals in $|\Phi\rangle$, we can form the total active space by combining the sets of $\{|\tilde{i}\rangle\}$ and $\{|\tilde{a}\rangle\}$ with non-zero projected overlap eigenvalues. Since the combined set includes *all* orbitals which have non-vanishing overlap with our target space $\text{span}(A)$, all of the selected AOs in A then lie within the span of this active space. Therefore, a multi-configurational wave function with this active space will be capable of representing arbitrary quantum resonances involving the target AOs—which was the goal of our construction.

2.3 Technical details of the construction

Sec. 2.2 discusses the formal framework of the active space construction. However, several practical aspects still need to be discussed: (a) How are the target AOs $A = \{|p\rangle\}$ chosen and represented? (b) How are the actual rotation matrices \mathbf{U} (eq. (6)) and $\bar{\mathbf{U}}$ (eq. (9)) computed in practice? (c) Can the active space (formally twice the number of the target AOs) be further reduced in size? (d) How should open-shell systems be handled? (In particular, what to do for restricted open-shell functions $|\Phi\rangle$?) We will discuss these questions in the current and next subsections.

Let us first assume that $|\Phi\rangle$ is a closed-shell Slater determinant obtained from an SCF calculation. Its occupied and virtual molecular orbitals are expressed as:

$$\begin{aligned} |i\rangle &= \sum_{\mu \in B_1} |\mu\rangle C_i^\mu \\ |a\rangle &= \sum_{\mu \in B_1} |\mu\rangle \bar{C}_a^\mu, \end{aligned} \quad (10)$$

where μ are basis functions from the (large) computational basis set B_1 (e.g., cc-pVTZ or def2-TZVPP), and $C_i^\mu = [\mathbf{C}_{\text{occ}}]_{\mu i}$ and $\bar{C}_a^\mu = [\mathbf{C}_{\text{vir}}]_{\mu a}$ are the coefficients of the basis function μ in the expansion of the occupied orbital i and virtual orbital a , respectively. \mathbf{C}_{occ} and \mathbf{C}_{vir} denote the

$|B_1| \times N_{\text{occ}}$ occupied and $|B_1| \times N_{\text{vir}}$ virtual sub-matrices of the $|B_1| \times |B_1|$ SCF orbital matrix \mathbf{C} (note that $N_{\text{occ}} + N_{\text{vir}} = |B_1|$ —each orbital is either occupied or virtual).

In general, computational basis sets such as B_1 do not contain basis functions directly corresponding to AOs of any sort. For this reason, we here select our target AOs $A = \{|p\rangle\}$ based on a second auxiliary basis set B_2 . This is a minimal basis set of tabulated free-atom AOs (here MINAO is used;⁴⁷ but other choices, such as subsets of ANO-RCC,^{58,59} ano-pVnZ,⁶⁰ or ANO-VT-XZ,⁶¹ could be considered; see also Sec. 5.1). This choice leads to simple expressions for the projected overlap matrices

$$S_{ij}^A = \langle i|\hat{P}|j\rangle = \sum_{\mu\mu'} C_i^\mu P_{\mu\mu'} C_j^{\mu'} \quad (11)$$

$$\bar{S}_{ab}^A = \langle a|\hat{P}|b\rangle = \sum_{\mu\mu'} \bar{C}_a^\mu P_{\mu\mu'} \bar{C}_b^{\mu'}, \quad (12)$$

where the matrix elements of the projector are

$$P_{\mu\mu'} = \sum_{pp'} \langle \mu|p\rangle [\sigma^{-1}]_{pp'} \langle p'|\mu'\rangle. \quad (13)$$

Combining all formulas into a numerical algorithm, the rotated orbitals are constructed as follows:

- Let $A \subset B_2$ denote the subset of AOs we choose as target AOs for the active space construction (for example, the five $3d$ AOs in a transition metal complex with one metal center).
- Form the overlap matrix σ with elements $\sigma_{pp'} = \langle p|p'\rangle$, where $p, p' \in A$, as well as σ 's inverse matrix, with elements $\sigma^{pp'} = [\sigma^{-1}]_{pp'}$. Both matrices have dimension of $|A| \times |A|$.
- Form the overlap matrix \mathbf{S}_{21} between the functions p of $A \subset B_2$ and the functions μ of the large basis set B_1 , with elements $[\mathbf{S}_{21}]_{p\mu} = \langle p|\mu\rangle$.
- Form the projector $P_{\mu\mu'} = \sum_{pp'} \langle \mu|p\rangle \sigma^{pp'} \langle p'|\mu'\rangle$, or $\mathbf{P} = \mathbf{S}_{21}^\dagger \sigma^{-1} \mathbf{S}_{21}$.
- Form the projected overlap matrices $\mathbf{S}^A = \mathbf{C}_{\text{occ}}^\dagger \mathbf{P} \mathbf{C}_{\text{occ}}$ for the occupied orbitals (eq. (11)), and $\bar{\mathbf{S}}^A = \mathbf{C}_{\text{vir}}^\dagger \mathbf{P} \mathbf{C}_{\text{vir}}$ for the virtual orbitals (eq. (12)).

- Finally, diagonalize both projected overlap matrices to obtain the transformation matrices separating the MO sets by overlap with $\text{span}(A)$. Concretely, diagonalize \mathbf{S}^A to obtain the eigenvector matrix \mathbf{U} , and use it to find the transformed occupied orbital matrix $\tilde{\mathbf{C}}_{\text{occ}} = \mathbf{C}_{\text{occ}}\mathbf{U}$. Then diagonalize $\tilde{\mathbf{S}}^A$ to obtain the eigenvector matrix $\tilde{\mathbf{U}}$, and use it to find the transformed virtual orbital matrix $\tilde{\mathbf{C}}_{\text{vir}} = \mathbf{C}_{\text{vir}}\tilde{\mathbf{U}}$. These are the expansion coefficients of $\{|\tilde{i}\rangle\}$ (eq. (6)) and $\{|\tilde{a}\rangle\}$ (eq. (9)), respectively.

Rather than using the minimal basis B_2 directly, one could consider choosing the target AOs from a set of polarized AOs which take the molecular environment into account, such as the Intrinsic Atomic Orbitals (IAOs).⁴⁷ If the IAOs are given as

$$|\tilde{p}\rangle = \sum_{\mu} |\mu\rangle T_{\mu p}, \quad (14)$$

where $T_{\mu p}$ denotes the elements of the $|B_1| \times |B_2|$ IAO transformation matrix,⁴⁷ this can be incorporated by updating the projection matrix in S_{ij}^A and S_{ab}^A in eqs. (11) and (12) as

$$\mathbf{P} = \mathbf{S}\mathbf{T}(\mathbf{T}^\dagger\mathbf{S}\mathbf{T})^{-1}\mathbf{T}^\dagger\mathbf{S}. \quad (15)$$

For simplicity, we do not follow this approach in this work, and choose the target AOs directly from the minimal basis B_2 as described above.

Additionally, in some cases one may wish to extend the target space by some AOs which lie outside the valence space. This scenario is discussed in Sec. 5.1.

2.4 Truncating the active space

The eigenvalues σ_i and σ_a of the projected overlap matrices in eqs. (6) and (9) reflect the degree to which the transformed orbitals $|\tilde{i}\rangle$ and $|\tilde{a}\rangle$ overlap with the space of our target AOs. If we include every such transformed orbital with $\sigma_a \neq 0$ and $\sigma_i \neq 0$ into our active space, the resulting CAS space will exactly include all electronic configurations which can be formed over the given AOs

1
2
3 and the maximum size of the CAS space is twice that of the target AO space. However, this CAS
4 space may be too large and we may need to truncate it. Here we can use the fact that often many of
5 the σ_i and σ_a are small. As a practical measure, we can set a threshold, such as 0.05–0.1, to exclude
6 MOs with negligible overlap with $\text{span}(A)$. In addition to reducing the size of the active space, this
7 can further improve the reproducibility of calculations in the case of very small eigenvalues. This
8 threshold becomes the only numerical parameter to be chosen by the user, and together with the
9 selection of target AOs and the type of the SCF wave function $|\Phi\rangle$, it fully determines the active
10 space.
11

12
13 Of course, if truncation is used, the active space no longer captures all possible configurations
14 involving the target AOs which can be formed. However, the truncation does *not* affect the quality
15 of the description of the rest of the molecule by the core determinant in Eq. (1). This guarantees
16 that the CASCI energy lies below the variational HF energy. We can also imagine the opposite
17 tradeoff, where one obtains a truncated active space which retains the ability to capture all possible
18 configurations involving the target AOs, at the cost of worsening the quality of the core determinant
19 which describes the rest of the molecule. (In DMET language, this would correspond to truncating
20 the “bath” orbitals, which is considered in Refs. 52 and 57). However, this may be a worse
21 truncation procedure in the current setting, as the energy gained by treating the fluctuations in
22 occupation number in the target AOs (such as a TM 3d shell) may not make up for the energy lost
23 in incompletely describing the mean-field hybridization between the target AOs and the rest of the
24 molecule. In particular, this second bath truncation procedure can, in principle, lead to a CASCI
25 energy above the variational HF energy.
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

48 2.5 Treatment of open-shell systems

49 If $|\Phi\rangle$ is a closed-shell determinant, then the active space construction algorithm can be directly
50 used as described in Sec. 2.3. However, in the case of open shell determinants, several choices can
51 be considered:
52
53
54
55

- 56 1. One may perform the algorithm separately for alpha and beta orbitals, thus creating active
57
58
59
60

1
2
3 orbitals with different spatial parts for alpha and beta-spin electrons. While this choice is the
4 most straight-forward and, arguably, creates the best initial active orbitals in the open-shell
5 case, this option is not directly feasible if a spin-adapted multiconfigurational calculation will
6 follow—most existing MCSCF programs cannot use such unrestricted orbitals (although this
7 is implemented in the code we use here⁶²). A possible remedy for this problem would be to
8 construct a single set of “corresponding orbitals”⁶³ from the separate alpha- and beta-orbital
9 sets, but this has not been tested here.

- 10
11
12
13
14
15
16
17
18
19 2. One may use *exclusively* the alpha orbitals to construct the active space orbitals (and inactive
20 orbitals determining the core determinant). This treatment can be applied to both restricted
21 and unrestricted SCF functions $|\Phi\rangle$ in a simple manner. The rationale for this is that in the
22 restricted open-shell case, the occupied beta orbitals lie entirely within the linear span of
23 the occupied alpha orbitals, so one can argue that this choice takes care of both spin cases.
24 However, this argument is somewhat misleading because it may lead to some *unoccupied*
25 beta orbitals being transformed into the core space, therefore enforcing their occupation
26 with two electrons. This error in the core means that the CASCI energy may be higher than
27 the variational HF energy. If the singly-occupied MOs in $|\Phi\rangle$ have only small components
28 on the target valence AOs, this can lead to very bad CASCI wavefunctions.
- 29
30
31
32
33
34
35
36
37
38
39
40 3. If a ROHF determinant $|\Phi\rangle$ is used, one may apply the construction of Sec. 2.3 exclusively
41 to the doubly-occupied and fully unoccupied orbitals of $|\Phi\rangle$, to form the core determinant
42 and initial part of the active space, and then include *additionally* all the singly-occupied
43 orbitals of $|\Phi\rangle$ into the active space. The CASCI energy is then guaranteed to be below the
44 variational HF energy, and further spin-adaptation can be used. The main drawback is that
45 the active space is usually larger in this procedure.

46
47
48
49
50
51
52
53 By default, we use method 2, however, we compare the different schemes in one of the systems
54 below.
55
56
57
58
59
60

3 Computational Details

We implemented the atomic valence active space (AVAS) construction within the PYSCF⁶² package. All Complete Active Space Self-Consistent Field (CASSCF), Complete Active Space Configuration Interaction (CASCI), and strongly contracted N -electron valence state perturbation theory (NEVPT2)^{64–67} calculations were carried out using PYSCF. For active spaces with more than 16–17 orbitals we used the BLOCK code⁶⁸ through the PYSCF-BLOCK interface to perform DMRG calculations in the active space. The AVAS construction is also being implemented in a development version of Molpro.

For simplicity, we used all-electron cc-pVTZ-DK^{69,70} basis sets for all systems, apart from the Fenton reaction (*vide infra*). For the auxiliary minimal basis B_2 used to choose the target AOs, we employed the MINAO basis,⁴⁷ which is a truncated subset of the cc-pVTZ basis; for most atoms, this set consists of spherically averaged ground-state Hartree-Fock orbitals for the free atoms. We did not use point group symmetry in the present calculations, since in a straight-forward implementation, the MOs do not necessarily retain symmetry-adaptation after rotation; however, if symmetry-adaptable sets of target AOs are chosen, symmetry respecting orbital rotations can in principle be constructed. For completeness, scalar relativistic effects were included using the exact-two-component (X2C) approach,^{71,72} but this did not lead to significant differences from the non-relativistic calculations in any of the considered examples. Spin-orbit coupling was not considered.

For the case of the Fenton reaction, the geometries were optimized with the DSCF and GRAD modules of Turbomole 7.0. This was done at the level of symmetry-broken unrestricted B3LYP⁷³ with def2-TZVP basis sets,⁷⁴ starting from the structures provided in Ref. 75. The RI approximation was not employed, and solvation effects were not considered. The characters of the starting geometry, transition state geometry, and product geometry were confirmed by computing the analytic nuclear Hessians at these points, via the AOFORCE module.⁷⁶ The reaction path was computed by tracing the intrinsic reaction coordinate⁷⁷ in both directions, starting at the transition state geometry. This was done using Turbomole's DRC module. Finally, the structures of the starting point,

1
2
3 transition state, product, and IRC segments (of both directions) were joined, aligned, and com-
4 pressed using the development version of IboView.^{78,79} Employing these geometries, the reported
5 multi-configuration AVAS calculations were performed with PYSCF, using cc-pVTZ orbital basis
6 sets and a non-relativistic Hamiltonian.
7
8
9

10
11 The used molecular geometries are supplied in the supporting information, as are selected
12 visualizations of the obtained active spaces. Orbital visualizations were made with IboView,^{78,79}
13 and show iso-surfaces enclosing 80% of the orbital's electron density.
14
15
16
17
18
19

20 4 Results and Discussion

21
22 Judging the quality of an initial active space is difficult. Here we employ two complementary
23 criteria:
24
25
26

- 27
28 1. In the course of a CASSCF calculation, the active space orbitals are optimized. If the overlap
29 of the optimized active space with our initial active space guess remains high, we here regard
30 our active space guess as being of high quality. To quantify this aspect, we compute the
31 $N_{\text{act}} \times N_{\text{act}}$ overlap matrix
32
33
34
35
36
37

$$38 \mathbf{S}_{\text{change}} = (\mathbf{C}_{\text{act}}^{\text{final}})^{\dagger} \mathbf{S}(\mathbf{C}_{\text{act}}^{\text{initial}}) \quad (16)$$

39
40
41 between the initial guess and the optimized final active orbitals, and compute its singular
42 value decomposition. If all singular values are close to 1.0, the active space remains mostly
43 unchanged during the optimization. On the contrary, each singular value close to 0.0 in-
44 dicates that an initial active orbital had to be completely replaced by an unrelated orbital.
45 The latter case may indicate that the initial active space lacks the capability to represent
46 some essential features of the strongly correlated electronic structure of the given molecule
47 (since the optimized active orbital, which presumably is needed to represent energetically
48 important electron configurations of the system, lies completely outside the linear span of
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 the initial active space guess); the initial active space guess may therefore need to be changed
4
5 or enlarged.
6
7

- 8
9 2. We compare our computed excitation energies to experimental results and other high-level
10
11 calculations reported in the literature. When constructing a well-behaved series of active
12
13 spaces, we should be able to see convergence or stability of the computed properties with
14
15 respect to the active space size.
16

17 We note that, strictly speaking, neither criterion can establish an initial active space as “*definitely*
18
19 *good*”: While criterion 1 tests whether the CASSCF-optimized active space is close to the initial
20
21 active space, there is no formal guarantee that the optimized CASSCF wave function itself is
22
23 the best representative of the sought after electronic state (e.g., the optimization algorithm could
24
25 be stuck in an undesirable local minimum). While criterion 2 establishes the compatibility with
26
27 some specific physical properties and reference systems, it cannot guarantee that no other physical
28
29 properties or systems exist which need different active spaces for a proper description. However,
30
31 both criteria *could* establish our active space construction *to be unfit* for its target applications—
32
33 by showing counter-examples and negative results. In order to investigate the basic properties
34
35 of the AVAS procedure, we now apply these criteria in calculations on various transition-metal
36
37 complexes.
38
39
40

41 **A. Ferrocene**

42
43
44 We begin by considering the electronic structure of ferrocene, $\text{Fe}(\text{C}_5\text{H}_5)_2$. The ground state of
45
46 ferrocene is dominated by a single configuration with Fe d^6 . Both the MO analysis in Ref. 80 and
47
48 our CI expansion coefficients indicate that the lowest excited states of ferrocene have significant
49
50 multiconfigurational character.
51

52
53 We carried out an initial restricted Hartree-Fock (ROHF) calculation for the singlet ground
54
55 state. We used the optimized D_{5h} geometry from Ref. 81 (using the cc-pw-CVTZ basis set at the
56
57 CCSD(T) level). In this geometry, the two cyclopentadienyl (Cp) rings are planar and the z-axis is
58
59
60

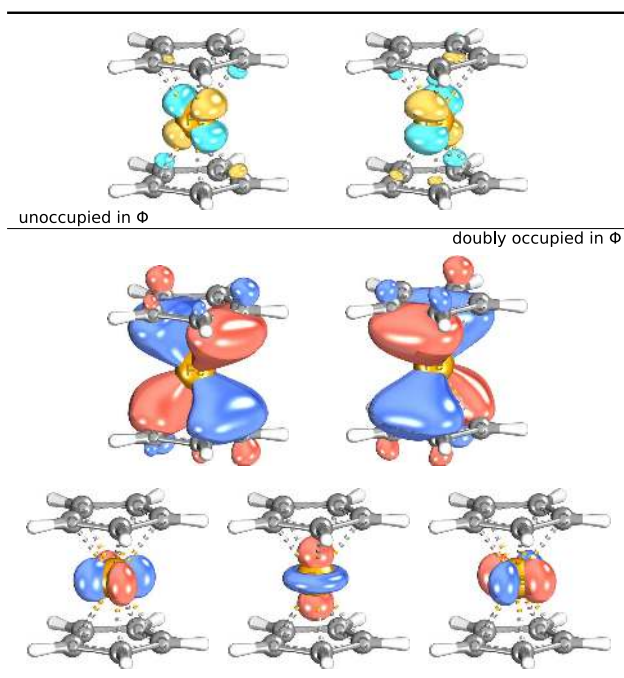


Figure 1: Ferrocene: The (10e, 7o) AVAS initial active space for the five $3d$ orbitals of the iron atom. It is generated with threshold 0.1, based on the RHF wave function $|\Phi\rangle$ of the singlet ground state of the ferrocene molecule. After the AVAS transformation, all other occupied orbitals have exactly zero percent d -character, and all other unoccupied orbitals have 2.7% Fe $3d$ -character or less (see text).

aligned with the Cp-Fe-Cp axis.

As our target set of AOs, we first chose the five $3d$ orbitals of Fe. Using a threshold of 0.1, the AVAS scheme produces a seven orbital active space: five orbitals from the occupied orbital space and two orbitals from the unoccupied orbital space are combined into a (10e, 7o) active space; the orbitals are visualized in Fig. 1. The five overlap eigenvalues above the threshold from the occupied space are 0.325, 0.325, 0.973, 0.973, 0.995, and the two eigenvalues from the unoccupied space are 0.675, 0.675. There are three omitted unoccupied orbitals: two of them have only 2.7% weight in the $3d$ AO space and $(3d_{xy}, 3d_{x^2-y^2})$ character, and one has 0.6% weight in the $3d$ AO space, corresponding to the $3d_{z^2}$ AO. The two orbitals from the occupied orbital space with eigenvalues 0.325 and 0.325, as well as the two selected MOs from the unoccupied orbital space with eigenvalues 0.675, 0.675, both have $(3d_{xz}, 3d_{yz})$ character; based on this, we conclude that $(3d_{xz}, 3d_{yz})$ are most strongly involved in bonding with the two Cp rings, in agreement with the

1
2
3 bonding picture established in earlier studies.^{80,82}

4
5 To construct a second, larger active space, we next included the ten p_z orbitals of the carbon
6 atoms in the Cp rings into the target AO list. The AVAS construction then yields 15 orbitals
7 above the 0.1 overlap threshold, giving an (18e,15o) active space: 9 of occupied character and 6
8 of unoccupied character. Of the 9 occupied orbitals, four do not carry Fe 3d character, instead
9 belonging to the π -system of the Cp rings; two more have strong mixing with the Fe ($3d_{xz}, 3d_{yz}$)
10 orbitals, and the final three represent almost-pure Fe $3d_{xy}$, $3d_{x^2-y^2}$, and $3d_{z^2}$ atomic orbitals (with
11 $\geq 95\%$ orbital weight on the Fe 3d AOs). Among six unoccupied orbitals, two have no Fe 3d
12 character, two have predominantly Fe ($3d_{xz}, 3d_{yz}$) character, and two have $< 5\%$ character of Fe
13 $3d_{xy}$ and $3d_{x^2-y^2}$ orbitals, respectively. This agrees with our observations for the smaller (10e,7o)
14 active space; namely, the additional orbitals in the (18e,15o) space are practically non-bonding in
15 character. Lowering the AVAS threshold from 0.1 to 0.05 gives a (22e,17o) active space, which
16 includes some further effectively non-bonding orbitals beyond the (18e,15o) active space.
17
18
19
20
21
22
23
24
25
26
27
28
29

30 We calculated the ground state and singlet and triplet excited states of ferrocene with these
31 active spaces at the CASCI and CASSCF levels. Previous studies⁸⁰ indicate that there are three
32 low-lying $d \rightarrow d$ singlet transitions ($1^1E_2'', 1^1E_1'', 2^1E_1''$) and three low-lying $d \rightarrow d$ triplet transi-
33 tions ($1^3E_1'', 1^3E_2'', 2^3E_1''$). These $d \rightarrow d$ transitions describe excitations from the three non-bonding
34 orbitals (predominantly of $3d_{x^2-y^2}$, $3d_{xy}$ and $3d_{z^2}$ metal character, as described above) to the two
35 antibonding orbitals having mostly ($3d_{xz}, 3d_{yz}$) metal character. All these excited states have multi-
36 configurational (but single-excitation) character. Note that the E_1'' and E_2'' states are doubly degen-
37 erate, thus there are 6 low-lying singlet and triplet excited states. In the CASSCF calculations we
38 therefore state-averaged over 7 roots and 6 roots in the singlet and triplet manifolds, respectively.
39
40
41
42
43
44
45
46
47

48 Table 1 displays the excitation energies, compared to experimental and theoretical numbers
49 from the literature, including singly excited configuration interaction (SECI),⁸³ symmetry adapted
50 cluster configuration interaction (SAC-CI)⁸⁰ and time-dependent density functional theory (TD-
51 DFT) calculations.⁸⁴ Table 2 compares the impact of using different active spaces with different
52 methods.
53
54
55
56
57
58
59
60

1
2
3
4 The performance of CASCI for the smallest (10e,7o) active space is reasonable for most of
5 the excited states, with the $2^1E_1''$ and $2^3E_1''$ states being exceptions. These two states have some
6 Rydberg character,⁸⁰ and therefore the valence CASCI overestimates these transitions by about 2
7 eV; this effect can also be seen in the errors of the SECI energies. Averaging over all the states
8 in the CASSCF seemed to spread the error over the states, lowering all the energies. An accurate
9 description of the differential correlation in the $2^1E_1''$ and $2^3E_1''$ states thus requires a dynamic
10 correlation treatment. We find excellent agreement for all states at the CASSCF+NEVPT2 level,
11 with a largest error of only 0.21 eV.

12
13 Using the larger (18e,15o) active space, which includes the ligand π -orbitals, worsens the
14 CASSCF excitation energies (except for the $2^1E_1''$ and $2^3E_1''$ states). However, incorporating dy-
15 namic correlation through NEVPT2 rebalances the states, improving CASCI and CASSCF ener-
16 gies (except for $1^1E_2''$ and $1^3E_1''$) and yielding better agreement with experiment and the (10e,7o)
17 CASSCF+NEVPT2 excitation energies. If we further go from the (18e,15o) active space to the
18 (22e,17o) active space, by additionally including the truncated orbitals with eigenvalues below
19 0.1, the excitation energies change by less than 0.01 eV. Together, these observations indicate that
20 the multiconfigurational character is already well converged in the smallest (10e,7o) active space.

21
22 As discussed above, a second test of the quality of the AVAS active space is provided by the
23 SVD decomposition of the overlap between the CASSCF-optimized active space and the active
24 space initial guess. In the case of ideal coincidence, the SVD eigenvalues should be equal to 1.
25 The smallest SVD eigenvalues we found were 0.927 for the (10e,7o) active space and 0.906 for
26 the (18e,15o) active space, respectively. This indicates that here the AVAS provides a stable and
27 accurate initial guess for the CASSCF procedure.

28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 **B. [Fe(NO)(CO)₃]⁻**

52
53 We next consider the complex anion [Fe(NO)(CO)₃]⁻, which exhibits catalytic activity in a range
54 of organic reactions, and has been extensively characterized both theoretically and experimentally
55 (see Ref. 87 and references therein). The complex features three-center bonds along both the
56
57
58
59
60

Table 1: Lowest singlet and triplet excitation energies of ferrocene: results from the (10e,7o) active space compared to other electronic structure methods. Energies are given in eV.

State	CASCI (10e,7o)	CASSCF (10e,7o)	CASSCF+NEVPT2 (10e,7o)	SAC-CI ⁸⁰	SECI ⁸³	TD-DFT ⁸⁴ PBE	Expt.*
$1^1E_2''$	2.85	2.10	2.79	2.11	2.63	2.90	2.8, ⁸⁵ 2.7 ⁸⁶
$1^1E_1''$	3.33	2.17	2.87	2.27	3.31	3.03	2.81, ⁸⁵ 2.98 ⁸⁶
$2^1E_1''$	5.82	4.07	3.99	4.03	5.74	3.60	3.82 ^{85,86}
$1^3E_1''$	1.81	0.97	1.88	1.40	1.81-1.87		1.74 ⁸⁵
$1^3E_2''$	1.84	1.07	2.03	1.68	1.81-1.87		2.05 ⁸⁵
$2^3E_1''$	4.26	2.25	2.50	2.60	4.56		2.29-2.34 ⁸⁵

* From electronic absorption spectra of ferrocene in EPA glass, formed by cooling a EPA solution of ferrocene to liquid nitrogen temperature.

Table 2: Lowest singlet and triplet excitation energies of ferrocene: comparison of active spaces. Energies are given in eV.

	CASCI		CASCI+NEVPT2		CASSCF		CASSCF+NEVPT2		Expt.*
	(10e,7o)	(18e,15o)	(10e,7o)	(18e,15o)	(10e,7o)	(18e,15o)	(10e,7o)	(18e,15o)	
$1^1E_2''$	2.85	3.51	3.10	3.20	2.10	2.88	2.79	2.93	2.8, ⁸⁵ 2.7 ⁸⁶
$1^1E_1''$	3.33	3.90	3.20	3.21	2.17	3.45	2.87	3.11	2.81, ⁸⁵ 2.98 ⁸⁶
$2^1E_1''$	5.82	5.81	4.24	4.00	4.07	4.61	3.99	4.06	3.82 ^{85,86}
$1^3E_1''$	1.81	2.46	2.14	2.25	0.97	1.91	1.88	2.09	1.74 ⁸⁵
$1^3E_2''$	1.84	2.48	2.25	2.34	1.07	2.37	2.03	2.26	2.05 ⁸⁵
$2^3E_1''$	4.26	4.33	2.78	2.54	2.24	2.64	2.50	2.50	2.29-2.34 ⁸⁵

* See note in Tab. 1.

Fe–N–O axis and between Fe and each pair of CO ligands;⁸⁷ its catalytic mode of action exhibits a highly unusual nitrosyl-ligand based oxidation in some cases,⁸⁸ and response to photo-activation in other cases.⁸⁹ Analysis of the ground-state CASSCF wavefunction and natural orbital occupations indicates that it has some multiconfigurational character, and that it should be thought of as a Fe(0) species bound via two covalent π -bonds to the $[\text{NO}^-]$.⁸⁷

As in the previous example, we started with a RHF calculation for the singlet ground state, and for the simplest active space we chose five 3d orbitals of Fe as the target AOs. We used the geometry of Ref. 89. Using an overlap threshold of 0.1 gives rise to five occupied orbitals and three unoccupied orbitals for the active space; two unoccupied orbitals with only 6% weight in the 3d orbital space lie below the threshold for active space inclusion. Of the three unoccupied orbitals included in the active space, two have ($3d_{yz}$, $3d_{xy}$) and ($3d_{xz}$, $3d_{x^2-y^2}$) character, respectively,

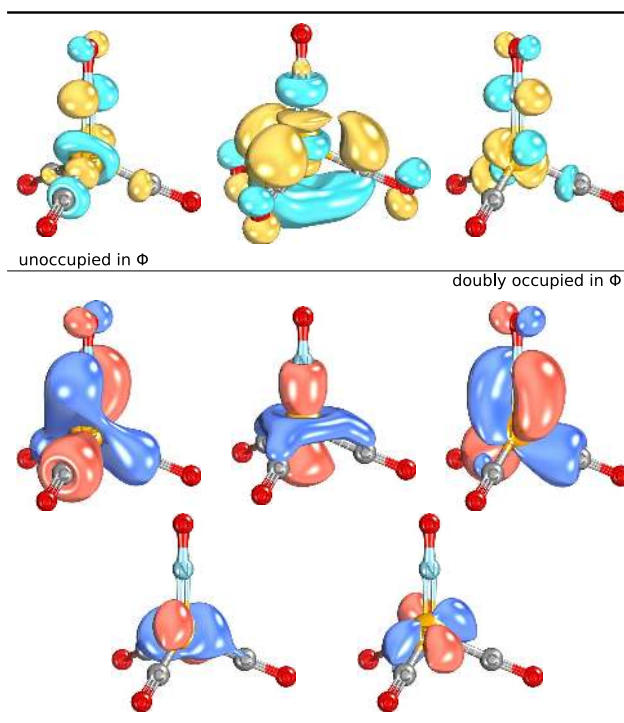


Figure 2: $[\text{Fe}(\text{NO})(\text{CO})_3]^-$: The (10e,8o) AVAS initial active space for the five $3d$ orbitals of the iron atom, generated with threshold 0.1 from the ground state singlet RHF wave function. Two virtual orbitals with 6% d -character lie below the 10% threshold and are omitted from the active space (see text); all other molecular orbitals are completely free of Fe $3d$ -character after the AVAS transformation.

while the third one has mostly $3d_{z^2}$ character. The resulting active space is visualized in Fig. 2. Additionally including the nitrogen $2p$ orbitals into the set of target AOs (with the same threshold) gives a (16e,14o) active space; this adds three MOs with $2p$ character, involved in the three-center Fe–N–O bonds, and three involved in N–Fe–CO type bonds.

Unfortunately, there is no gas phase experimental excitation data for this system. However, theoretical vertical excitation energies from state-averaged CASSCF calculations followed by MRCI+Q, using the def2-TZVPP basis set (omitting g -functions) have previously been reported,⁸⁹ which we can compare against. We used CASSCF and NEVPT2 to compute the vertical transition energies averaging over five singlet and four triplet states, as in Ref. 89. The smallest SVD eigenvalue for the active space overlap with the initial guess for the (10e,8o) active space is 0.806, indicating that AVAS provides a good initial guess. For the (16e,14o) active space, the lowest SVD eigenvalue decreases to 0.652—apparently adding only the nitrogen $2p$ orbitals, without also adding the car-

1
2
3 bon $2p$ orbitals, leads to a less balanced active space compared to the (10e,8o) AVAS initial active
4 space.
5
6

7 The vertical excitation energies from CASSCF with the (10e,8o) active space are in better
8 agreement with the CASSCF/MRCI+Q excitation energies than with CASSCF results from Ref.
9 89. This indicates that the (10e,8o) active space constructed using AVAS provides a more balanced
10 description of electron correlation than the larger active spaces.
11
12
13
14

15 However, the NEVPT2 dynamical correlation treatment significantly raises the obtained ex-
16 citations energies above the MRCI+Q values. Similarly, CASSCF calculations with the larger
17 (16e,14o) active space also yield significantly higher excitation energies than with the (10e,8o)
18 active space, and the CASSCF+NEVPT2 excitation energies in this larger space are also fairly dif-
19 ferent from the values obtained from the (10e,8o) active space. This lack of stability with respect
20 to the active space size indicates that the excited states are not benign electronically, and that their
21 accurate description requires a more sophisticated dynamic correlation treatment beyond 2nd order
22 perturbation theory. This is supported by the MRCI+Q study in Ref. 89, where the (empirical)
23 Q-contribution to the excitation energy is as large as 0.2 eV.
24
25
26
27
28
29
30
31
32
33

34 To further substantiate these claims, we also computed CASSCF+NEVPT2 results for the same
35 manually selected (14e,9o) initial active space as described in Ref. 89 (in these calculations, the
36 initial 14 active orbitals were manually selected for Fe d and NO π and π^* -character by visual
37 inspection of KS-DFT/PBE orbitals computed with the def2-TZVPP basis set; the CASSCF ex-
38 citation energies thus obtained reproduced the results reported in the supporting information of
39 Ref. 89 with better than 0.01 eV accuracy). By comparing the results of (our) NEVPT2 and (the
40 referenced) MRCI+Q for the same active space of Ref. 89, we can separate the effect of the dy-
41 namic correlation treatment from the quality of the active space. We see that CASSCF+NEVPT2
42 calculations performed with our automatically constructed (10e,8o) active space and the manually
43 selected (14e,9o) active space show a fair agreement in the case of singlet excited states; however,
44 the CASSCF+NEVPT2 method overestimates the energies of the triplet excited states with the
45 (14e,9o) active space by 0.5–0.75 eV more than with the (10e,8o) active space, compared to the
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

CASSCF/MRCI+Q transition energies. Combined, these facts strongly suggest that the approximate NEVPT2 correlation treatment is the primary cause of deviation from the MRCI+Q reference values, rather than our automatically constructed active space, and that the smaller AVAS provides a more balanced description of this system.

Table 3: Vertical excitation energies from the ground state to the lowest singlet and triplet states of $[\text{Fe}(\text{NO})(\text{CO})_3]^-$. Energies are given in eV.

State	CASSCF		CASSCF+NEVPT2		CASSCF ⁸⁹	CASSCF/MRCI+Q ⁸⁹	CASSCF+NEVPT2
	(10e,8o)	(16e,14o)	(10e,8o)	(16e,14o)	(14e,9o)*	(14e,9o)*	(14e,9o)*
¹ A ₂	3.14	4.37	4.04	3.76	2.78	3.26	4.13
¹ E	3.58	4.86	4.29	3.98	3.18	3.53	4.29
¹ A ₁	3.54	5.17	4.33	4.25	3.22	3.64	4.40
³ A ₁	2.27	2.44	2.63	2.43	1.76	2.32	3.40
³ E	2.82	3.41	3.51	3.20	2.44	2.96	4.02
³ A ₂	2.89	4.04	3.80	3.82	2.56	3.15	4.39

* The (14e,9o) active space consisted of Fe *d* orbitals and the NO π and π^* (manually chosen according to Ref. 89).

C. FeO₄²⁻

As our next system, we consider the bare tetraoxoferrate (VI) ion, FeO₄²⁻. We assume a tetrahedral FeO₄²⁻ cluster with an Fe–O distance of 1.660 Å.

We started with a ROHF calculation for the ³A₂ ground state. Including only 3*d* orbitals into the target AO set gives a (8e,8o) active space. Three unoccupied MOs have 34.5% weight in the 3*d* orbital space, with 3*d*_{yz}, 3*d*_{xz}, and 3*d*_{x²-y²} character, but are mostly centred on the ligands; the occupied MOs have mostly metal character. This indicates that some of the low-lying excitations are charge-transfer excitations.

An earlier study⁹⁰ found that the ground and excited states cannot be described by a simple Ligand Field Theory *d*² model; rather, they contain superpositions of a large number of configurations, including ligand-to-metal excitations. From this it has been argued that it is insufficient to only consider molecular orbitals with Fe 3*d* character in the active space to describe excited states. Indeed, we find that CASSCF+NEVPT2 calculations with the (8e,8o) active space (generated only with the 3*d* orbitals in the target AO set) significantly overestimate the excited states energies, by

about $\approx 6500 \text{ cm}^{-1}$ ($\approx 0.81 \text{ eV}$), compared to experiment.

For this reason, we expanded the target AO list to the five $3d$ orbitals of Fe and $2p$ orbitals of all four O atoms. Using option 2 to transform the alpha orbitals, our scheme with the 0.1 overlap threshold produced 14 occupied orbitals and 3 unoccupied orbitals, resulting in a (26e,17o) active space.

We calculated the vertical excitation energies of FeO_4^{2-} , namely transition energies from the ground 3A_2 state to the first two excited states, 1E and 1A_1 (see Table 4), using CASCI and CASSCF (state-averaged over three singlet states and one triplet state) and CASCI+NEVPT2, comparing to previously reported RASSCF and experimental numbers. The CASCI calculations significantly overestimate the excitation energies; however, this is significantly improved by optimizing the orbitals using state-averaged CASSCF. The smallest SVD singular value for the active space overlap between the initial and optimized active orbitals is 0.968, indicating that our initial active orbitals provide a very good guess for the CASSCF procedure and only require a little relaxation to yield good agreement with experiment.

Including dynamic correlation by means of NEVPT2 on top of CASCI or CASSCF significantly improves the results. Note that the difference between our CASSCF excitation energies and those in Ref. 90 with 17 orbitals, obtained with the RASSCF method, reflects both the slightly different basis set as well as the truncated CI configuration space in RAS.

Table 4: Calculated and experimental low-lying excitation energies of FeO_4^{2-} . Energies are given in cm^{-1} .

State	CASCI (26e,17o)	CASSCF (26e,17o)	CASCI+NEVPT2 (26e,17o)	CASSCF+NEVPT2 (26e,17o)	RASSCF ⁹⁰ (26,4,0;12,5,0)	Expt.*
1E	9882, 9899	7462	6252, 6268	6548, 6550	6300	6209, 6219 ⁹¹ 6219, 6230 ⁹²
1A_1	14703	10710	9006	9471	9200	9119, ⁹¹ 9176 ⁹²

* From single-crystal polarized electronic absorption spectra.

D. VOCl_4^{2-}

We now consider the oxotetrachlorovanadate(IV) anion, VOCl_4^{2-} . We use a square pyramidal geometry for VOCl_4^{2-} , as in Ref. 93 (although we use a different orientation: the V atom is at the origin, the O atom is on the z axis above the x-y plane and the Cl atoms are below the x-y plane).

In this complex, vanadium is in a d^1 configuration. As in the next example, here the $d \rightarrow d$ excitations do not have much multiconfigurational character. However, it is important that multireference methods (and their active spaces) provide a balanced description of all states, not just multiconfigurational ones. The vanadium complex provides a system to test this in an early transition metal (single-reference) problem.

If we choose a set of five AOs, representing the five $3d$ orbitals of a vanadium atom, we obtain five occupied and four unoccupied MOs from the ROHF reference wavefunction using the AO-projector option 2. One of the occupied MOs is a non-bonding $3d_{xy}$ atomic orbital, while the other V $3d$ AOs strongly mix with the valence orbitals of the oxygen atom and four chlorine atoms. This results in four doubly occupied bonding MOs and four anti-bonding MOs which are unoccupied in the ground state. The two unoccupied MOs have 69.4% overlap with the $3d$ orbital space and carry ($3d_{xz}, 3d_{yz}$) character, other two have 54.0% and 70.3% overlap with the $3d$ orbital space and have ($3d_{z^2}$ and $3d_{x^2-y^2}$) character, respectively; all four unoccupied MOs have mostly metal character.

Using the (9e,9o) active space we calculated the lowest transitions, which are essentially $d \rightarrow d$ in nature. There are four possible ligand-field transitions from the highest non-bonding $3d_{xy}$ orbital to four unoccupied MOs, thus in CASSCF we averaged over five doublet states. Table 5 summarizes the low-energy vertical $d \rightarrow d$ energies. The CASSCF method with the small (9e,9o) active space gives an accurate 2B_1 state, but strongly overestimates the other excited states. Using NEVPT2 to treat the dynamic correlation on top of CASSCF significantly improves the excited state energies, resulting in a good agreement with the experimental values and with CASPT2 results obtained with a (11e,10o) space.⁹³ This (11e,11o) space is similar to ours, with the addition of the oxygen $2p$ shell.

We also construct a larger (33e,21o) active space, including the $2p$ orbitals of O and the $3p$

orbitals of Cl into the target AO list. In this larger space, to reduce computational cost, we used CASCI+NEVPT2 rather than CASSCF+NEVPT2. The excited states from CASCI+NEVPT2 with the (33e,21o) active space are also in excellent agreement with the experimental data. The stability of the CASCI/CASSCF+NEVPT2 excitations with respect to expanding the active space confirms that the correlation is well converged by all these treatments.

In the CASSCF calculation with the (9e,9o) active space, the smallest SVD eigenvalue for the active space overlap between the initial and the optimized active orbitals amounts to 0.821. This implies that, in this case, AVAS provides a reasonable, but not perfect, initial guess for CASSCF.

We also used this complex to test and compare options 2 and 3, as described in Sec. 2.5, for constructing the active space with a ROHF reference determinant $|\Phi\rangle$. The excitation energies, calculated with these two options, differ by less than 100 cm^{-1} ($\approx 0.01\text{ eV}$). However, VOCl_4^{2-} 's ground state has only one singly occupied orbital, and it is possible that larger differences will occur for systems with ground states of higher spin.

Table 5: Vertical excitation energies to the lowest doublet excited states from the 2B_2 state ground of VOCl_4^{2-} . Energies are given in cm^{-1} .

State	CASSCF (9e,9o)	CASSCF+NEVPT2 (9e,9o)	CASCI+NEVPT2 (33e,21o)	CASPT2 ⁹³ (11e,10o)	Exp.*
2B_1	11736	13044	12123	12667	11600 ⁹⁴
2E	18650	14537	13722	13620	13700 ⁹⁴
2A_1	33808	28586	26824		

* From single-crystal polarized electronic absorption spectrum

E. $[\text{CuCl}_4]^{2-}$

We finally consider the D_{4h} $[\text{CuCl}_4]^{2-}$ complex, with a Cu–Cl bond length of 2.291 Å, as in Ref. 93. As in the vanadium system, the $d \rightarrow d$ transitions are single-reference in character: this complex provides a late transition metal example.

Using the 3d AOs of Cu as the target AOs and a default cutoff of 0.1, we obtain only 5 occupied MOs and no unoccupied MOs. This result might surprise at first glance. However, in this case,

the antibonding orbitals have ligand character, and thus there are no unoccupied MOs having more than 5% 3d character. The lowest ligand-field transitions arise from the excitation of electrons from the doubly-occupied MOs with dominant 3d character to the singly occupied MOs with 3d character.

Despite the fact that the lowest transitions happen mostly *within* the 3d orbital space, such a small (9e, 5o) active space is insufficient to describe them at the CASSCF level—because the nearly filled space leaves no room for electron correlation. CASSCF+NEVPT2 however, provides good agreement with the experimental numbers (see Table 6). To see the effect of a larger active space, we also included the 3p AOs of Cl in the target AO list, obtaining a (33e,17o) active space. The corresponding CASSCF excitation energies are still poor, indicating that the necessary correlation is not of valence character. The CASSCF+NEVPT2 excitation energies in this larger space, however, remain in very good agreement with experiment. The insensitivity to active space indicates that correlations are well converged in the CASSCF+NEVPT2 treatment.

In CASSCF calculations we averaged over five doublet states, each having four doubly occupied *d* orbitals and one *d*-orbital with a single electron. The smallest SVD eigenvalue for the active space overlap with the initial guess in converged CASSCF calculations is equal to 0.930 in the case of the small (9e,5o) active space and 0.985 for the (33e,17o) active space, indicating that AVAS provides a good initial guess for CASSCF.

Table 6: Low-lying excitation energies of $[\text{CuCl}_4]^{2-}$. Energies are given in cm^{-1} .

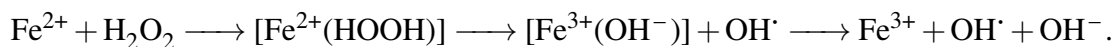
State	CASSCF		CASSCF+NEVPT2		CASPT2 ⁹³	Exp.*
	(9e,5o)	(33e,17o)	(9e,5o)	(33e,17o)	(11e,11o)	
1^2B_{2g}	6588	6588	10675	10666	11321	10500 ⁹⁵
1^2E_g	8727	8728	12832	12947	13379	12800 ⁹⁵
1^2A_{1g}	9690	9692	14021	14135	14597	

* From polarized absorption spectrum of a single-crystal D_{4h} $[\text{CuCl}_4]^{2-}$

F. Fenton reaction: an example of homolytic bond dissociation

When modeling chemical reactions with multi-configuration methods, one particularly challenging problem is the selection of active spaces capable of representing all relevant configurations in a homolytic bond dissociation process. In fact, aside from special cases, where e.g. a reasonable active space choice along the reaction path is enabled by “accidental” factors, such as molecular symmetry, or all the relevant orbitals are energetically well-separated from and unmixed with other orbitals, or the molecule is simply small enough to allow for a full-valence active space treatment, active space selection constitutes the primary bottleneck in the real-world applications of multi-configurational methods to reaction chemistry.

However, as rationalized in Sec. 2.1, we expect the AVAS scheme to be applicable when studying homolytic bond dissociation processes. To illustrate this capability, we here consider one key step of a Fenton reaction. Concretely, we consider the homolytic splitting of hydrogen peroxide (H_2O_2) by aqueous ferrous (Fe(II)) iron at low pH, to yield OH^\cdot radicals and an aqueous ferric (Fe(III)) species. Summarized, we examine the inner step of



Variants of this reaction have been proposed as possibly relevant elementary steps in biochemical processes involving iron-oxo (Fe(IV)=O) species (including such as performed by cytochromes P450⁹⁶), but this chemistry has been the subject of intense mechanistic debate since its inception in 1894.⁹⁷ The history, background, and significance of this reaction, as well as the surrounding controversies, are discussed in Refs. 96, 98, and 75.

Here no attempt is made at quantitatively describing the reaction, or at resolving any of the controversies it involves; rather, we only examine whether AVAS is capable of producing an active space capable of *qualitatively* describing the reaction mechanism of a non-trivial homolytic bond dissociation along the entire IRC. In particular, no solvation effects are considered.

To model the Fenton reaction, we considered the $[\text{Fe}(\text{H}_2\text{O})_5(\text{H}_2\text{O}_2)]^{2+}$ complex as an initial reagent. It represents an octahedral $[\text{Fe}(\text{H}_2\text{O})_6]^{2+}$ complex, a model for the aqueous ferrous ion,

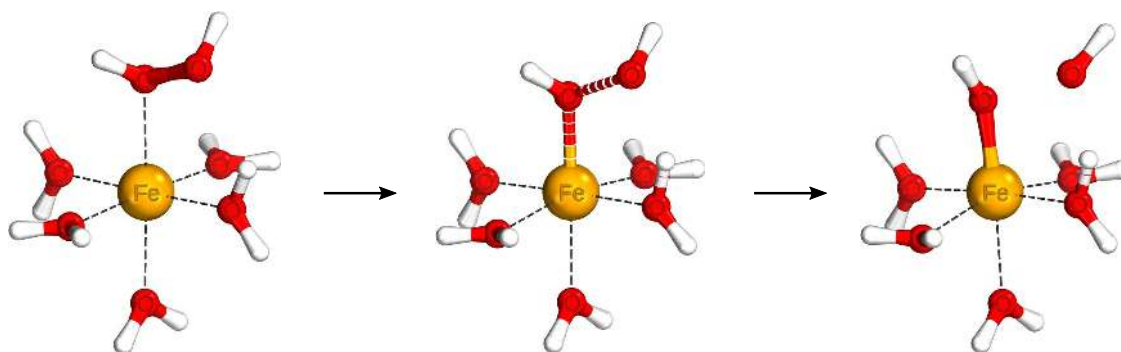


Figure 3: Structural model for the homolytic bond dissociation process of the Fenton reaction: a process of homolytic bond dissociation of hydrogen peroxide H_2O_2 , catalyzed by aqueous iron(II). The $[\text{Fe}(\text{H}_2\text{O})_5(\text{H}_2\text{O}_2)]^{2+}$ complex (left) is formed when H_2O_2 enters the coordination sphere of an octahedral $[\text{Fe}(\text{H}_2\text{O})_6]^{2+}$ complex, a model complex for the aqueous $\text{Fe}^{2+}(\text{aq})$. The homolytic bond cleavage of the coordinated hydrogen peroxide in the $[\text{Fe}(\text{H}_2\text{O})_5(\text{H}_2\text{O}_2)]^{2+}$ complex leads to the formation of two spin-coupled radical fragments: a neutral hydroxide radical OH^\cdot and an iron(III) complex $[\text{Fe}(\text{OH})(\text{H}_2\text{O})_5]^{2+}$. The center picture depicts the transition state (UB3LYP/def2-TZVP), the left and the right pictures depict two structures along the computed IRC on the reactant and product side of the reaction model.

with one water molecule substituted by the hydrogen peroxide molecule. Such a complex is expected to be formed when H_2O_2 enters the coordination sphere of $[\text{Fe}(\text{H}_2\text{O})_6]^{2+}$. The homolytic bond cleavage/dissociation of the $[\text{Fe}(\text{H}_2\text{O})_5(\text{H}_2\text{O}_2)]^{2+}$ complex leads to the formation of two spin-coupled radical fragments: ferric $[\text{Fe}(\text{H}_2\text{O})_5(\text{OH})]^{2+}$ and hydroxyl OH^\cdot radicals.⁷⁵ For this process we computed a reaction path along the intrinsic reaction coordinate, at the level of UB3LYP/def2-TZVP as described in Sec. 3 (the geometry optimizations were started with structures from Ref. 75). The reaction model is visualized in Fig. 3. The reactant complex has four unpaired electrons on iron and a total spin quantum number of $S = 4/2$.

To build the active space, we first computed ROHF wave functions (with four unpaired electrons) for each structure along the IRC. To ensure the convergence for the ROHF method and retain a continuous character of the ROHF solution along the entire reaction path, we used the orbitals obtained for the previous geometry as an initial guess for the next one, starting at the reactant side. The actual active space was then formed by the AVAS procedure, in which we used the three $2p$ orbitals of two dissociating oxygen atoms and five $3d$ orbitals of Fe as target AOs. For the open-shell treatment, we invoked option 3 from Sec. 2.5; that is, we applied the AVAS projection only on

1
2
3 doubly-occupied and virtual orbitals from ROHF wavefunction, and the four singly-occupied $3d$
4 orbitals of Fe were added unchanged to the active space to keep the correct spin S number for the
5 entire system. For the initial complex, AVAS with a 10% threshold produced 12 occupied orbitals
6 and 1 unoccupied orbital to form the (20e,13o) active space. To retain consistency of the active
7 space along the reaction path, we then fixed the size of the active space to the values thus obtained;
8 that is, for each geometry we chose the 12 occupied orbitals corresponding to the 12 largest overlap
9 eigenvalues with the target space, and the one virtual orbital with the largest overlap with the target
10 space.
11
12

13
14
15
16
17
18
19
20 With this active space, multi-configuration calculations along the reaction path were executed.
21 We wish to probe the capability of the AVAS to describe the qualitative features of the electronic
22 structure on *both* sides of the reaction, despite the fact that AVAS was constructed from a ROHF
23 wave function, which is itself incorrect on the product side. For this reason, the orbitals *were not*
24 *optimized* in the multiconfiguration treatment—that is, we use (approximate) CASCI with ROHF
25 orbitals as a multi-configuration method, and not (orbital-optimized) CASSCF. Concretely, the
26 CASCI energies were approximated with DMRG-CI ($M=1600$), which should be essentially exact
27 with this active space.
28
29
30
31
32
33
34
35

36 The obtained energies are displayed in Fig. 4 (top). During the process, the ROHF energy rises,
37 because a ROHF determinant is incapable of qualitatively describing the electronic structure of the
38 product complex. In contrast, with the same ROHF orbitals, the computed CASCI energies show
39 an energy maximum, and a decreasing energy towards the product side. In Fig. 4 (bottom) we
40 also plot the CASCI-occupation numbers of two active MOs participating in this homolytic bond
41 cleavage along the IRC. As expected for a homolytic bond cleavage, in which one doubly occupied
42 molecular orbital splits into two singly occupied molecular orbitals, the occupation numbers go
43 from $\approx 2.0/0.0$ to $\approx 1.0/1.0$ during the process. This confirms the formation of two radicals on the
44 product side. Concretely, the active orbital obtained from the virtual side of the AVAS construction
45 at ROHF level changes its occupation number from 0.06 to 0.96 along the IRC, at the DMRG-CI
46 level.
47
48
49
50
51
52
53
54
55
56
57
58
59
60

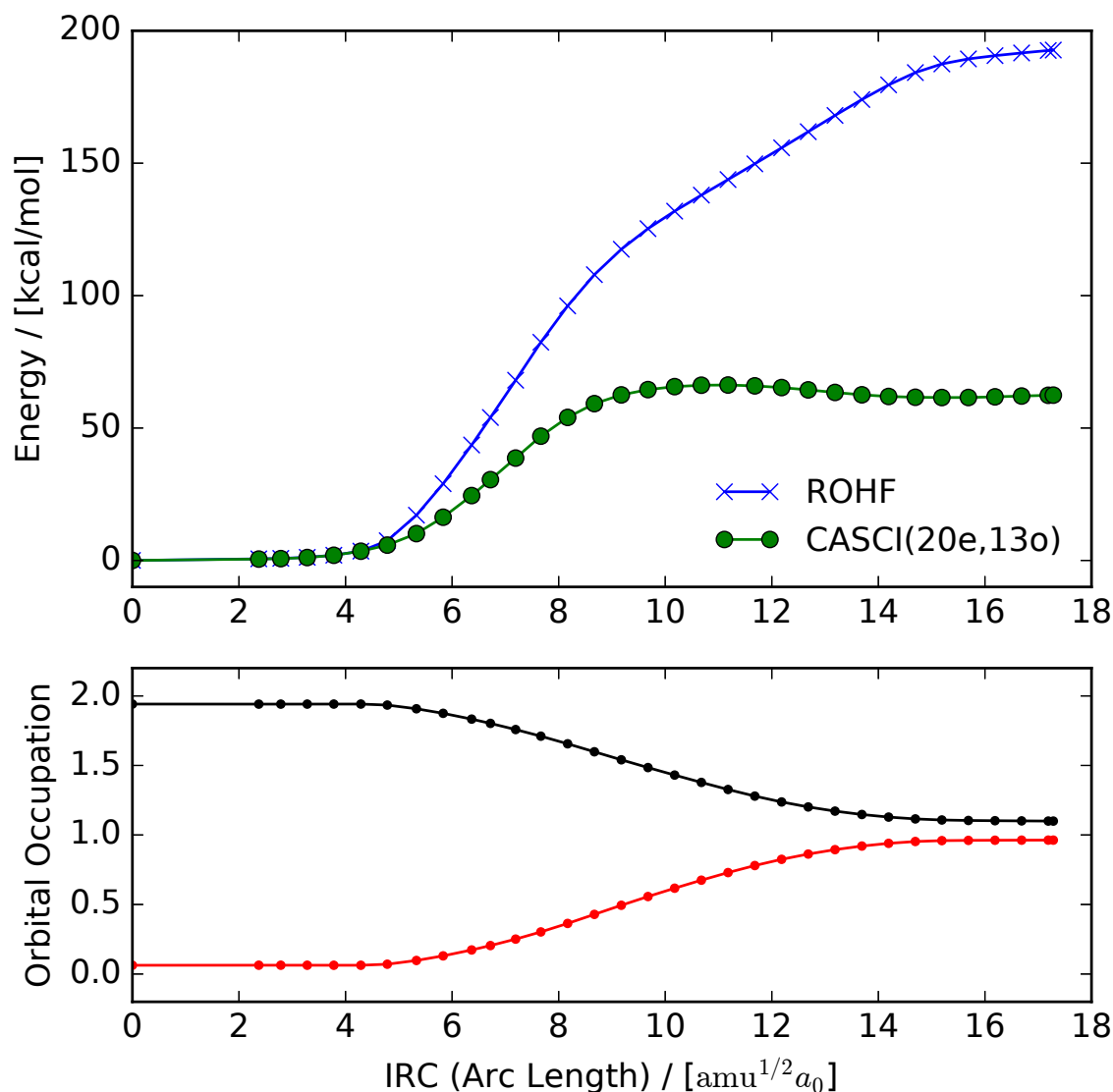


Figure 4: Computed data along the intrinsic reaction coordinate of the homolytic bond dissociation process visualized in Fig. 3. cc-pVTZ basis sets were employed, and CASCI was executed with DMRG-CI ($M=1600$), and the (20e,13o)-active space (from AVAS with the two O $2p$ electron shells and the Fe $3d$ electron shell as target AOs; see text). Top: ROHF and CASCI energies. Bottom: occupation numbers of the two active orbitals in the cleaved O–O bond (all other active orbitals retain an orbital configuration of ≈ 1.0 (for the Fe $3d$ active orbitals) and ≈ 2.0 (for the other active orbitals) along the entire IRC).

1
2
3
4 As expected at this level of theory, the absolute energetics of the process are not well re-
5 produced (compared to accurate CCSD(T)-F12 calculations⁷⁵). However, we conclude that the
6 qualitative features of the process are correctly described with the given active space. In partic-
7 ular, the AVAS contains the relevant active orbitals for describing the electronic structure of both
8 the reactant side and product side of the reaction—even though the AVAS is constructed from
9 a ROHF wave function which is qualitatively incorrect on the product side. This observation is
10 consistent with the earlier finding in DMET that relevant entanglement spaces can be constructed
11 from Hartree-Fock wave functions even in cases where the latter are qualitatively incorrect.⁵⁰ All
12 geometries, energies and occupation numbers are available in Supporting Information.
13
14
15
16
17
18
19
20
21
22
23

24 **5 Speculative extensions for complex cases**

25
26
27 We believe that AVAS is useful for the application cases fulfilling the premises described in
28 Sec. 2.1. However, it is clear that various application scenarios fall outside of this range. We
29 here outline two straight-forward extensions of the AVAS concept, which may extend its range
30 of applicability. Investigations of the details and practical performance of these approaches are
31 beyond the scope of this work and will be done elsewhere.
32
33
34
35
36
37
38
39

40 **5.1 Active spaces beyond the valence space**

41
42 For some combinations of electronic structure methods and/or application scenarios, it may be
43 desirable to construct active spaces capable of explicitly representing electronic degrees of freedom
44 which cannot be represented within a (minimal) valence-AO basis. One such case is the treatment
45 of Rydberg excited states: describing such states on even a qualitative level will require diffuse
46 orbitals in the active space.
47
48
49
50
51
52

53 The other, more common, case, is the extension of the active space in order to improve the
54 *quantitative* accuracy of computations in certain cases. For example, the accuracy of CASPT2 (and
55 related PT2 methods) for transition metal complexes is often substantially improved if a correlating
56
57
58
59
60

1
2
3
4 second radial shell of d -AOs is included into the active space. This is known as the “double-shell
5 effect”,^{99–101} and rests on the fact that doing so, effectively, relocates the treatment of d -electron
6 radial electron correlation from the (very approximate) “PT2 part” of the method to the (very
7 accurate) “CAS part”.
8
9

10
11 While in this work only valence-spanned active spaces are considered, the AVAS procedure
12 should be able to also create such extended active spaces—by including the additional desired
13 non-valence AOs in the target AO space. This can be done by simply taking any non-minimal
14 atomic natural orbital basis set (such as ANO-RCC,^{58,59} ano-pVnZ,⁶⁰ ANO-VT-XZ,⁶¹ or any
15 other) as source for these target AOs—rather than the MINAO basis used here. For example, a
16 “double-shell” active space for a $3d$ transition metal atom could be created by including *both* $3d$
17 and $4d$ AOs in the “target” AO space of AVAS. One could even add an additional shell of tight f
18 functions, to follow the idea of “correlation consistency”^{102–104} of dynamic correlation—although
19 doing so is not common practice.
20
21
22
23
24
25
26
27
28
29
30

31 32 **5.2 Handling complex metal/ligand interactions or multi-nuclear coordina-** 33 **tion complexes** 34 35

36
37 We expect that normally the simple d -electron AVAS will be sufficient to qualitatively represent
38 the multi-reference wave functions of ground states and (at least) $d \mapsto d$ -excited states of single-
39 metal coordination complexes. However, there are many application scenarios in which either
40 complicated and non-transparent metal-ligand interactions, or the presence of multiple metal cores
41 in coordination complexes, would make a straightforward active space construction via AVAS
42 either questionable (if ligand AOs or additional metal AOs are not included in the target space) or
43 infeasible (if such additional AOs are included, but the so-created active space becomes too large
44 for convenient quantitative handling).
45
46
47
48
49
50
51
52

53 In the presence of such effects it may be warranted to *combine* AVAS with either one of several
54 methods of constructing approximate FCI wave functions in large active spaces (*vide infra*), or
55 with the entanglement-based active-space construction procedure of Reiher and coworkers.^{26,34}
56
57
58
59
60

1
2
3
4 In the simplest case of suspected complex metal-ligand interactions in a single-metal complex,
5 one might, for example, first create an AVAS by including as target AOs both the d -electrons of
6 the transition metal, as well as *all* ligand AOs suspected to play an important role (e.g., all ligand
7 valence AOs from the first coordination sphere, or all p_z -orbitals of a large π -system coordinated to
8 the metal atom). In practice, the active space created from this choice will frequently be too large
9 for an accurate quantitative multi-reference calculation including dynamic correlation. However,
10 there are methods capable of providing qualitative wave functions and first-order reduced density
11 matrices at an approximate CASCI level for such active spaces. For large active spaces without
12 too many open shells, methods such as SHCI^{105,106} or FCIQMC^{107–110} can provide an efficient
13 description, while for active spaces with a larger number of open shells, but fewer than 50 orbitals,
14 QC-DMRG^{111–122} is a reliable approach. Once such a qualitative wave function is present, it can
15 be used to determine which orbitals are *not needed* in the active space.
16
17
18
19
20
21
22
23
24
25
26

27
28 Concretely, we suggest the following procedure of handling these complex cases:
29

- 30 1. Make an initial AVAS, with a target space including *all* AOs suspected to be possibly rele-
31 vant.
32
- 33 2. Compute an approximate CASCI wave function $|\Psi_{\text{approx}}\rangle$ in this active space—by using, for
34 example, SHCI¹⁰⁵ (with high threshold ϵ) or QC-DMRG¹²¹ (with low bond dimension M).
35
- 36 3. Compute the natural orbitals $\{\psi_i\}$ and their occupation numbers $\{n_i\}$ from the approximate
37 CASCI wave function $|\Psi_{\text{approx}}\rangle$.
38
- 39 4. Do the *actual* quantitative calculation with an active space $\{\psi_i; t \leq n_i \leq (2-t)\}$, where t
40 is some numeric threshold (e.g., $t = 0.05$). That is, use an active space composed of the
41 natural orbitals ψ_i from step 3 for which the occupation numbers n_i differ significantly from
42 0.0 (empty in all configurations of $|\Psi_{\text{approx}}\rangle$) and 2.0 (doubly-occupied in all configurations
43 of $|\Psi_{\text{approx}}\rangle$).
44
45
46
47
48
49
50
51
52
53
54

55 While not investigated here, we expect this combination of methods to be capable of identifying
56 suitable active spaces in many application scenarios.
57
58
59
60

5.3 Quantitative treatment of homolytic bond dissociation processes

The obtained data in Sec. 4 shows that the AVAS contains the necessary orbitals for describing the electron configurations relevant in a homolytic bond dissociation process, along the entire reaction path. However, we note that for actual quantitative calculations, particularly in *simple* concerted reaction processes without hidden intermediates,^{123–125} modified procedures may be beneficial. For example, if the chemical transformation of the reaction is sufficiently characterized by a single transition state (as in Sec. 4), a more economical way of treating the reaction might be to compute an AVAS active space only at the transition state, use it to initialize a CASSCF calculation there, and then propagate the initial orbitals from geometry to geometry. The main benefit of this latter approach would be that in the AVAS procedure, not only the rotated occupied and virtual orbitals with *very low* overlap with the target AO space could be eliminated from the active space, but also the orbitals with *very high* overlap (say, $\geq 98\%$) could be eliminated, as such values indicate that they *stay* doubly occupied/unoccupied during the reaction. The active space size could therefore be substantially reduced in such cases, therefore allowing the application of more powerful electron correlation methods. However, a detailed study of this is beyond the scope of this work.

We also wish to explicitly note that there are *many* reactions in which the chemical transformations are *not* characterized by only the transition state (e.g., see Ref. 79 Fig. 3, or Refs. 123–125), and in these cases the ability to construct a consistent active space along the entire reaction path, such as afforded by the AVAS procedure, may be valuable.

6 Conclusions

In this work, we investigated how to systematically and automatically construct molecular active spaces solely from a single determinant wavefunction together with a list of atomic valence orbitals. The atomic valence active space (AVAS) procedure is based on a straightforward linear algebraic rotation of the occupied and unoccupied molecular orbital spaces which maximizes their given atomic valence character. The method automatically detects the valence bonding partners of

1
2
3 a given atomic valence orbital, and, by using a single small threshold, can also detect non-bonding
4 orbitals without constructing spurious partners (either occupied or unoccupied).
5
6

7 To test our scheme, we tested both the quality of our orbitals as initial guesses for CASSCF
8 optimization, as well as the accuracy and stability of the valence excitation energies calculated
9 within our spaces. We find high overlap of our orbitals with fully optimized CASSCF orbitals,
10 demonstrating their high quality. We can also obtain good CASSCF excitation energies in cases
11 where the excitations are dominated by valence correlation. In molecules where the excitations
12 are not of this character, we find that the addition of dynamic correlation (through the *N*-electron
13 valence perturbation theory) yields quantitative agreement with experiment.
14
15
16
17
18
19
20
21

22 No doubt it will still be necessary to experiment with active spaces in the modeling of large
23 and very complex molecules. Our study provides two reasons to believe that the difficulty of per-
24 forming reliable multireference calculations in complex problems can be reduced using the AVAS
25 technique. First, the simple procedure makes it trivial to obtain not only the minimal active spaces,
26 but also extended active spaces, for example, including additional ligand orbitals. This makes it
27 simpler to systematically explore different active spaces, eliminating user error and subjectivity in
28 their definition, and allowing for convergence of properties with respect to the active space size.
29
30 Second, systematically varying active space size, while including a dynamic correlation treatment
31 (such as NEVPT2) provides a straightforward way to assess whether our active space is converged,
32 as computed observables should become insensitive to the active space size. For these reasons, we
33 believe that the AVAS construction provides a simple route to painless multireference calculations
34 by non-experts, particularly in complex systems involving transition metals.
35
36
37
38
39
40
41
42
43
44
45
46
47
48

49 **Supporting Information**

50

51
52 Supporting information: Additional computational details; geometric structures of the studied tran-
53 sition metal complexes; additional figures of studied active orbital spaces; numerical and structural
54 data of the studied reaction path of the Fenton reaction. This material is available free of charge
55
56
57
58
59
60

1
2
3 via the Internet at <http://pubs.acs.org>.
4
5
6

7 8 **Acknowledgments** 9

10
11 We acknowledge the US National Science Foundation for funding this research primarily through
12 the award NSF:CHE-1665333. Additional support for software development and to support QS
13 was provided through NSF:CHE-1657286. We acknowledge additional support for GKC from the
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

21 22 **References** 23 24

- 25 (1) Veryazov, V.; Malmqvist, P.; Roos, B. How to Select Active Space for Multiconfigurational
26 Quantum Chemistry? *Int. J. Quantum Chem.* **2011**, *111*, 3329–3338.
- 27
28 (2) Roos, B. O. The Complete Active Space Self-Consistent Field Method and its Applications
29 in Electronic Structure Calculations. *Adv. Chem. Phys.* **1987**, *69*, 339–445.
- 30
31 (3) Pierloot, K. In *Computational Organometallic Chemistry*; Cundari, T. R., Ed.; Marcel
32 Dekker: New York, 2001; pp 123–158.
- 33
34 (4) Olivares-Amaya, R.; Hu, W.; Nakatani, N.; Sharma, S.; Yang, J.; Chan, G. K.-L. The ab-
35 initio density matrix renormalization group in practice. *J. Chem. Phys.* **2015**, *142*, 034102.
- 36
37 (5) Sharma, S.; Chan, G. K.-L. A flexible multi-reference perturbation theory by minimizing
38 the Hylleraas functional with matrix product states. *J. Chem. Phys.* **2014**, *141*, 111101.
- 39
40 (6) Chan, G. K.-L.; Sharma, S. The density matrix renormalization group in quantum chemistry.
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
Ann. Rev. Phys. Chem. **2011**, *62*, 465–481.
- (7) Sharma, S.; Alavi, A. Multireference linearized coupled cluster theory for strongly corre-
lated systems using matrix product states. *J. Chem. Phys.* **2015**, *143*, 102815.

- 1
2
3
4 (8) Sharma, S.; Knizia, G.; Guo, S.; Alavi, A. Combining Internally Contracted States and
5 Matrix Product States To Perform Multireference Perturbation Theory. *J. Chem. Theory*
6 *Comput.* **2017**, *13*, 488–498.
7
8
9
10 (9) Ma, Y.; Knecht, S.; Keller, S.; Reiher, M. Second-Order Self-Consistent-Field Density-
11 Matrix Renormalization Group. *arXiv preprint arXiv:1611.05972* **2016**,
12
13
14
15 (10) Freitag, L.; Knecht, S.; Angeli, C.; Reiher, M. Multireference Perturbation Theory with
16 Cholesky Decomposition for the Density Matrix Renormalization Group. *J. Chem. Theory*
17 *Comput.* **2017**, *13*, 451–459.
18
19
20
21 (11) Gomez, J. A.; Henderson, T. M.; Scuseria, G. E. Singlet-paired coupled cluster theory for
22 open shells. *J. Chem. Phys.* **2016**, *144*, 244117.
23
24
25
26 (12) Henderson, T. M.; Bulik, I. W.; Stein, T.; Scuseria, G. E. Seniority-based coupled cluster
27 theory. *J. Chem. Phys.* **2014**, *141*, 244104.
28
29
30
31 (13) Lehtola, S.; Parkhill, J.; Head-Gordon, M. Cost-effective description of strong correlation:
32 Efficient implementations of the perfect quadruples and perfect hexuples models. *J. Chem.*
33 *Phys.* **2016**, *145*, 134110.
34
35
36
37 (14) Zhang, T.; Evangelista, F. A. A deterministic projector configuration interaction approach
38 for the ground state of quantum many-body systems. *J. Chem. Theory Comput.* **2016**, *12*,
39 4326–4337.
40
41
42
43 (15) Schriber, J. B.; Evangelista, F. A. An adaptive configuration interaction approach for
44 strongly correlated electrons with tunable accuracy. *J. Chem. Phys.* **2016**, *144*, 161106.
45
46
47
48 (16) Li, C.; Evangelista, F. A. Towards numerically robust multireference theories: The driven
49 similarity renormalization group truncated to one-and two-body operators. *J. Chem. Phys.*
50 **2016**, *144*, 164114.
51
52
53
54
55
56
57
58
59
60

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- (17) Li, C.; Evangelista, F. A. Multireference Driven Similarity Renormalization Group: A Second-Order Perturbative Analysis. *J. Chem. Theory. Comput.* **2015**, *11*, 2097–2108.
- (18) Sand, A. M.; Truhlar, D. G.; Gagliardi, L. Efficient algorithm for multiconfiguration pair-density functional theory with application to the heterolytic dissociation energy of ferrocene. *J. Chem. Phys.* **2017**, *146*, 034101.
- (19) Gagliardi, L.; Truhlar, D. G.; Li Manni, G.; Carlson, R. K.; Hoyer, C. E.; Bao, J. L. Multiconfiguration pair-density functional theory: A new way to treat strongly correlated systems. *Acc. Chem. Res.* **2017**, *50*, 66–73.
- (20) Thomas, R. E.; Sun, Q.; Alavi, A.; Booth, G. H. Stochastic multiconfigurational self-consistent field theory. *J. Chem. Theory Comput.* **2015**, *11*, 5316–5325.
- (21) Fosso-Tande, J.; Nguyen, T.-S.; Gidofalvi, G.; DePrince III, A. E. Large-scale variational two-electron reduced-density-matrix-driven complete active space self-consistent field methods. *J. Chem. Theory Comput.* **2016**, *12*, 2260–2271.
- (22) Ma, D.; Manni, G. L.; Gagliardi, L. The generalized active space concept in multiconfigurational self-consistent field methods. *J. Chem. Phys.* **2011**, *135*, 044128.
- (23) Manni, G. L.; Ma, D.; Aquilante, F.; Olsen, J.; Gagliardi, L. SplitGAS method for strong correlation and the challenging case of Cr₂. *J. Chem. Theory Comput.* **2013**, *9*, 3375–3384.
- (24) Manni, G. L.; Smart, S. D.; Alavi, A. Combining the Complete Active Space Self-Consistent Field Method and the Full Configuration Interaction Quantum Monte Carlo within a Super-CI Framework, with Application to Challenging Metal-Porphyrins. *J. Chem. Theory Comput.* **2016**, *12*, 1245–1258.
- (25) Manni, G. L.; 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.

- 1
2
3
4 (26) Keller, S.; Boguslawski, K.; Janowski, T.; Reiher, M.; Pulay, P. Selection of active spaces
5 for multiconfigurational wavefunctions. *J. Chem. Phys.* **2015**, *142*, 244104.
6
7
8 (27) Stein, C. J.; Reiher, M. Automated Selection of Active Orbital Spaces. *J. Chem. Theory*
9 *Comput.* **2016**, *12*, 1760–1771.
10
11
12 (28) Pulay, P.; Hamilton, T. P. UHF natural orbitals for defining and starting MC-SCF calcula-
13 tions. *J. Chem. Phys.* **1988**, *88*, 4926–4933.
14
15
16 (29) Bofill, J. M.; Pulay, P. The unrestricted natural orbital–complete active space (UNO–
17 CAS) method: An inexpensive alternative to the complete active space–self-consistent-field
18 (CAS–SCF) method. *J. Chem. Phys.* **1989**, *90*, 3637–3646.
19
20
21 (30) Jensen, H. J. A.; Jørgensen, P.; Ågren, H.; Olsen, J. Second-order Møller-Plesset pertur-
22 bation theory as a configuration and orbital generator in multiconfiguration self-consistent
23 field calculations. *J. Chem. Phys.* **1988**, *88*, 3834–3839.
24
25
26 (31) Abrams, M. L.; Sherrill, C. D. Natural orbitals as substitutes for optimized orbitals in com-
27 plete active space wavefunctions. *Chem. Phys. Lett.* **2004**, *395*, 227–232.
28
29
30 (32) Wouters, S.; Bogaerts, T.; Van Der Voort, P.; Van Speybroeck, V.; Van Neck, D. DMRG-
31 SCF study of the singlet, triplet, and quintet states of oxo-Mn (Salen). *J. Chem. Phys.* **2014**,
32 *140*.
33
34
35 (33) Stein, C. J.; von Burg, V.; Reiher, M. The Delicate Balance of Static and Dynamic Electron
36 Correlation. *J. Chem. Theory Comput.* **2016**, *12*, 3764–3773.
37
38
39 (34) Stein, C. J.; Reiher, M. Automated Identification of Relevant Frontier Orbitals for Chemical
40 Compounds and Processes. *CHIMIA Int. J. Chem.* **2017**, *71*, 170–176.
41
42
43 (35) Iwata, S. Valence type vacant orbitals for configuration interaction calculations. *Chem. Phys.*
44 *Lett.* **1981**, *83*, 134–138.
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- (36) Schmidt, M. W.; Hull, E. A.; Windus, T. L. Valence Virtual Orbitals: An unambiguous ab initio quantification of the LUMO concept. *J. Phys. Chem. A* **2015**, *119*, 10408–10427.
- (37) Ruedenberg, K.; Schmidt, M. W.; Gilbert, M. M.; Elbert, S. T. Are atoms intrinsic to molecular electronic wavefunctions? I. The FORS model. *Chem. Phys.* **1982**, *71*, 41–49.
- (38) Ruedenberg, K.; Schmidt, M. W.; Gilbert, M. M. Are atoms intrinsic to molecular electronic wavefunctions? II. Analysis of FORS orbitals. *Chem. Phys.* **1982**, *71*, 51–64.
- (39) Lee, M. S.; Head-Gordon, M. Polarized atomic orbitals for self-consistent field electronic structure calculations. *J. Chem. Phys.* **1997**, *107*, 9085–9095.
- (40) Lee, M. S.; Head-Gordon, M. Extracting polarized atomic orbitals from molecular orbital calculations. *Int. J. Quantum Chem.* **2000**, *76*, 169–184.
- (41) Auer, A. A.; Nooijen, M. Dynamically screened local correlation method using enveloping localized orbitals. *J. Chem. Phys.* **2006**, *125*, 024104.
- (42) Lu, W. C.; Wang, C. Z.; Schmidt, M. W.; Bytautas, L.; Ho, K. M.; Ruedenberg, K. Molecule intrinsic minimal basis sets. I. Exact resolution of ab initio optimized molecular orbitals in terms of deformed atomic minimal-basis orbitals. *J. Chem. Phys.* **2004**, *120*, 2629–2637.
- (43) Subotnik, J. E.; Dutoi, A. D.; Head-Gordon, M. Fast localized orthonormal virtual orbitals which depend smoothly on nuclear coordinates. *J. Chem. Phys.* **2005**, *123*, 114108.
- (44) Ivanic, J.; Atchity, G.; Ruedenberg, K. Intrinsic local constituents of molecular electronic wave functions. I. Exact representation of the density matrix in terms of chemically deformed and oriented atomic minimal basis set orbitals. *Theor. Chem. Acc.* **2008**, *120*, 281–294.
- (45) Glezakou, V.-A.; Elbert, S. T.; Xantheas, S. S.; Ruedenberg, K. Analysis of Bonding Patterns in the Valence Isoelectronic Series O₃, S₃, SO₂, and OS₂ in Terms of Oriented Quasi-Atomic Molecular Orbitals. *J. Phys. Chem. A* **2010**, *114*, 8923–8931.

- 1
2
3
4 (46) Laikov, D. N. Intrinsic minimal atomic basis representation of molecular electronic wave-
5 functions. *Int. J. Quantum Chem.* **2011**, *111*, 2851–2867.
6
7
8 (47) Knizia, G. Intrinsic Atomic Orbitals: An Unbiased Bridge between Quantum Theory and
9 Chemical Concepts. *J. Chem. Theory Comput.* **2013**, *9*, 4834–4843.
10
11
12 (48) West, A. C.; Schmidt, M. W.; Gordon, M. S.; Ruedenberg, K. A comprehensive analysis
13 of molecule-intrinsic quasi-atomic, bonding, and correlating orbitals. I. Hartree-Fock wave
14 functions. *J. Chem. Phys.* **2013**, *139*, 234107.
15
16
17 (49) Knizia, G.; Chan, G. K.-L. Density Matrix Embedding: A Simple Alternative to Dynamical
18 Mean-Field Theory. *Phys. Rev. Lett.* **2012**, *109*, 186404.
19
20
21 (50) Knizia, G.; Chan, G. K.-L. Density Matrix Embedding: A Strong-Coupling Quantum Em-
22 bedding Theory. *J. Chem. Theory Comput.* **2013**, *9*, 1428–1432.
23
24
25 (51) Zheng, B.-X.; Chan, G. K.-L. Ground-state phase diagram of the square lattice Hubbard
26 model from density matrix embedding theory. *Phys. Rev. B* **2016**, *93*, 035126.
27
28
29 (52) Wouters, S.; Jiménez-Hoyos, C. A.; Sun, Q.; Chan, G. K.-L. A practical guide to density
30 matrix embedding theory in quantum chemistry. *J. Chem. Theory Comput.* **2016**, *12*, 2706–
31 2719.
32
33
34 (53) Bulik, I. W.; Scuseria, G. E.; Dukelsky, J. Density matrix embedding from broken symmetry
35 lattice mean fields. *Phys. Rev. B* **2014**, *89*, 035140.
36
37
38 (54) Bulik, I. W.; Chen, W.; Scuseria, G. E. Electron correlation in solids via density embedding
39 theory. *J. Chem. Phys.* **2014**, *141*, 054113.
40
41
42 (55) Tsuchimochi, T.; Welborn, M.; Voorhis, T. V. Density matrix embedding in an antisym-
43 metrized geminal power bath. *J. Chem. Phys.* **2015**, *143*, 024107.
44
45
46 (56) Welborn, M.; Tsuchimochi, T.; Voorhis, T. V. Bootstrap embedding: An internally consistent
47 fragment-based method. *J. Chem. Phys.* **2015**, *145*, 074102.
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3
4 (57) Sun, Q.; Chan, G. K.-L. Exact and Optimal Quantum Mechanics/Molecular Mechanics
5 Boundaries. *J. Chem. Theory Comput.* **2014**, *10*, 3784–3790.
6
7
8
9 (58) Widmark, P.-O.; Malmqvist, P.-Å.; Roos, B. O. Density matrix averaged atomic natural
10 orbital (ANO) basis sets for correlated molecular wave functions. *Theor. Chem. Acc.* **1990**,
11 *77*, 291–306.
12
13
14
15 (59) Roos, B. O.; Lindh, R.; Malmqvist, P.-Å.; Veryazov, V.; Widmark, P.-O. Main group atoms
16 and dimers studied with a new relativistic ANO basis set. *J. Phys. Chem. A* **2004**, *108*,
17 2851–2858.
18
19
20
21
22 (60) Neese, F.; Valeev, E. F. Revisiting the atomic natural orbital approach for basis sets: Robust
23 systematic basis sets for explicitly correlated and conventional correlated ab initio methods?
24 *J. Chem. Theory Comput.* **2010**, *7*, 33–43.
25
26
27
28
29 (61) Claudino, D.; Gargano, R.; Bartlett, R. Coupled-cluster based basis sets for valence correla-
30 tion calculations. *J. Chem. Phys.* **2016**, *144*, 104106.
31
32
33
34 (62) Sun, Q.; Berkelbach, T. C.; Blunt, N. S.; Booth, G. H.; Guo, S.; Li, Z.; Liu, J.; McClain, J.;
35 Sayfutyarova, E. R.; Sharma, S.; Wouters, S.; Chan, G. K.-L. The Python-based Simulations
36 of Chemistry Framework (PySCF). *arXiv:1701.08223v2* **2017**,
37
38
39
40
41 (63) Amos, A.; Hall, G. Single determinant wave functions. *Proc. R. Soc. London, Ser. A.* 1961;
42 pp 483–493.
43
44
45
46 (64) Angeli, C.; Cimiraglia, R.; Evangelisti, S.; Leininger, T.; Malrieu, J. P. Introduction of n-
47 electron valence states for multireference perturbation theory. *J. Chem. Phys.* **2001**, *114*,
48 10252–10264.
49
50
51
52
53 (65) Angeli, C.; Cimiraglia, R.; Malrieu, J. P. N-electron valence state perturbation theory: a fast
54 implementation of the strongly contracted variant. *Chem. Phys. Lett.* **2001**, *350*, 297–305.
55
56
57
58
59
60

- 1
2
3
4 (66) Angeli, C.; Cimiraglia, R.; Malrieu, J. P. n-electron valence state perturbation theory: A
5 spinless formulation and an efficient implementation of the strongly contracted and of the
6 partially contracted variants. *J. Chem. Phys.* **2002**, *117*, 9138–9153.
7
8
9
10 (67) Guo, S.; Watson, M. A.; Hu, W.; Sun, Q.; Chan, G. K.-L. N-Electron Valence State Pertur-
11 bation Theory Based on a Density Matrix Renormalization Group Reference Function, with
12 Applications to the Chromium Dimer and a Trimer Model of Poly(p-Phenylenevinylene). *J.*
13 *Chem. Theory Comput.* **2016**, *12*, 1583–1591.
14
15
16
17 (68) Sharma, S.; Chan, G. K.-L. Spin-adapted density matrix renormalization group algorithms
18 for quantum chemistry. *J. Chem. Phys.* **2012**, *136*.
19
20
21
22 (69) Dunning, Jr., T. H. Gaussian basis sets for use in correlated molecular calculations. I. The
23 atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
24
25
26
27 (70) Balabanov, N.; Peterson, K. Systematically convergent basis sets for transition metals. I.
28 All-electron correlation consistent basis sets for the 3d elements Sc–Zn. *J. Chem. Phys.*
29 **2005**, *123*, 064107.
30
31
32
33 (71) Liu, W.; Peng, D. Infinite-order quasirelativistic density functional method based on the
34 exact matrix quasirelativistic theory. *J. Chem. Phys.* **2006**, *125*, 044102.
35
36
37
38 (72) Peng, D.; Middendorf, N.; Weigend, F.; Reiher, M. An efficient implementation of two-
39 component relativistic exact-decoupling methods for large molecules. *J. Chem. Phys.* **2013**,
40 *138*, 184105.
41
42
43
44 (73) Becke, A. D. Density functional thermochemistry. III. The role of exact exchange. *J.*
45 *Chem. Phys.* **1993**, *98*, 5648–5652.
46
47
48
49 (74) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadru-
50 ple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem.*
51 *Phys.* **2005**, *7*, 3297–3305.
52
53
54
55
56
57
58
59
60

- 1
2
3
4 (75) Petit, A. S.; Penniford, R. C.; Harvey, J. N. Electronic structure and formation of simple
5 ferryloxo complexes: Mechanism of the Fenton reaction. *Inorg. Chem.* **2014**, *53*, 6473–
6 6481.
7
8
9
10 (76) Deglmann, P.; Furche, F.; Ahlrichs, R. An efficient implementation of second analytical
11 derivatives for density functional methods. *Chem. Phys. Lett.* **2002**, *362*, 511–518.
12
13
14 (77) Fukui, K. Formulation of the reaction coordinate. *J. Phys. Chem.* **1970**, *74*, 4161–4163.
15
16
17 (78) Knizia, G. IboView; see <http://www.iboview.org> for information, source code, and executables (accessed: 2017-02-05). 2017.
18
19
20
21
22 (79) Knizia, G.; Klein, J. E. M. N. Electron Flow in Reaction Mechanisms—Revealed from First
23 Principles. *Angew. Chem. Int. Ed.* **2015**, *54*, 5518–5522.
24
25
26
27 (80) Ishimura, K.; Hada, M.; Nakatsuji, H. Ionized and excited states of ferrocene: Symmetry
28 adapted cluster-configuration-interaction study. *J. Chem. Phys.* **2002**, *117*, 6533–6537.
29
30
31
32 (81) Harding, M.; Metzroth, T.; Gauss, J.; Auer, A. A. Parallel calculation of CCSD and
33 CCSD(T) analytic first and second derivatives. *J. Chem. Theory Comput.* **2008**, *4*, 64–74.
34
35
36
37 (82) Lauher, J. W.; Hoffmann, R. Structure and chemistry of Bis(cyclopentadienyl)-ML_n com-
38 plexes. *J. Am. Chem. Soc.* **1976**, *98*, 1729–1742.
39
40
41
42 (83) Rohmer, M.-M.; Veillard, A.; Wood, M. H. Excited states and electronic spectrum of fer-
43 rocene. *Chem. Phys. Lett.* **1974**, *29*, 466–468.
44
45
46
47 (84) Fromager, E.; Knecht, S.; Jensen, H. J. A. Multiconfiguration time-dependent density-
48 functional theory based on range separation. *J. Chem. Phys.* **2013**, *138*, 084101.
49
50
51
52 (85) Armstrong, A. T.; Smith, F.; McGlynn, E. E. S. P. Electronic Absorption Spectrum of Fer-
53 rocene. *J. Chem. Phys.* **1967**, *46*, 4321.
54
55
56
57
58
59
60

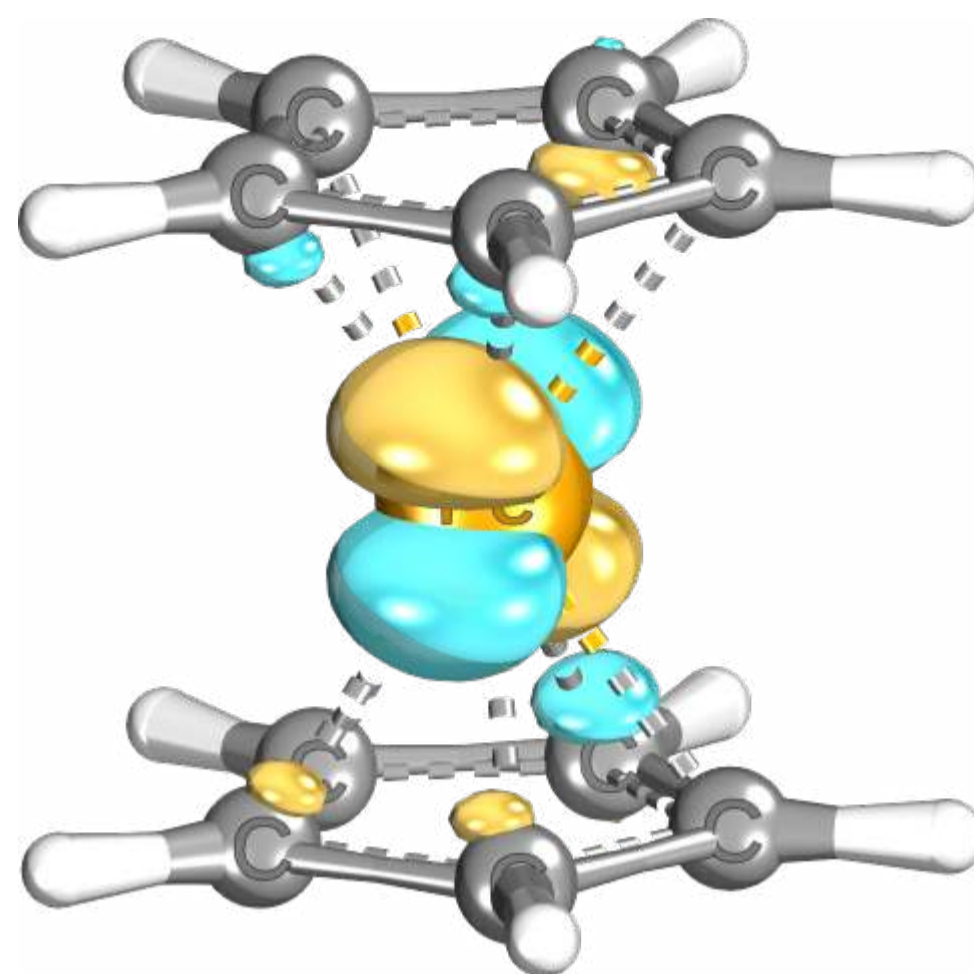
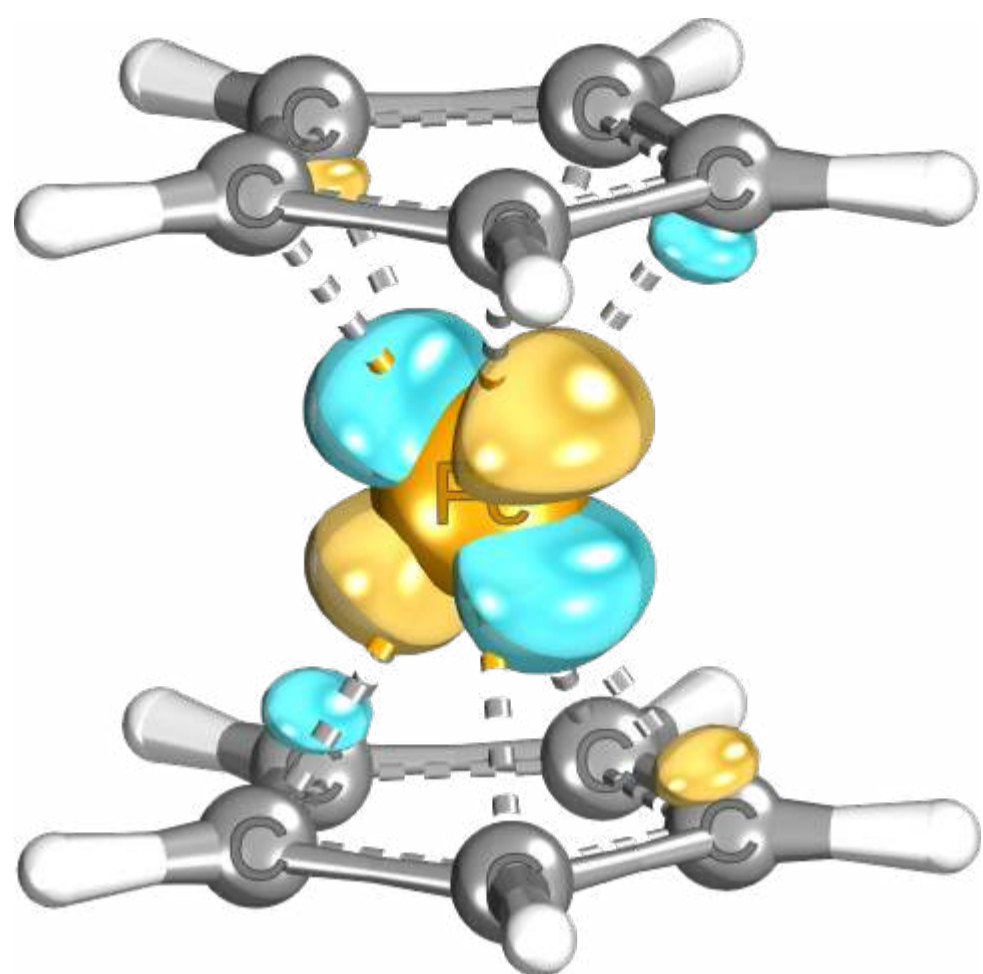
- 1
2
3
4 (86) Sohn, Y. S.; Hendrickson, D. N.; Gray, B. Electronic Structures of Metallocenes. *J. Am.*
5 *Chem. Soc.* **1971**, *93*, 3603.
6
7
8
9 (87) Klein, J. E. M. N.; Miehlisch, B.; Holzwarth, M. S.; Bauer, M.; Milek, M.; Khusni-
10 yarov, M. M.; Knizia, G.; Werner, H.-J.; Plietker, B. The Electronic Ground State of
11 $[\text{Fe}(\text{CO})_3(\text{NO})]^-$: A Spectroscopic and Theoretical Study. *Angew. Chem. Int. Ed.* **2014**,
12 *53*, 1790–1794.
13
14
15
16
17 (88) Klein, J. E. M. N.; Knizia, G.; Miehlisch, B.; Kästner, J.; Plietker, B. Fe or Fe–NO Cataly-
18 sis? A Quantum Chemical Investigation of the $[\text{Fe}(\text{CO})_3(\text{NO})]^-$ -Catalyzed Cloke-Wilson
19 Rearrangement. *Chem. Eur. J.* **2014**, *20*, 7254–7257.
20
21
22
23
24 (89) Lin, C.-H.; Pursley, D.; Klein, J. E. M. N.; Teske, J.; Allen, J. A.; Rami, F.;
25 Köhn, A.; Plietker, B. Non-decarbonylative photochemical versus thermal activation of
26 $\text{Bu}_4\text{N}[\text{Fe}(\text{CO})_3(\text{NO})]$ —the Fe-catalyzed Cloke-Wilson rearrangement of vinyl and arylcy-
27 clopropanes. *Chem. Sci.* **2015**, *6*, 7034.
28
29
30
31
32
33 (90) Al-Abdalla, A.; Seijo, L.; Barandiaran, Z. A new interpretation of the bonding and spec-
34 troscopy of the tetraoxoferrate(VI) FeO_4^{2-} ion. *J. Chem. Phys.* **1998**, *109*, 6396.
35
36
37
38 (91) Brunold, T. C.; Hauser, A.; Güdel, H. U. Absorption and luminescence spectroscopy of
39 ferrate(VI) doped into crystals of K_2MO_4 (M = S, Se, Cr, Mo). *J. Lumin.* **1994**, *59*, 321–
40 332.
41
42
43
44
45 (92) Brunold, T. C.; Güdel, H. U.; Kück, S.; Huber, G. Excited state properties of ferrate(VI)
46 doped crystals of K_2SO_4 and K_2CrO_4 . *J. Lumin.* **1995**, *65*, 293–301.
47
48
49
50 (93) Vancoillie, S.; Malmqvist, P.-Å.; Pierloot, K. Calculation of EPR g Tensors for Transition-
51 Metal Complexes Based on Multiconfigurational Perturbation Theory (CASPT2). *Chem.*
52 *Phys. Chem.* **2007**, *8*, 1803–1815.
53
54
55
56
57
58
59
60

- 1
2
3
4 (94) Collison, D.; Gahan, B.; Garner, C. D.; Mabbs, F. E. Electronic absorption spectra of some
5 oxovanadium(IV) compounds. *J. Chem. Soc., Dalton Trans.* **1980**, 667–674.
6
7
8 (95) Willett, R. D.; Liles, O. L.; Michelson, C. Electronic absorption spectra of monomeric
9 copper(II) chloride species and the electron spin resonance spectrum of the square-planar
10 CuCl_4^{2-} ion. *Inorg. Chem.* **1967**, 6, 1885–1889.
11
12
13 (96) Dunford, H. B. Oxidations of iron (II)/(III) by hydrogen peroxide: from aquo to enzyme.
14 *Coord. Chem. Rev.* **2002**, 233, 311–318.
15
16
17 (97) Fenton, H. Oxidation of tartaric acid in presence of iron. *J. Chem. Soc., Trans.* **1894**, 65,
18 899–910.
19
20
21 (98) Kremer, M. Mechanism of the Fenton reaction. Evidence for a new intermediate. *Phys.*
22 *Chem. Chem. Phys.* **1999**, 1, 3595–3605.
23
24
25 (99) , Jr., T. H. D.; Botch, B. H.; Harrison, J. F. On the orbital description of the $4s3d^{n+1}$ states
26 of the transition metal atoms. *J. Chem. Phys.* **1980**, 72, 3419–3420.
27
28
29 (100) Botch, B. H.; , Jr., T. H. D.; Harrison, J. F. Valence correlation in the $s^2 d^n$, sd^{n+1} , and d^{n+2}
30 states of the first-row transition metal atoms. *J. Chem. Phys.* **1981**, 75, 3466–3471.
31
32
33 (101) Pierloot, K. In *Computational Organometallic Chemistry*; Cundari, T. R., Ed.; Marcel
34 Dekker: New York, 2001; pp 123–158.
35
36
37 (102) Dunning Jr, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The
38 atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, 90, 1007–1023.
39
40
41 (103) Almlöf, J.; Taylor, P. R. General contraction of Gaussian basis sets. I. Atomic natural orbitals
42 for first-and second-row atoms. *J. Chem. Phys.* **1987**, 86, 4070–4077.
43
44
45 (104) Almlöf, J.; Taylor, P. R. General contraction of Gaussian basis sets. II. Atomic natural or-
46 bitals and the calculation of atomic and molecular properties. *J. Chem. Phys.* **1990**, 92,
47 551–560.
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3
4 (105) Sharma, S.; Holmes, A. A.; Jeanmairet, G.; Alavi, A.; Umrigar, C. J. Semistochastic Heat-
5 bath Configuration Interaction method: selected configuration interaction with semistochas-
6 tic perturbation theory. *J. Chem. Theory Comput.* **2017**, *13*, 1595–1604.
7
8
9
10
11 (106) Holmes, A. A.; Tubman, N. M.; Umrigar, C. Heat-Bath Configuration Interaction: An Ef-
12 ficient Selected Configuration Interaction Algorithm Inspired by Heat-Bath Sampling. *J.*
13 *Chem. Theory Comput.* **2016**, *12*, 3674–3680.
14
15
16
17
18 (107) Booth, G. H.; Thom, A. J.; Alavi, A. Fermion Monte Carlo without fixed nodes: A game of
19 life, death, and annihilation in Slater determinant space. *J. Chem. Phys.* **2009**, *131*, 054106.
20
21
22 (108) Cleland, D.; Booth, G. H.; Alavi, A. Survival of the fittest: Accelerating convergence in full
23 configuration-interaction quantum Monte Carlo. *J. Chem. Phys.* **2010**, *132*, 041103.
24
25
26
27 (109) Booth, G. H.; Chan, G. K.-L. Excited states, dynamic correlation functions and spectral
28 properties from full configuration interaction quantum Monte Carlo. *J. Chem. Phys.* **2012**,
29 *137*, 191102.
30
31
32
33
34 (110) Overy, C.; Booth, G. H.; Blunt, N.; Shepherd, J. J.; Cleland, D.; Alavi, A. Unbiased re-
35 duced density matrices and electronic properties from full configuration interaction quantum
36 Monte Carlo. *J. Chem. Phys.* **2014**, *141*, 244117.
37
38
39
40
41 (111) White, S. R.; Martin, R. L. Ab initio quantum chemistry using the density matrix renormal-
42 ization group. *J. Chem. Phys.* **1999**, *110*, 4127–4130.
43
44
45
46 (112) Mitrushenkov, A. O.; Fano, G.; Ortolani, F.; Linguerri, R.; Palmieri, P. Quantum chemistry
47 using the density matrix renormalization group. *J. Chem. Phys.* **2001**, *115*, 6815–6821.
48
49
50
51 (113) Mitrushenkov, A. O.; Linguerri, R.; Palmieri, P.; Fano, G. Quantum chemistry using the
52 density matrix renormalization group II. *J. Chem. Phys.* **2003**, *119*, 4148–4158.
53
54
55
56 (114) Chan, G. K.-L.; Head-Gordon, M. Highly correlated calculations with a polynomial cost
57
58
59
60

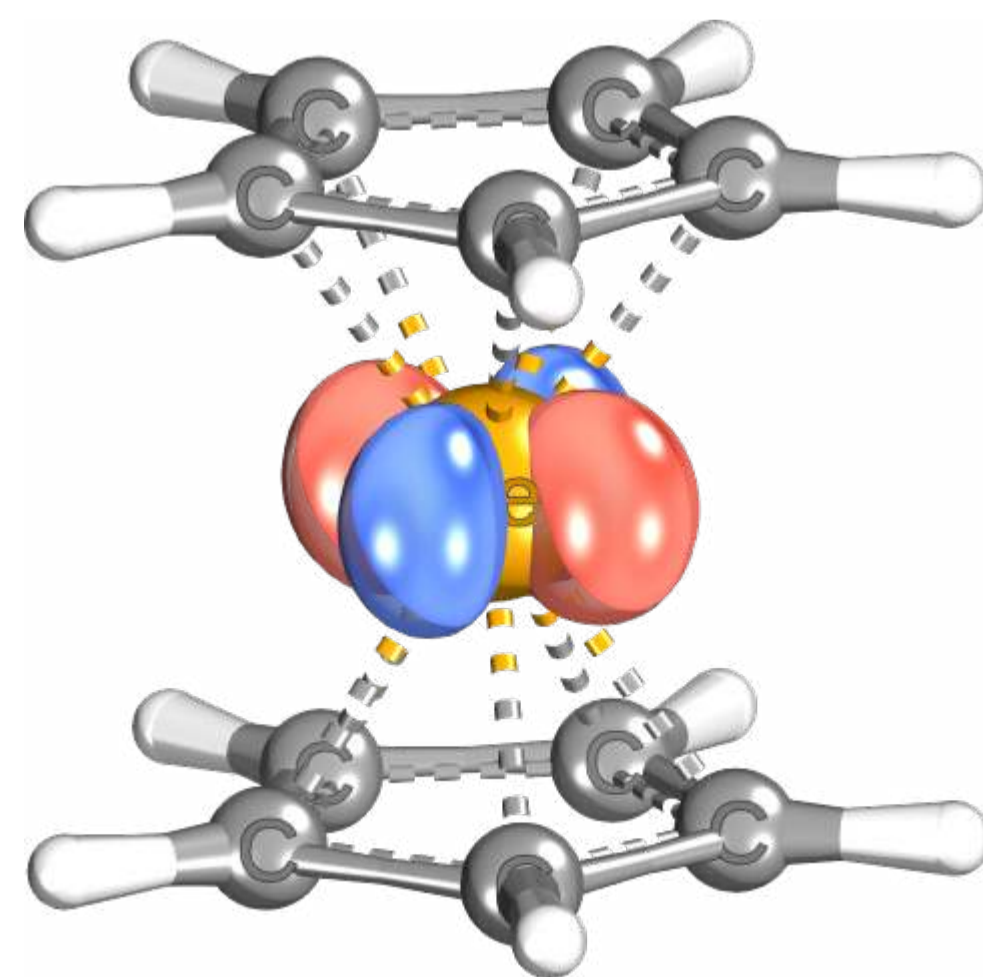
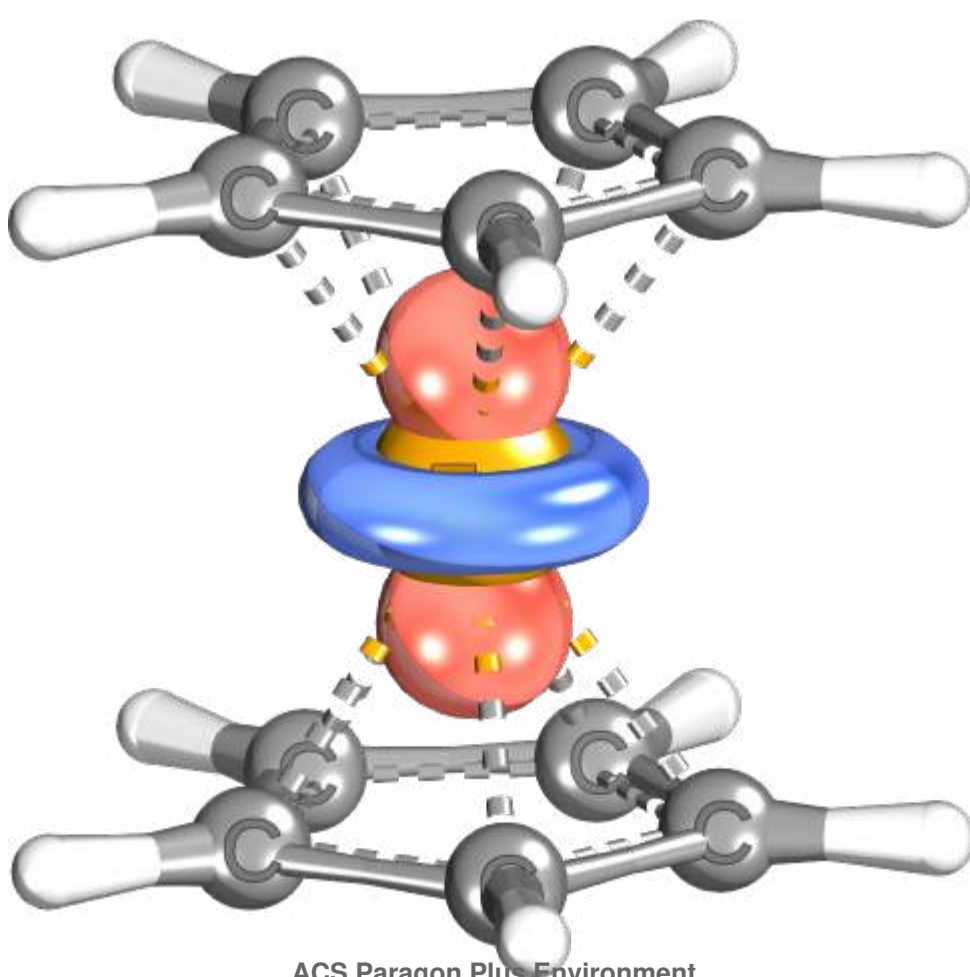
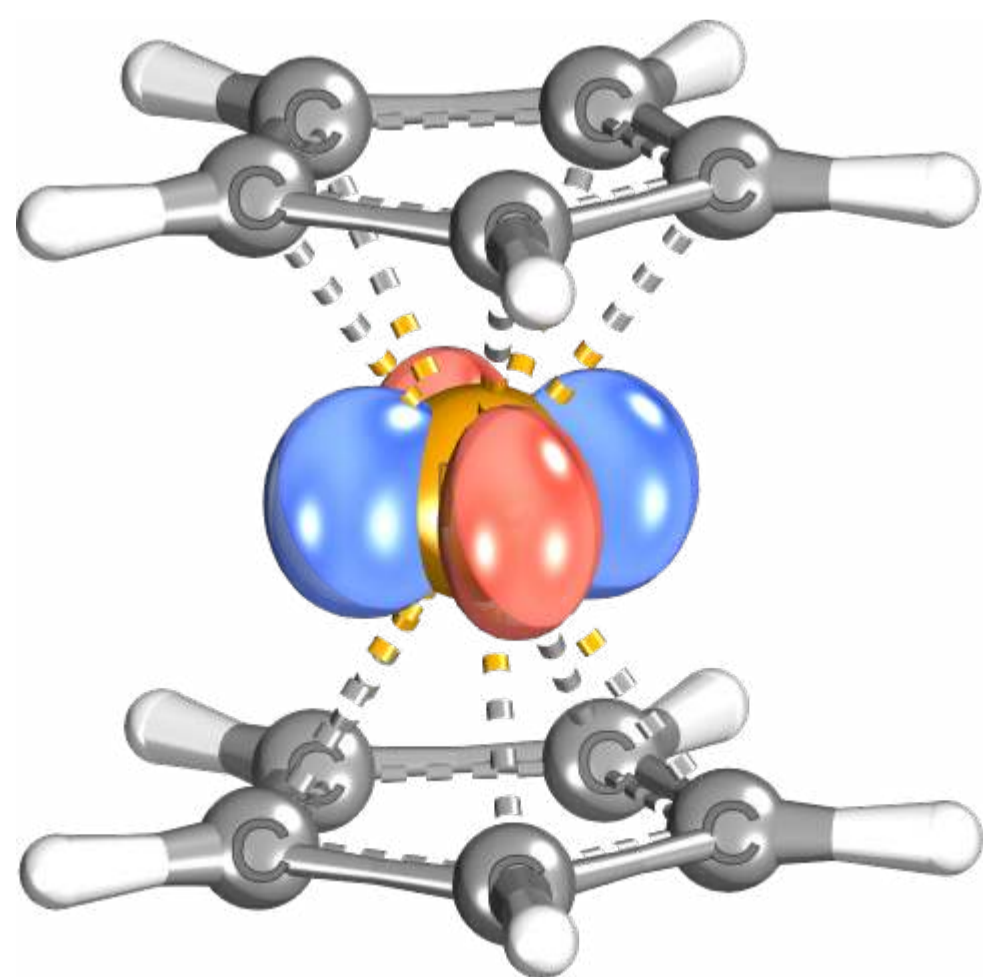
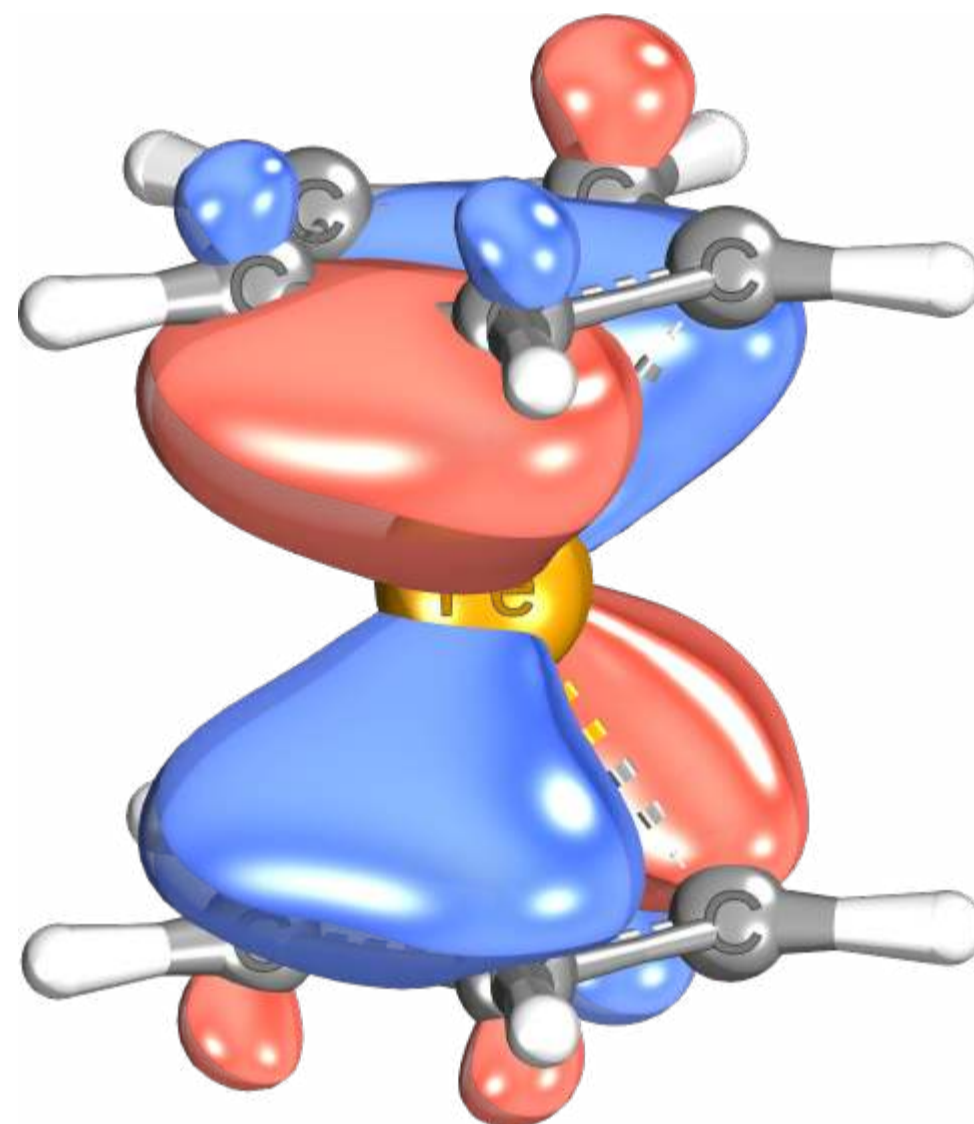
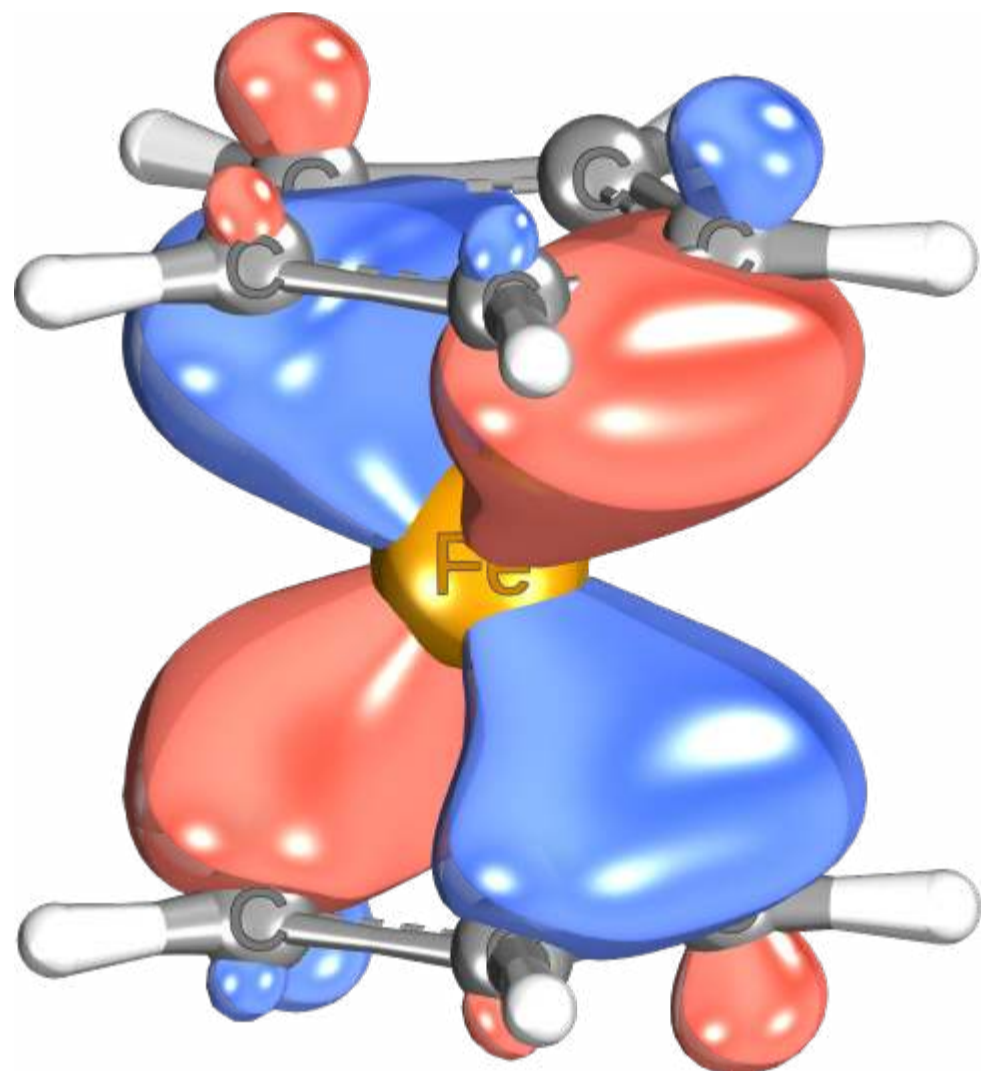
- 1
2
3
4 algorithm: A study of the density matrix renormalization group. *J. Chem. Phys.* **2002**, *116*,
5 4462–4476.
6
7
8
9 (115) Legeza, Ö.; Röder, J.; Hess, B. Controlling the accuracy of the density-matrix
10 renormalization-group method: The dynamical block state selection approach. *Phys. Rev.*
11 *B* **2003**, *67*, 125114.
12
13
14
15 (116) Chan, G. K.-L.; Head-Gordon, M. Exact solution (within a triple-zeta, double polarization
16 basis set) of the electronic Schrödinger equation for water. *J. Chem. Phys.* **2003**, *118*, 8551–
17 8554.
18
19
20
21
22 (117) Moritz, G.; Hess, B. A.; Reiher, M. Convergence behavior of the density-matrix renormal-
23 ization group algorithm for optimized orbital orderings. *J. Chem. Phys.* **2005**, *122*, 024107.
24
25
26
27 (118) Chan, G. K.-L.; Van Voorhis, T. Density-matrix renormalization-group algorithms with
28 nonorthogonal orbitals and non-Hermitian operators, and applications to polyenes. *J. Chem.*
29 *Phys.* **2005**, *122*, 204101.
30
31
32
33
34 (119) Moritz, G.; Reiher, M. Construction of environment states in quantum-chemical density-
35 matrix renormalization group calculations. *J. Chem. Phys.* **2006**, *124*, 034103.
36
37
38
39 (120) Chan, G. K.-L.; Sharma, S. The density matrix renormalization group in quantum chemistry.
40 *Ann. Rev. Phys. Chem.* **2011**, *62*, 465–481.
41
42
43
44 (121) Sharma, S.; Chan, G. K.-L. Spin-adapted density matrix renormalization group algorithms
45 for quantum chemistry. *J. Chem. Phys.* **2012**, *136*, 124121.
46
47
48
49 (122) Olivares-Amaya, R.; Hu, W.; Nakatani, N.; Sharma, S.; Yang, J.; Chan, G. K.-L. The ab-
50 initio density matrix renormalization group in practice. *J. Chem. Phys.* **2015**, *142*, 034102.
51
52
53
54 (123) Kraka, E.; Cremer, D. Computational analysis of the mechanism of chemical reactions in
55 terms of reaction phases: Hidden intermediates and hidden transition states. *Acc. Chem.*
56 *Res.* **2010**, *43*, 591–601.
57
58
59
60

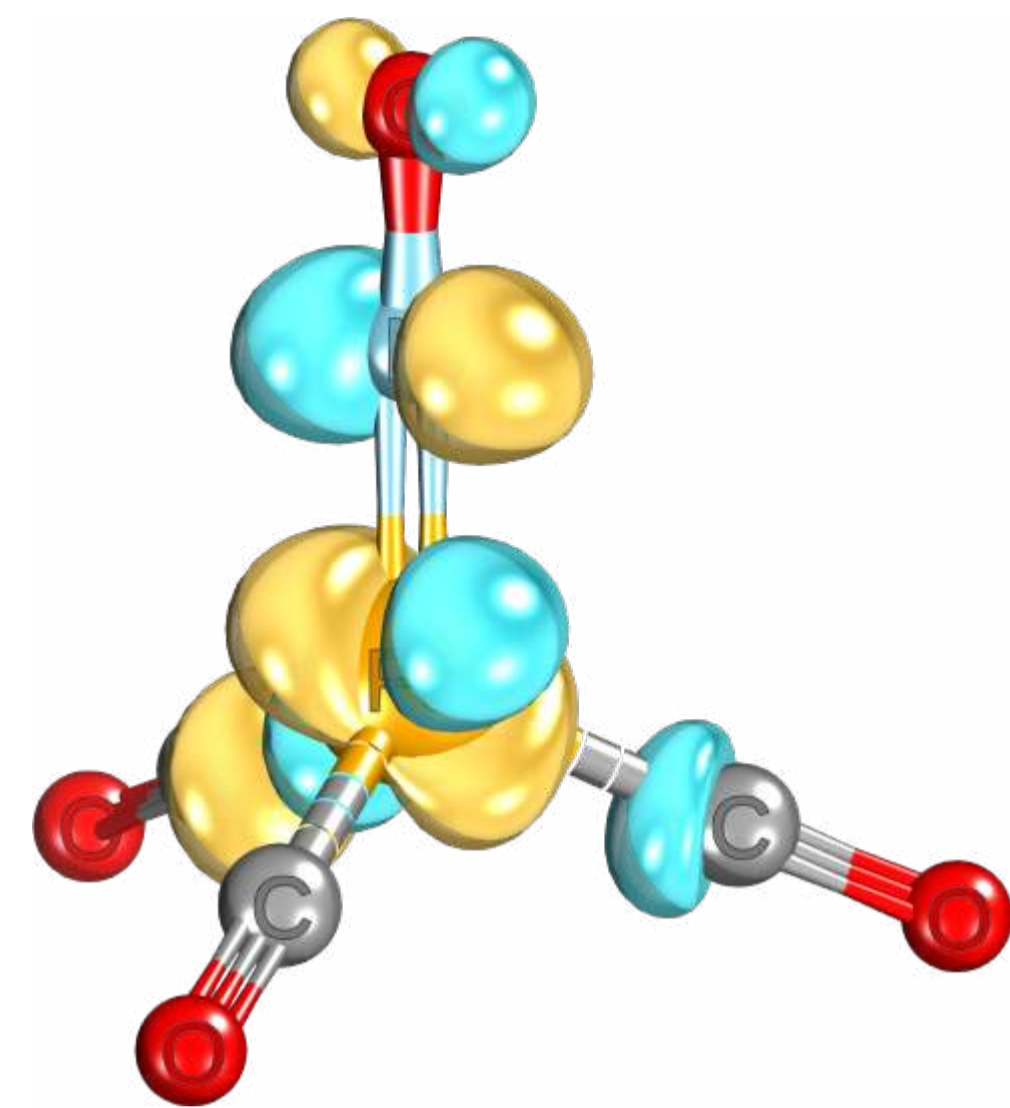
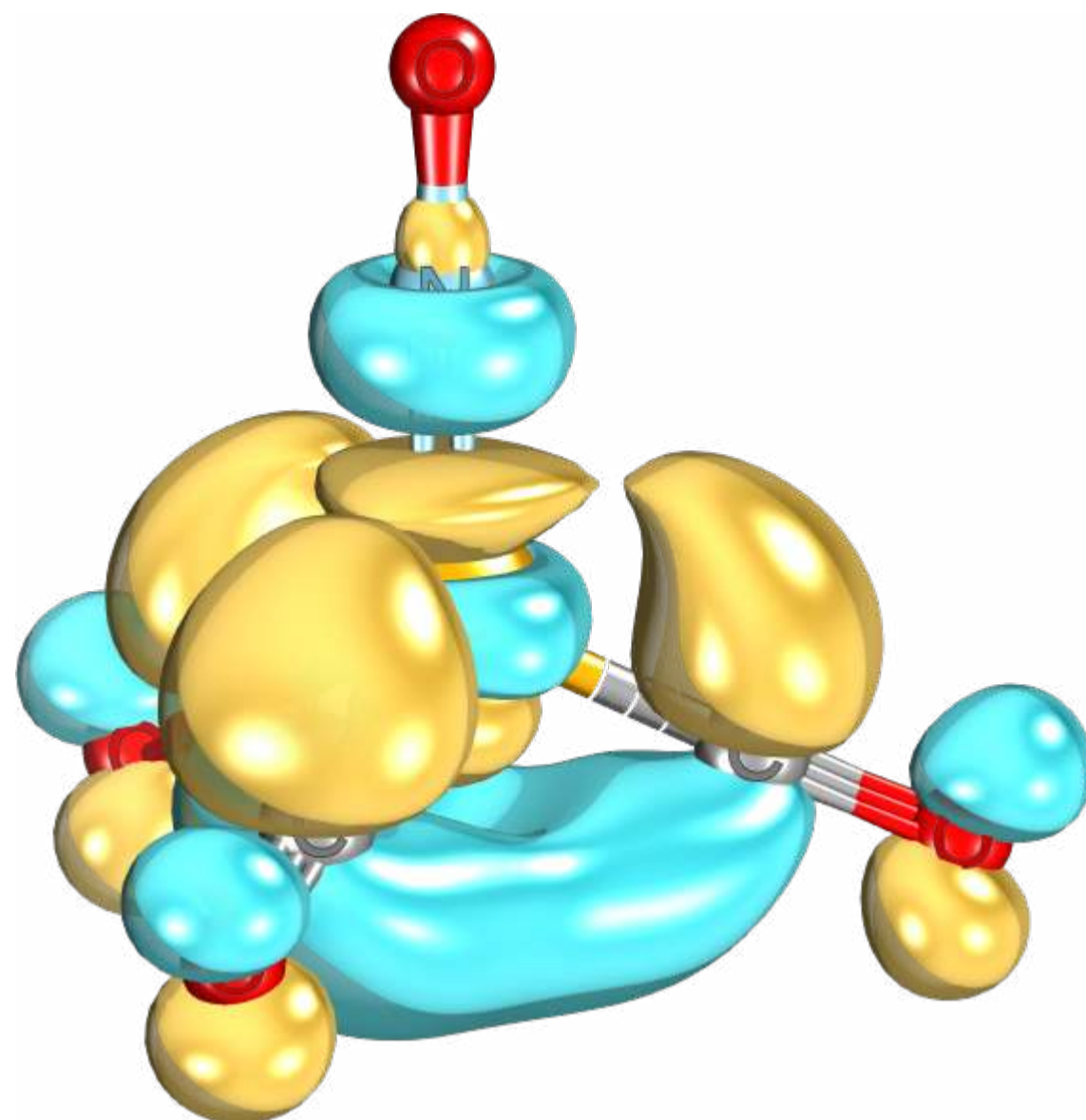
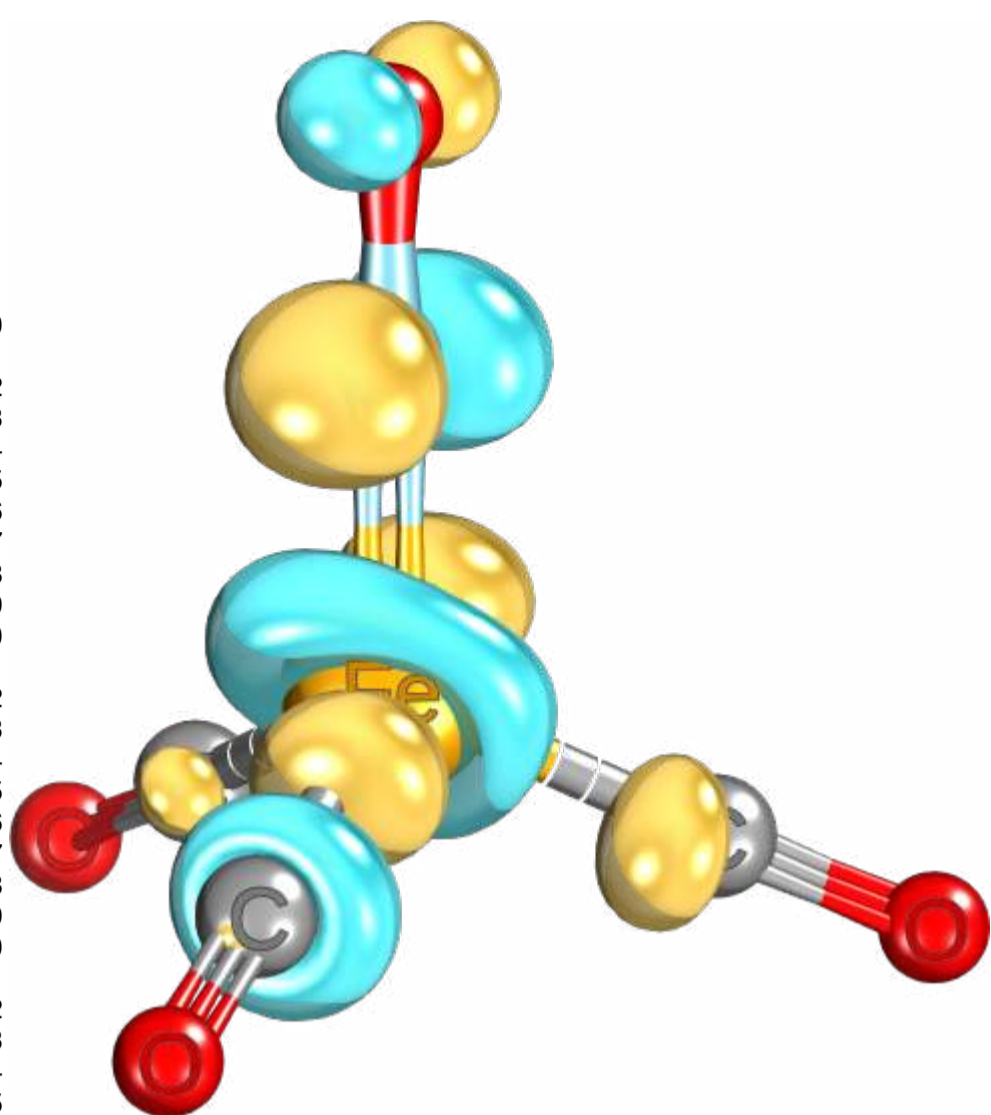
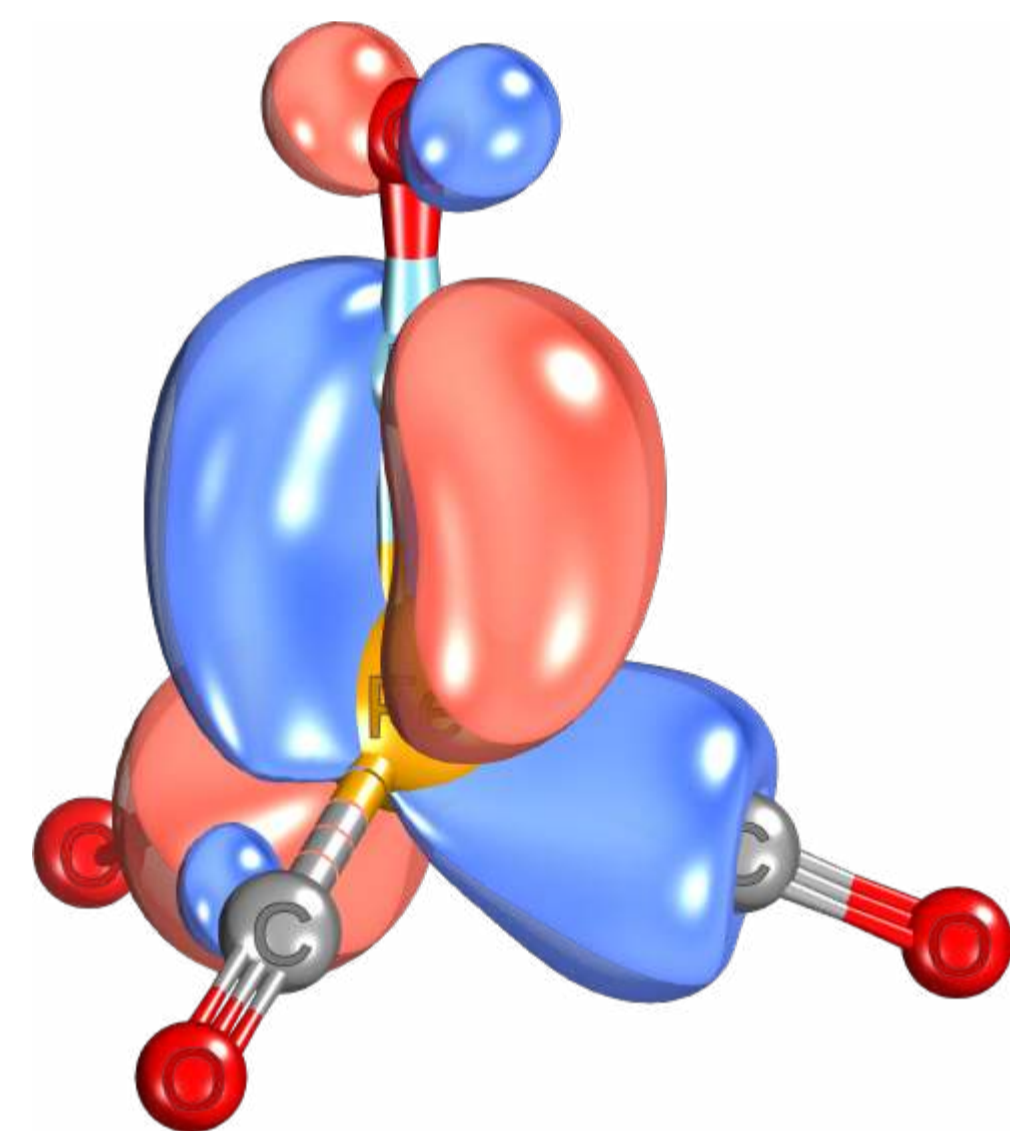
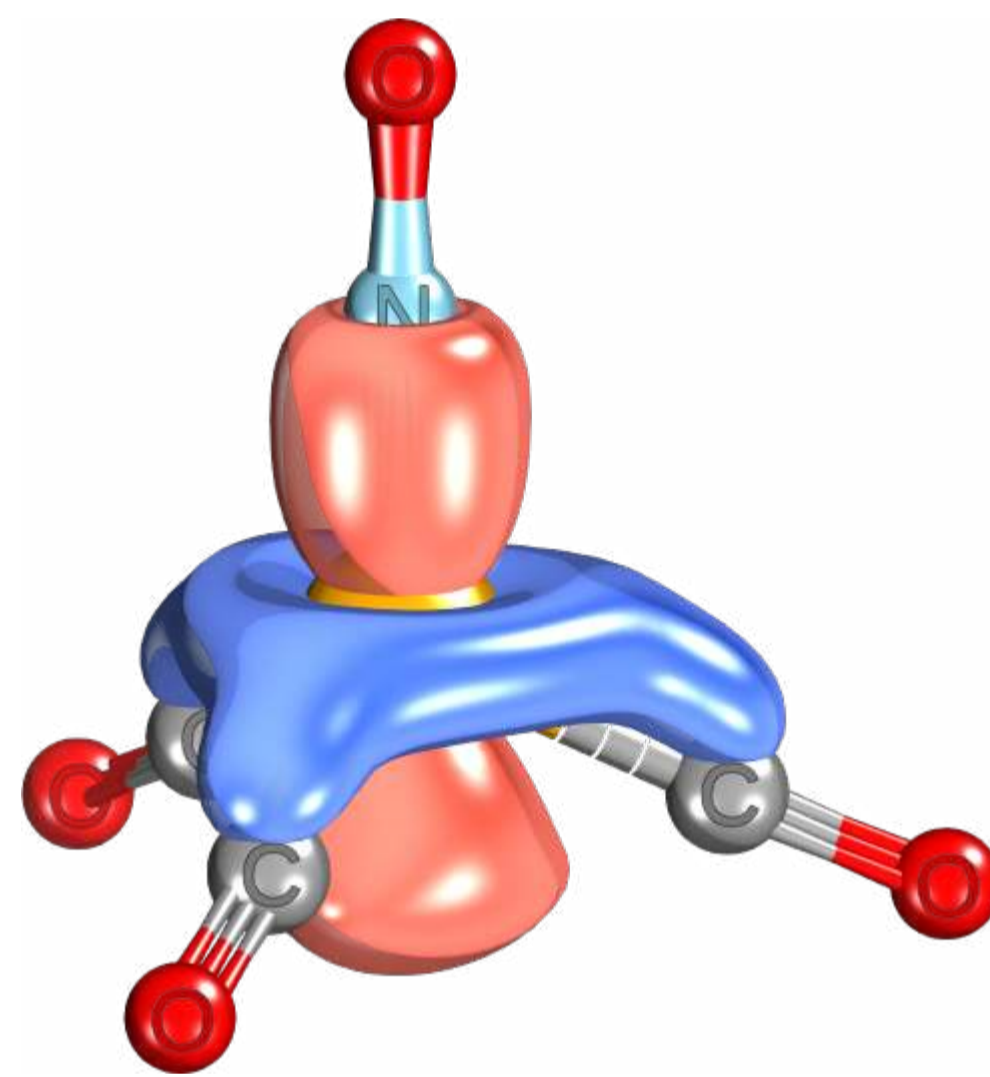
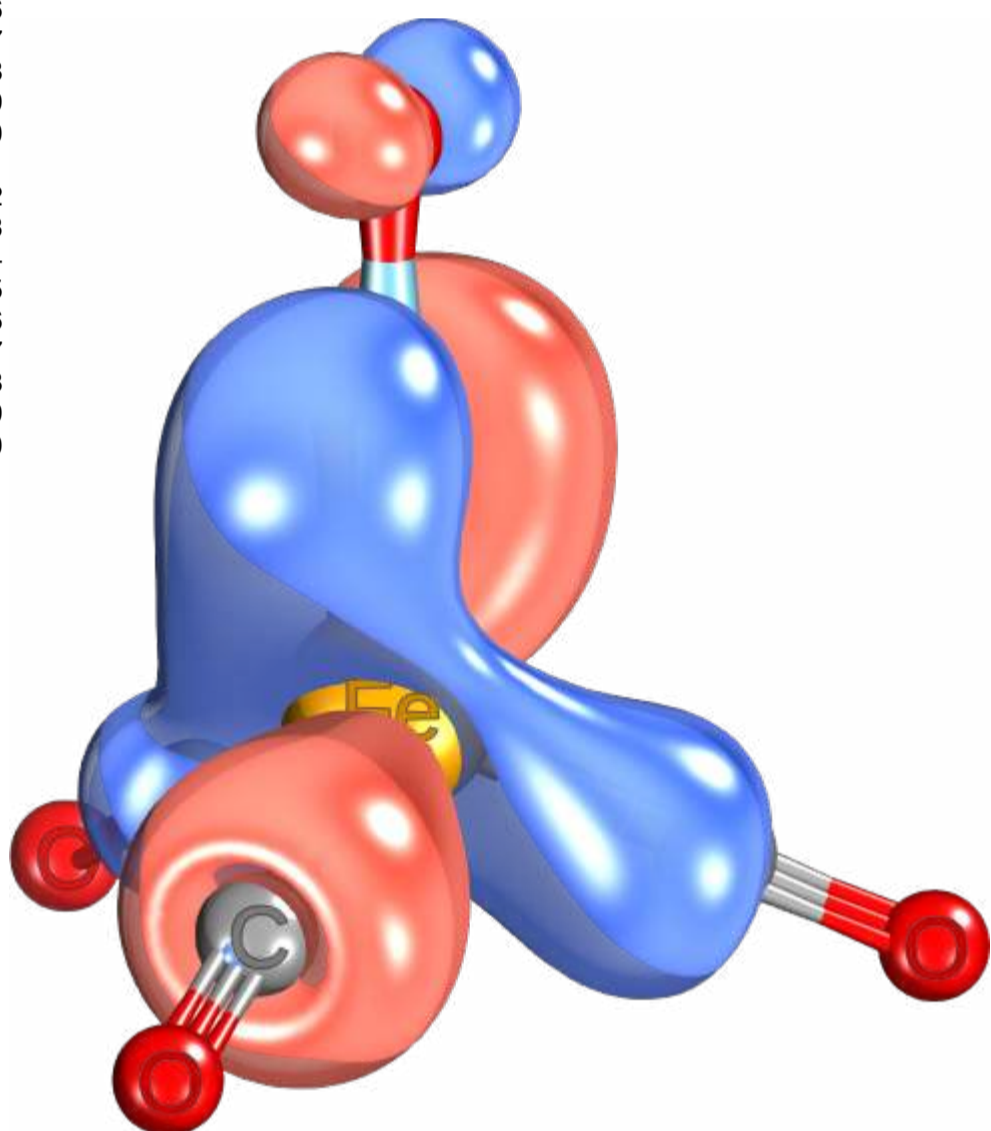
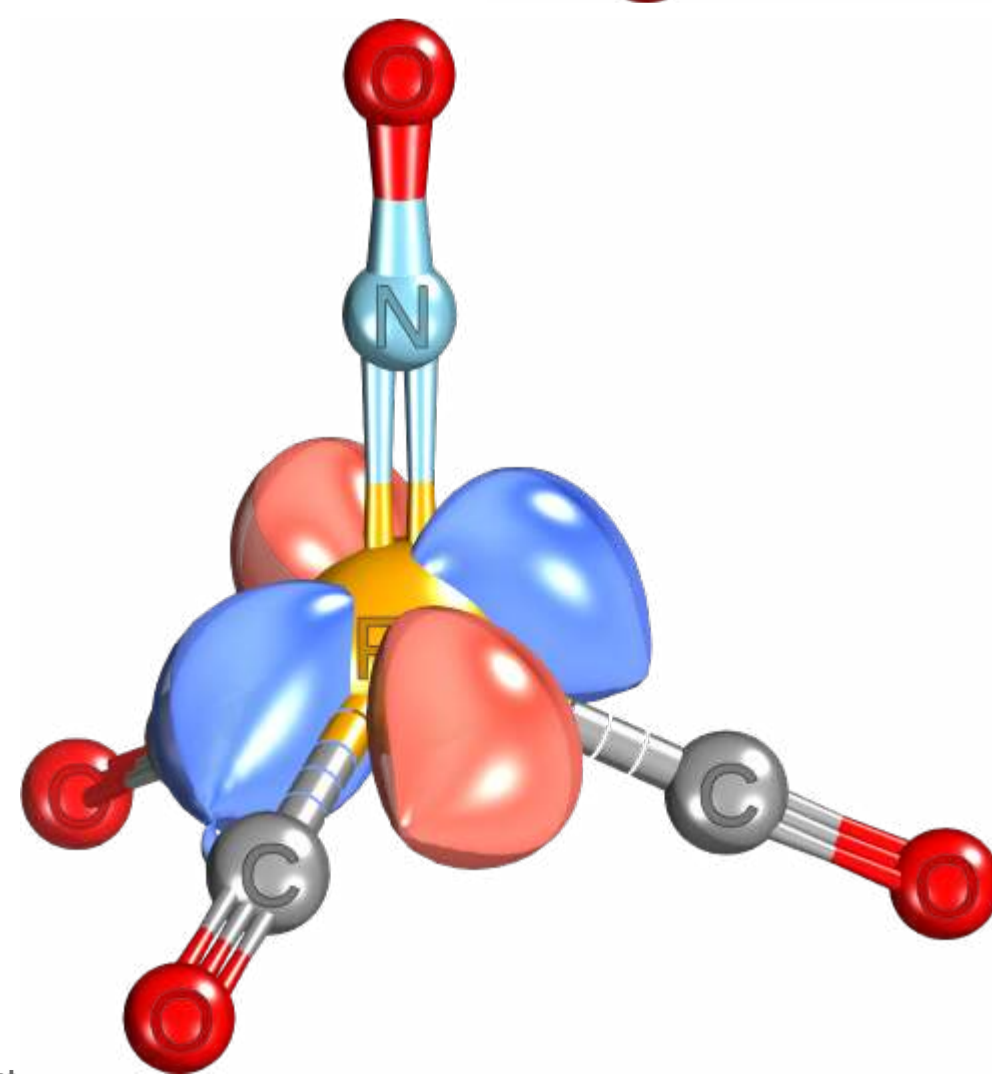
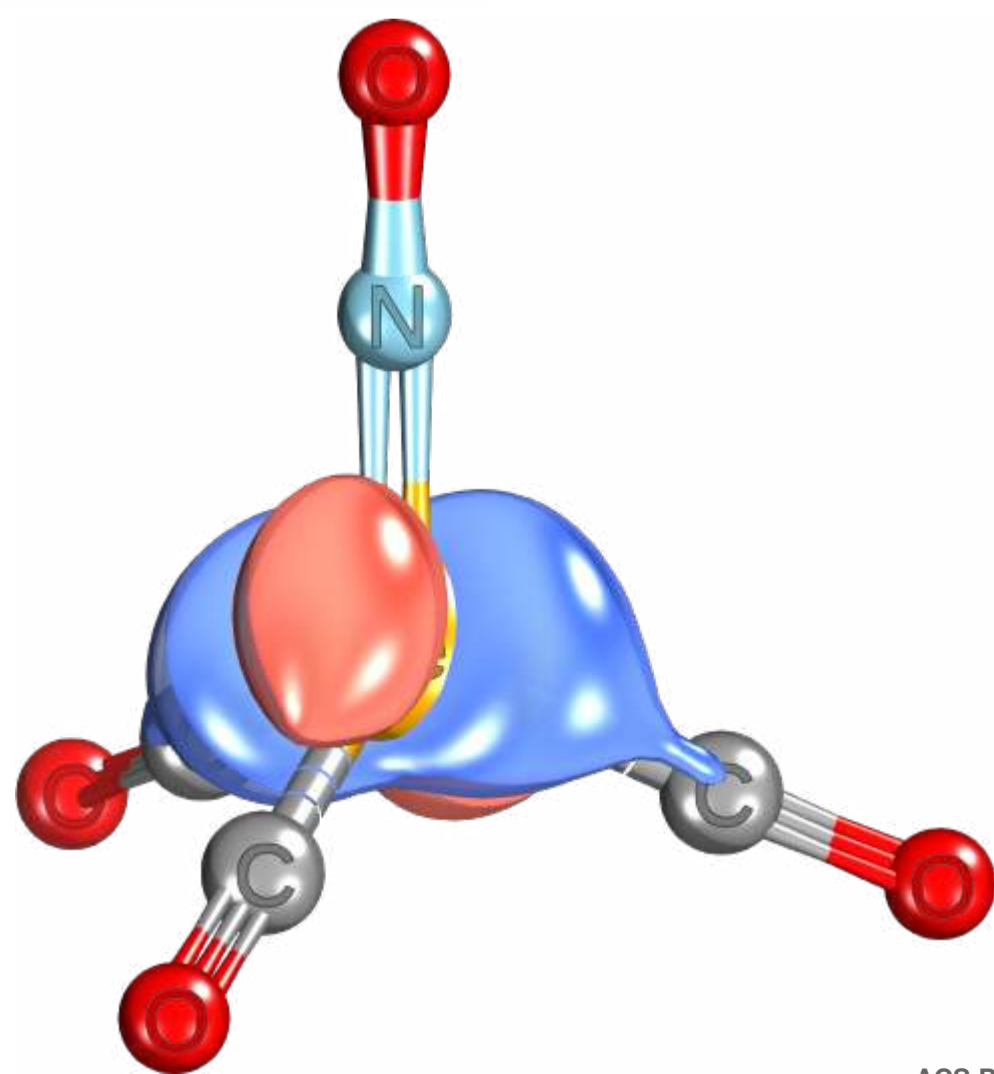
- 1
2
3
4 (124) Joo, H.; Kraka, E.; Quapp, W.; Cremer, D. The mechanism of a barrierless reaction: Hidden
5 transition state and hidden intermediates in the reaction of methylene with ethene. *Mol.*
6 *Phys.* **2007**, *105*, 2697–2717.
7
8
9
10 (125) Kraka, E.; Joo, H.; Cremer, D. A stunning example for a spontaneous reaction with a com-
11 plex mechanism: the vinylidene–acetylene cycloaddition reaction. *Mol. Phys.* **2010**, *108*,
12 2667–2685.
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

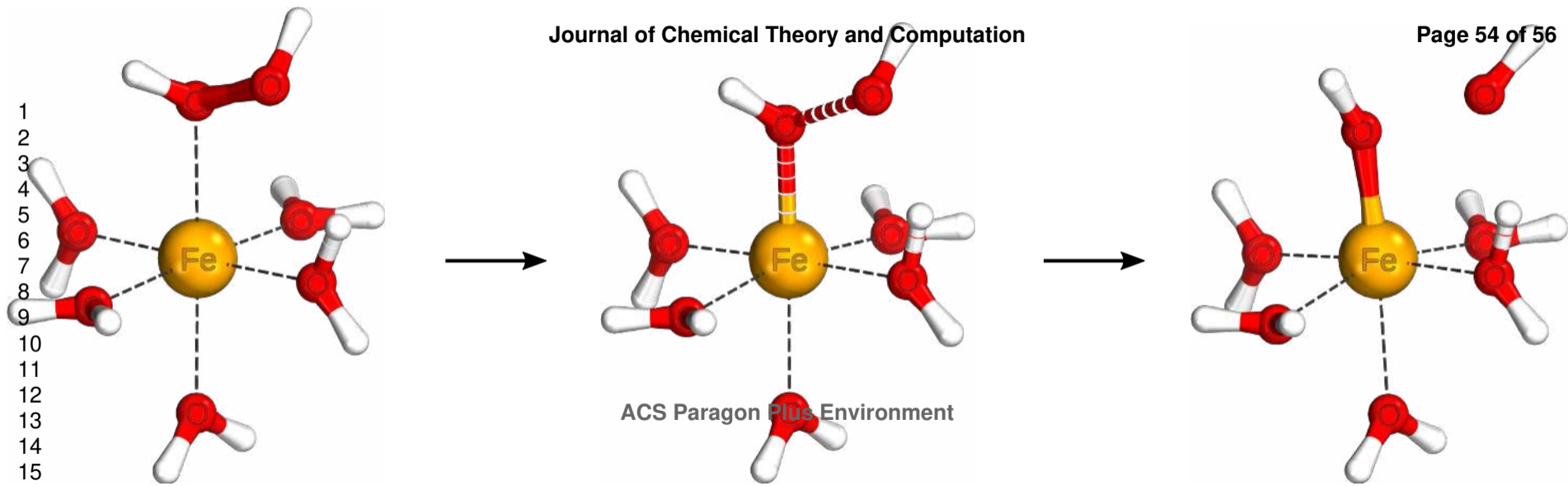


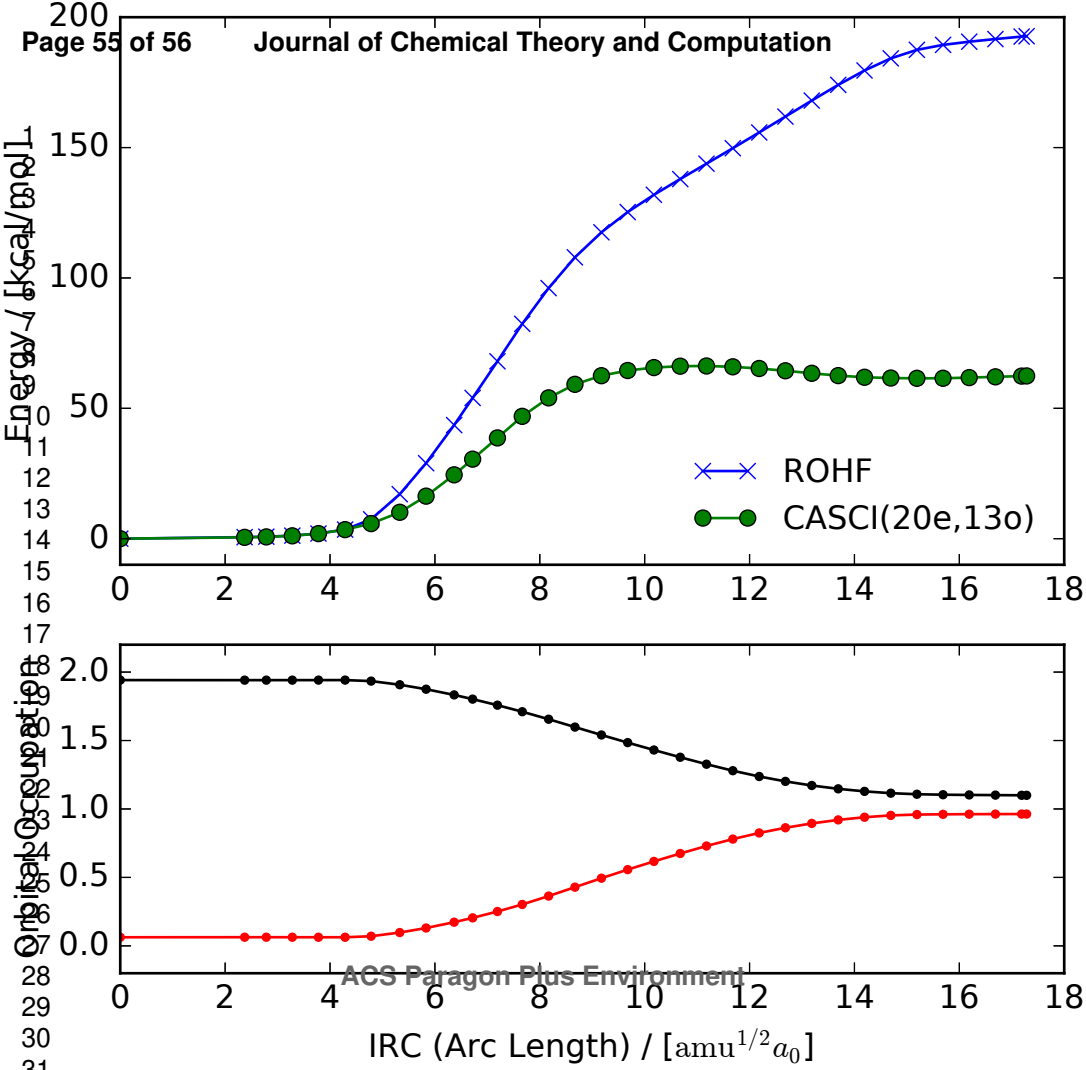
unoccupied in Φ

doubly occupied in Φ

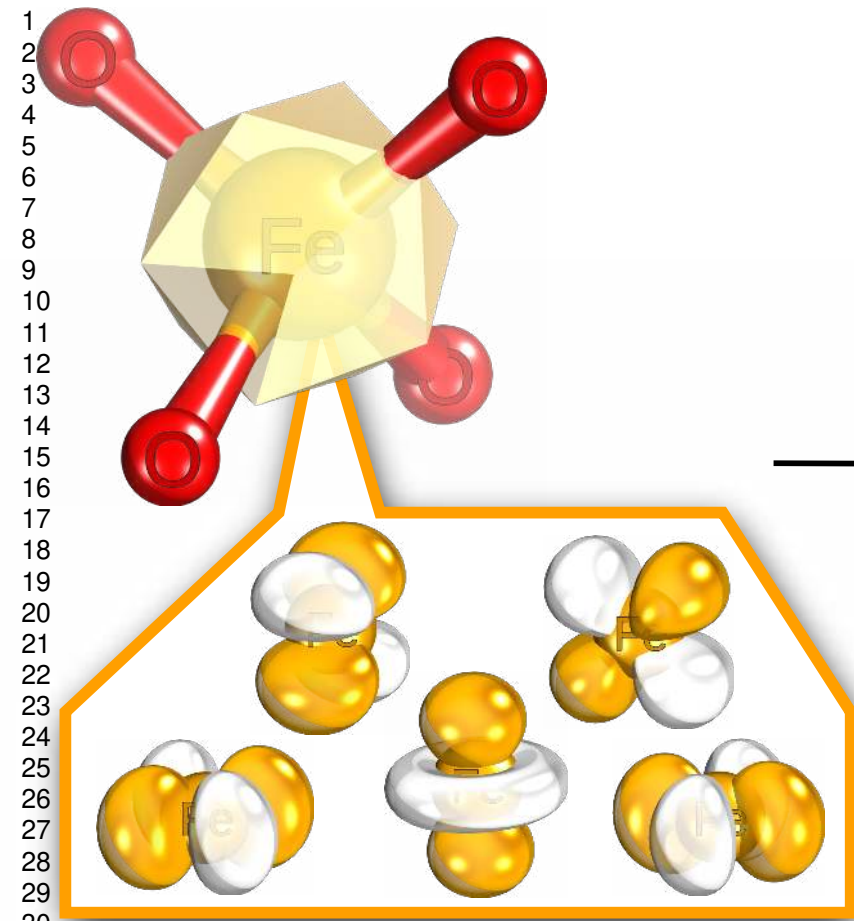


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41unoccupied in Φ 42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60doubly occupied in Φ 

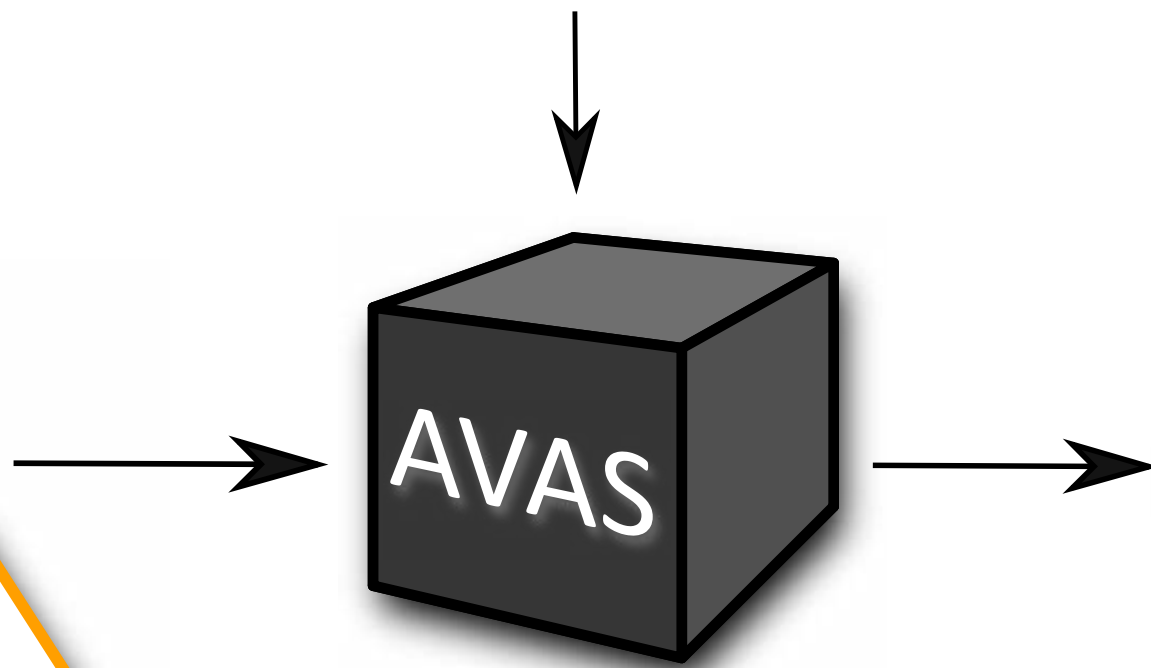




1) Select Target AOs



2) Compute $\Phi_{\text{HF/KS}}$



ACS Paragon Plus Environment

3) Get Active Space

