Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

Nan He Date

Develop multireference quantum embedding theories for chemical applications

By

Nan He
Doctor of Philosophy

Department of Chemistry

Francesco A. Evangelista, Ph.D.
Advisor

James T. Kindt, Ph.D.
Committee Member

Joel M. Bowman, Ph.D.
Committee Member

Accepted:

Kimberly Jacob Arriola, PhD, MPH
Dean of the James T. Laney School of Graduate Studies

Date

Develop multireference quantum embedding theories for chemical applications

By

Nan He
B.Sc. Chemistry, Fudan University, Shanghai, China, 2014
M.Sc. Physical Chemistry, Fudan University, Shanghai, China, 2017

Advisor: Francesco A. Evangelista, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Department of Chemistry
2022

Abstract

Develop multireference quantum embedding theories for chemical applications

By Nan He

Multireference computations of large-scale chemical systems are typically limited by the computational cost of quantum chemistry methods. Quantum embedding schemes are a promising way to extend multireference computations to large molecules with strong correlation effects localized on a small number of atoms. In this dissertation, we develop a series of quantum embedding schemes named active space embedding theory (ASET) for multireference computations. The schemes include mean-field active space embedding theory [ASET(mf)], a simple and automatic approach for embedding any multireference dynamical correlation method based on a frozen-orbital treatment of the environment; and second-order active-space embedding theory [ASET(2)] which improves upon mean-field frozen embedding by treating fragment–environment interactions via an approximate canonical transformation. We benchmark ASET(mf) and ASET(2) on various systems, including the N=N bond dissociation in pentyldiazene, the S0 to S1 excitation in 1-octene, and the interaction energy of the O2–benzene complex. In addition, the ASET schemes are used to study the singlet-triplet gap of p-benzyne and 9,10-anthracyne diradicals adsorbed on a NaCl surface; the inversion of CO on NaCl(100) surface. Despite their simplicity, our results show that ASET schemes are robust and sufficiently accurate, applicable when the coupling between the fragment and the environment is in the weak to medium regime. The ASET(2) explicit treatment of fragment–environment interactions beyond the mean-field level generally improves the accuracy of embedded computations. However, it becomes necessary to achieve an accurate description in several cases. This dissertation also discussed the potential future developments of ASET.

Develop multireference quantum embedding theories for chemical applications

By

Nan He
B.Sc. Chemistry, Fudan University, Shanghai, China, 2014
M.Sc. Physical Chemistry, Fudan University, Shanghai, China, 2017

Advisor: Francesco A. Evangelista, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Department of Chemistry
2022

Acknowledgments

I would like to express my deepest gratitude to my advisor, Dr. Francesco Evangelista, for his patient help and guidance throughout my five-year research. I would like to express my deepest appreciation to my committee member, Dr. James Kindt, and Dr. Joel Bowman, for their help on my research and proposal exams. Special thanks to Dr. Chenyang Li's help in the ASET(2) research and Dr. Meng Huang's assistance in CO inversion analysis. I am also thankful to Dr. Ziwei Guo, Dr. Tianyuan Zhang, Dr. Nick Stairs, and Dr. Shuhe Wang for valuable discussions and daily support. I'd like to acknowledge all group members of the Evangelista lab and the department of chemistry, it's a pleasure of working with them. Lastly, I'd like to mention my father, Jilin He, for encouraging me to study chemistry, and my mother, Qian Meng, for continuous moral support.

This work includes reused paper (Noted in each section). Permissions have been acquired from the publisher and appended in the reference.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Multireference electronic-structure theories

Electronic Schrödinger Equation Theoretical modeling has become one of the essential tools for chemists to understand properties and reactions of molecules. For problems of different scales, different theoretical methods apply. Electronic structure theories focus on the highly accurate prediction of electronic states. They are powerful tools for studying reactions of small to middle-scale molecules. Electronic structure theories are based on the time-independent electronic Schrödinger equation:

$$\hat{H}\left|\Psi\right\rangle = E\left|\Psi\right\rangle \tag{1.1}$$

where for N-electron systems, the Hamiltonian operator can be written as either

$$\hat{H} = \sum_{n=1}^{N} \hat{T}_n + \hat{V} \tag{1.2}$$

where $\hat{T}_n$ are kinetic operators and $\hat{V}$ is the potential energy function, or using second-quantized notation as

$$\hat{H} = \sum_{pq} h_p^q \hat{a}_p^\dagger \hat{a}_q + \frac{1}{4} \sum_{pqrs} \langle pq \| rs \rangle \, \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_s \hat{a}_r \tag{1.3}$$

The first and second term of Eq. 1.3 account for one- and two-body electron interactions, respectively. h_p^q is the one-electron integral and $\langle pq \| rs \rangle$ is the two-electron integral. While $\hat{a}^\dagger$ and $\hat{a}$ are creation and annihilation operators, respectively.

Solving the electronic Schrödinger equation is not an easy task. One way to solve this equation is by separating the variables. By separating the variables, every electron i will have its own eigenvalue and eigenfunction (ϕ_i), and the total wavefunction (Ψ) is the product of all those single electron eigenfunctions ($\Psi = \prod_{i=1}^{N} \phi_i$) However, since the N-electron wavefunction is not generally a product, this separation step is an approximation, which means that we assume that the single-electron movement is independent of the movement of other electrons. This treatment is called the mean-field approximation. Due to the antisymmetric nature of the fermions, it is more convenient to represent the N-electron wavefunction Ψ using a basis of Slater determinants:[158]

$$
\begin{aligned}
\Psi(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_N) &= \frac{1}{\sqrt{N!}}\begin{vmatrix} \phi_1(\mathbf{x}_1) & \phi_2(\mathbf{x}_1) & \cdots & \phi_N(\mathbf{x}_1) \\ \phi_1(\mathbf{x}_2) & \phi_2(\mathbf{x}_2) & \cdots & \phi_N(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(\mathbf{x}_N) & \phi_2(\mathbf{x}_N) & \cdots & \phi_N(\mathbf{x}_N) \end{vmatrix} \\
&\equiv |\phi_1,\phi_2,\cdots,\phi_N\rangle \\
&\equiv |1,2,\ldots,N\rangle.
\end{aligned}
\tag{1.4}
$$

The next step to solving the Schrödinger equation is to make assumptions about the structure of the solution. The most straightforward form is a solution function containing only one Slater determinant. In the next section, we will discuss this approach.

Single-Reference Methods Here we will start discussing the Hartree-Fock approach, together with all the context used in the following chapters. The Hartree-Fock method utilizes the mean-field approximation mentioned above and uses the single Slater determinant as the guess to the wave function.[69, 53] The core procedure of the Hartree-Fock approach is the variational optimization of the orbital wavefunctions, which minimize the total energy. This variational optimization is equivalent to solving a set of self-consistent field equations for the Fock matrix and orbitals:[145]

$$
\hat{F}\phi_i = \varepsilon_i\phi_i. \tag{1.5}
$$

In practice, the molecular orbitals are expressed on a linear combination of atomic orbital (AO) basis functions,

$$\phi_p(\mathbf{r}) = \sum_{\mu}^{\mathrm{AO}} \chi_\mu(\mathbf{r}) C_{\mu p}, \tag{1.6}$$

where $C_{\mu p}$ denotes the coefficient matrix. Since the AOs are not orthogonal, the Hartree-Fock self-consistent field equations in the AO basis is represented as,

$$\mathbf{FC} = \mathbf{SCE}, \tag{1.7}$$

where $\mathbf{S}$ is the AO overlap matrix, and $\mathbf{E}$ is the eigenvalues (orbital energies). We can then obtain the final orbital energies $\mathbf{E}$ and corresponding orbital wavefunctions $\mathbf{C}$ by solving this general eigenvalue problem.

Methods like Hartree-Fock are referred to as single-reference (SR) methods, since they build the solution starting from a single Slater determinant. Due to the mean-field approximation, the Hartree-Fock solution ignores the electron correlation completely. However, there are different approaches to reclaim those correlations. Starting from the Hartree-Fock solution, post-SCF theories like Møller–Plesset perturbation (MP2) theory, configuration interaction (CI) theory, coupled-cluster (CC) theory, and many other theories can partially or fully recover the ignored correlation.[128, 156, 154] It is worth mentioning that besides those wavefunction theories (WFT) mentioned above, there are also other approaches to approximately solve the electronic Schrödinger equation. One alternative solution is density functional theory (DFT), which uses density instead of energy to formulate an SCF scheme. In analogy to the Hartree-Fock equation, the self-consistent scheme is formulated to the Kohn-Sham equation.[153] In this equation, free electron-gas density is treated as the zeroth-order solution, and the higher-order interactions are added by including local density, the gradient of local density, and other corrections to the exchange-correlation functional. Though the assumption of DFT is different from the mean-field approximation, DFT is still generally considered a single-reference method.[40]

Though by increasing the order of CI and CC, single reference methods can approach the full numerical solution of the electronic Schrödinger equation, the cost is forbiddingly expensive. In practice, we generally truncate theories to explicit two-body interactions. The coupled-cluster with singles, doubles, and perturbative triples (CCSD(T)) is regarded as the "golden standard" in computational chemistry.[154]

Single-reference methods are cheap and straightforward. Those WFT and DFT methods have been used in various applications. However, the computational modeling of chemical processes that involve the formation and breaking of bonds, transition metal catalysts, and electronically excited states remains challenging. In these cases, a single-reference (mean-field) description of electron correlation cannot correctly describe the multideterminantal nature of the wave function and often yield qualitatively in-correct predictions.[117] A minimal example is the H_2 bond dissociation. When H_2 is at its equilibrium bonding distance, a single Slater determinant $|1\underline{1}00\rangle$ (underline marks spin-down) is a good approximation since other determinants are not important. However, when H_2 is stretched to the dissociation distance, any one of $|1\underline{1}00\rangle$, $|100\underline{1}\rangle$, $|0\underline{1}10\rangle$, and $|001\underline{1}\rangle$ is not a good approximation. In this case, the more physical electron configuration should be a linear combination of these determinants.

Multi-Reference Methods Many multireference (MR) methods have been developed to overcome the limitations of single-reference methods. One of the basic approach is the multi-configurational self-consistent-field (MCSCF) method.[144, 146, 183] In MCSCF, a linear combination of Slater determinants forms the configuration state function (CSF). During the SCF iterations, we not only optimize the MO coefficients but also the coefficients of configuration state functions.

The cost of multireference methods depends on the number of determinants. A general practice to apply multireference theories is to select a set of MO as active orbitals and the electrons inside as active electrons (Sometimes their combination is called "active space"), among which all possible determinants are treated explicitly, while determinants involving inactive orbitals/electrons are ignored. For n active orbitals and N active electrons

[CAS(n,N)] with S total spin, there are

$$N_{\text{dets}} = \binom{n}{N/2+S}\binom{n}{N/2-S}, \tag{1.8}$$

determinants, and

$$N_{\text{CSF}} = \frac{2S+1}{N+1}\binom{n+1}{N/2-S}\binom{n+1}{N/2+S+1}. \tag{1.9}$$

configuration state functions. This N_{CSF} is the cost factor that will be multiplied by the computational cost of single-reference counterparts. It can be seen that the cost grows exponentially with the size of the active space. Therefore, we choose the minimal possible active spaces for the problem we study in practice. The active space should capture the strong electron correlation but keep its size small. The electron correlation not captured by an active space, mainly dynamical correlation, will then be corrected using perturbation theory,[52, 59, 5, 6] configuration interaction,[167, 168, 169] coupled cluster theory,[134, 121, 85, 130, 93, 44] and other alternatives.[124, 162, 22, 56, 113, 142] Their cost is also scaled both with the size of the active space (n and N) and the number of inactive orbitals (core C and virtual V). Additionally, it is worth mentioning that numerous generalizations of single-reference methods have been proposed that are applicable beyond the domain of conventional single-reference methods.[140, 87, 97, 170, 155, 36]

1.2 Quantum embedding theories

As mentioned in the previous section, the computational cost of multireference methods is generally higher than that of single-reference methods. In many cases, this vital limitation prevents applications to real-world chemical problems involving large transition metal complexes, solvated species, enzyme reaction centers, and molecules adsorbed on surfaces. Nonetheless, in complex systems, the central quantity of interest is often a local property, such as a reaction barrier for rearranging a small number of atoms or an electronic excitation localized to a chromophore. These situations can often be accurately described by an embedding scheme that partitions the entire system into a set of atoms or localized orbitals (fragment, A) treated at a higher level of theory and a complementary set (environment, B) treated with a lower-level method. Such a multi-level treatment can significantly reduce the computational cost, and it offers an avenue for large-scale multireference computations on complex chemical systems.

Embedding schemes mainly vary in three aspects: i) the way a system is partitioned into a fragment and the environment, ii) the level at which individual subsystems are treated, and iii) the approximation of the interaction between the fragment and the environment. Among the earliest embedding methods, classical schemes treat the environment and its interaction with the fragment at the level of molecular mechanics (MM).[166, 33, 81, 143, 160, 135, 151, 136] In QM/MM approach, the fragments are generally computed with QM theory adding an MM-level external potential field generated by the environment. For example, if the environment is characterized as point charges, the fragment Hamiltonian is formulated as

$$h_i^{\text{QM-MM}} = h_i^{QM} - \sum_{j \in \mathrm{L}_1} \frac{q_j}{\mathbf{r_i} - \mathbf{R_j}}, \tag{1.10}$$

where h_i is the Hamiltonian components of the ith electron, q_j are environment point charges, $\mathbf{r_i}$ and $\mathbf{R_j}$ are their corresponding positions. In practice, to keep the fragment and the environment separated, any bond between the fragment and the environment will be broken, and the dangling electrons are capped using hydrogen.[166, 33] Classical QM/MM approaches are suitable for massive systems like protein and solute-solvent systems. However, their rough treatment and uncontrolled approximation of the fragment-environment boundaries leads to inaccuracy description of fragment-environment electronic interactions. Quantum embedding schemes improve upon classical approaches by treating both the fragment and the environment quantum-mechanically (QM).[114, 165] In quantum embedding, the wavefunction or density of both the fragment and the environment are generated explicitly. Among the simplest quantum embedding schemes one counts density embedding based on density functional theory (DFT) pioneered by Carter and co-workers,[64, 89, 88, 79, 78] and the more recent projector-based embedding methods based on DFT.[122, 63, 105, 106, 73, 26, 74] In the basic DFT embedding schemes, the nonadditive kinetic potential (NAKP) enforces Pauli exclusion between the electrons of fragments and the environment.[80] However, this approach will break down when fragment-environment densities significantly overlap with each other, limiting the usage of basic density embedding schemes to weakly interacting systems.[105] The projector-based embedding solves this problem by creating mutually orthogonal fragment and environment MO orbitals.[105] A simple way to enforce the orthogonality is to introduce a level-shift operator to the environment's local orbitals:

$$\mu \mathbf{P}^B = \mu \mathbf{S} \gamma^B \mathbf{S}, \tag{1.11}$$

where μ is a positive scalar, $\mathbf{S}$ and γ^B are AO overlap matrix and environment one-body density matrix, respectively. It is worth mentioning that there are many variants on how to construct $\mathbf{P}^B$ and how to localize MOs.[73, 26, 74] If WFT methods are used for the fragment and DFT for the environment (WFT-in-DFT), the fragment one-electron Hamiltonian will be formed as:

$$\mathbf{h}^{\text{A-in-B}}(\gamma^A, \gamma^B) = \mathbf{h} + \mathbf{v}_{emb}(\gamma^A, \gamma^B) + \mu \mathbf{P}^B, \tag{1.12}$$

where $\mathbf{h}$ is the standard one-electron hamiltonian, and $\mathbf{v}_{emb}$ is the embedding potential from the corresponding DFT theory. Higher level WFT computations are then performed using $\mathbf{h}^{\text{A-in-B}}$ as the effective one-body Hamiltonian.

Other quantum embedding methods focuse on capturing the electron correlation between fragments and the environment to higher levels. Green's function-based embedding schemes, such as dynamical mean-field theory (DMFT)[194, 58, 94] and self-energy embedding (SEE)[101] generalize static schemes and can be used in situations where the fragment and the environment are strongly coupled, both in applications to molecular systems and solid-state problems. DMFT features a self-consistent scheme involving the fragment (impurity) local Green's function $\mathbf{G}(\mathbf{R}_0, \omega)$ and local self energy $\Sigma(\omega)$ via a hybridization $\Delta(\omega)$. Details of this scheme can be found in Ref. 194. This scheme is powerful at capturing strong electron correlation. However, its complexity and cost make its application difficult. Density matrix embedding theory (DMET),[190, 92, 91, 18] is a simpler and more economical alternative to DMFT in which self-consistent boundary conditions are imposed on the one-body density matrix. DMET was formally justified using a Schmidt decomposition of the wave function, and it is usually implemented as an iterative scheme. The wavefunction is factorized into the fragment (A) and the environment (B) using a Singular Value Decomposition (SVD) as:

$$|\Psi\rangle = \sum_{\alpha}^{N_A} \lambda_\alpha \, |\tilde{A}_\alpha\rangle \otimes |\tilde{B}_\alpha\rangle \tag{1.13}$$

where N_A is the fragment basis size, $\tilde{A}_\alpha$ and $\tilde{B}_\alpha$ are bath wavefunction transformed using the SVD transformation matrices. Note that only additional N_A orbitals from the environment are needed to form the embedding bath in this orthogonalization scheme. The self-consistent scheme enforces the condition

$$\min ||D_{A+B}^{\text{bath, low}} - D_{\tilde{A}\tilde{B}}^{\text{bath, high}}|| \tag{1.14}$$

where $D_{A+B}^{\text{bath, low}}$ is the density evaluated using low-level theory for the fragment and the environment limited to the embedding bath; while $D_{\tilde{A}\tilde{B}}^{\text{bath, high}}$ is the density in the embedding bath computed using high-level theory. Iteratively minimizing the density differences gives the final density, and the embedding energy is then evaluated using this density and the high-level theory.

The above mentioned quantum embedding schemes, together with many other variants,[20, 32, 179, 193, 54, 39, 28, 50, 192] can be combined with both density functional theory and wave function approaches; their usefulness has been demonstrated in several recent applications.[74, 2, 16, 184, 133, 21, 98, 61, 14, 89, 88, 82, 57]

Most of the embedding approaches mentioned above focus on treating the fragment with a high-level single-reference method, with a few exceptions.[60] In contrast, embedding schemes explicitly tailored for multireference methods have received less attention. Pham, Bernales, and Gagliardi,[139] recently proposed a complete-active-space DMET (CAS-DMET) that generalizes the original formalism of Knizia and Chan to a multireference embedding approach. This goal is achieved by partitioning the active and inactive orbitals inside the DMET fragment and then performing a complete active space self-consistent field (CASSCF) computation as a higher-level method. Following the traditional DMET treatment,[91] the DMET environment (the "bath") is identified using a mean-field (single determinant) description of the union of the fragment and environment. However, this CAS-DMET strategy was found to be problematic in dissociation limit, yielding qualitatively inaccurate dissociation curves. In a follow-up paper,[76] Hermes and Gagliardi overcame this problem by replacing the low-level mean-field description with one based on a localized active space self-consistent field (LASSCF). LASSCF corresponds to a simplified version of CASSCF in the DMET framework, where the active space is a direct product of active spaces on fragments, and orbitals are optimized self consistently. The LASSCF-based DMET elevates the description of the bath to a multi-determinantal level, which provides better results for bond dissociation curves. A recent work by Coughtrie and co-workers[30] proposed an elegant and straightforward frozen-orbital embedding scheme that focused on post-MCSCF calculations. In this scheme, the CASSCF reference is optimized for the entire system (fragment + environment). Dynamical correlation in the fragment was treated with internally-contracted multireference coupled-cluster theory (icMRCC),[67, 45], and an additive scheme based on multireference second-order perturbation theory (MRPT2)[52, 6] was used to account for the interaction and dynamical correlation in the environment.

1.3 Limitations of current embedding theories

An essential problem in the generalization of mean-field-based embedding schemes to multireference theories is reconciling the conventional partitioning of orbitals according to a fragment-based criterion (e.g., as defined in DMET) with the active space partitioning adopted in multireference calculations. In mean-field-based embedding methods, molecular orbitals are partitioned into two separate spaces that span the fragment and environment, with each of these divided into occupied and virtual sets. In approaches based on a mean-field picture, this step is commonly done by first localizing occupied and virtual orbitals separately and then assigning them to the fragment or environment according to some property, e.g., partial atomic charges.[73, 26, 180, 27, 15] Virtual space projection and truncation techniques have been explored to analyze excited-state orbitals.[116, 35] These techniques are also applied in embedding, like projected atomic orbitals (PAOs),[141] intrinsic atomic orbitals (IAOs),[90] or concentrically local orbitals,[28] often applied to identify a set of virtual orbitals that span the fragment. However, the orbital space is further partitioned into the core, active, and virtual spaces in multireference computations. In most cases, this separation is done manually, though there are techniques that can automate this step.[161, 7, 148, 41, 147] Previous work on CAS-DMET[139] suggests that orbital localization and projection based on a mean-field reference[91] is suboptimal since the resulting orbitals are not variationally optimal. This observation suggests that the use of a correlated reference (e.g., MCSCF) may be a better starting point for multireference embedding schemes.

The other significant problem is the correlation treatment. A common feature of quantum embedding methods is to approximate the fragment-environment interaction with a static or frequency-dependent effective one-body term (for some exceptions see Refs. 49, 132, and 152). However, a one-electron treatment of the fragment-environment interaction ignores cross fragment-environment correlation and pure environment correlation effects and may neglect orbital relaxation effects if the environment is frozen. Consequently, an embedding approach could lead to inaccurate results when the orbital response or correlation effects contribute significantly to differential properties of the fragment, for example, in excited states or weakly bound systems. The orbital response may be captured using methods like DFMT, DMET, and SEET, but this generally involves a self-consistent iterative procedure that raises the risk of numerical issues related to the convergence of these procedures. Within a given embedding scheme, the only way to systematically reduce the error introduced by this approximation is to make the fragment larger so as to include more atoms into the higher-level computation.

1.4 Structure of this dissertation

In this dissertation, we develop a set of systematically improvable active space embedding theories (ASET). Chapter 2 focuses on the orbital projection and the simplest form of ASET, which we denote as ASET(mf), a zeroth-order embedding scheme based on a mean-field treatment of the environment. In Chapter 3, a theory to systematically improve ASET is presented, and an ASET(2) scheme is formulated using the perturbation theory. In Chapter 4, ASET is applied to a practical chemistry problem: CO inversion on NaCl(100). In Chapter 5, some ongoing works and possible further developments of ASET are discussed.

Chapter 2

Mean-field Active Space Embedding Theory: ASET(mf)

2.1 Introduction

This chapter discusses the simplest form of ASET, ASET(mf). This chapter is mainly based on a published work; see Ref. 71. (Publisher: AIP; Publishing Date: Mar 7, 2020; Rights managed by AIP Publishing.)

ASET(mf) is a simple non-iterative frozen-orbital embedding scheme that assumes a zeroth-order CASSCF or MCSCF reference as a starting point that correctly describes the qualitative features of the entire system under study. Like in the embedding scheme of Coughtrie and co-workers,[30] in ASET(mf), the active orbitals from a CASSCF/MCSCF reference are treated separately and kept fixed. However, instead of using intrinsic bonding orbitals (IBOs)[90] and PAOs[141] to define occupied and virtual orbitals, our approach employs a generalization of the DMET orbital partitioning scheme. As our study shows, the use of a correlated reference to partition the orbitals significantly improves the convergence of properties with respect to the number of fragment orbitals. Given an MCSCF reference, the ASET(mf) scheme is designed so that all steps are fully automated. Our procedure treats the interaction of the fragment and the environment via a frozen-orbital embedding. However, future extensions of this method may be devised in which a low-level correlated treatment is extended to the environment (Chapter 3) or in which the environment is treated at the DFT level. We note that a similar single-reference version of frozen-orbital embedding based on a mean-field reference was also recently studied by Knizia[102] and Claudino and Mayhall.[27]

We combine ASET(mf) with the driven similarity renormalization group multireference second-order perturbation theory (DSRG-MRPT2) recently developed in our group,[107, 109] which we use to describe correlation effects in the fragment. Several benchmark computations are presented to test the accuracy of the ASET(mf). We study the C-C dissociation curve in ethane interacting with a methane molecule, and the N=N bond-breaking curve in pentyldiazene, as examples of problems with weak and strong coupling to the environment. To demonstrate the potential of this embedding scheme, we apply it to study the single-triplet gap of small arynes[181] adsorbed on a NaCl surface, which was recently examined experimentally by

Schuler and *et al.* [150] in a study of Bergman cyclization.

This chapter is structured as follows. In Sec. 2.2, we introduce the details of embedding orbital partitioning based on mean-field and MCSCF references. In Sec. 2.3, we go over the ASET(mf) equations and implementations. In Sec. 2.4, we benchmark ASET(mf) using bond-breaking reactions and apply it to study the singlet-triplet gap of *p*-benzyne and 9,10-anthracyne adsorbed on a NaCl surface. Finally, in Sec. 2.6, we summarize this chapter and discuss extensions of ASET(mf).

2.2 Orbital Partition

The ASET(mf) procedure generally starts with a CASSCF computation on the full system. In this section, we will discuss the partition procedure of ASET(mf), which is used to partition inactive orbitals into the fragment and environment. We assume that the molecular orbitals $\{\phi_p(\mathbf{r}), p = 1, \ldots, N\}$ are represented by a linear combinations of an atomic orbitals (AO) basis $\{\chi_\mu(\mathbf{r}), \mu = 1, \ldots, K\}$,

$$\phi_p(\mathbf{r}) = \sum_{\mu}^{\mathrm{AO}} \chi_\mu(\mathbf{r}) C_{\mu p}, \tag{2.1}$$

where $C_{\mu p}$ denotes the coefficient matrix. The CASSCF orbitals are subdivided into core (**C**), active (**A**), and virtual (**V**) sets. In the embedding procedure, we rotate and partition the core and virtual orbitals into subspaces belonging to the fragment (A) and the environment (B). To define a projector onto the fragment space, the user selects a list of atoms that belong to this set. Based on this choice, a projector can be constructed as

$$\hat{P}^{\mathrm{A}} = \sum_{\mu,\nu\in\mathrm{A}} |\chi_\mu\rangle\, S^{-1}_{\mu\nu}\, \langle\chi_\nu| , \tag{2.2}$$

where the summation is restricted to AOs χ_μ and χ_ν centered on atoms in A and the $S_{\mu\nu} = \langle\chi_\mu|\chi_\nu\rangle$ are AO overlap matrix elements. Following the DMET partitioning procedure,[190] we perform separate unitary rotations of core and virtual orbitals to maximize the overlap with the fragment projector. To this end, we form the overlap matrix of the MOs projected onto A (P^{A}_{pq})

$$P^{\mathrm{A}}_{pq} = \langle\hat{P}^{\mathrm{A}}\phi_p|\hat{P}^{\mathrm{A}}\phi_q\rangle = \langle\phi_p|\, \hat{P}^{\mathrm{A}}\, |\phi_q\rangle\, . \tag{2.3}$$

and separately solve the following eigenvalue problem for the core-core and virtual-virtual blocks of P^{A}_{pq},

$$\sum_q P^{\mathrm{A}}_{pq} U_{qr} = \lambda_r U_{pr} \quad p,q \in \mathbf{X}, \tag{2.4}$$

where $\mathbf{X} \in \{\mathbf{C},\mathbf{V}\}$. The eigenvectors U_{qp} define unitary transformations onto the basis of rotated MOs

$$\phi'_p(\mathbf{r}) = \sum_q \phi_q(\mathbf{r}) U_{qp}, \tag{2.5}$$

while the eigenvalues $\lambda_p = \langle\phi'_p|\, \hat{P}^{\mathrm{A}}\, |\phi'_p\rangle$ lie in the range $\lambda_p \in [0,1]$ and measure the overlap of the density $|\phi'_p|^2$ with the fragment. In DMET, the standard partitioning procedure assigns orbitals ϕ'_p that satisfy the condition $\lambda_p > \tau$ to the fragment, where τ is a given threshold. By the end of this procedure the orbitals are partitioned into five sets: 1) fragment doubly occupied ($\mathbf{C}_\mathrm{A}$), 2) fragment active ($\mathbf{A}_\mathrm{A}$), 3) fragment virtual ($\mathbf{V}_\mathrm{A}$), 4) environment occupied ($\mathbf{O}_\mathrm{B}$), and 5) environment virtual ($\mathbf{V}_\mathrm{B}$). To include cases where the environment contains unpaired electrons, the environment occupied ($\mathbf{O}_\mathrm{B}$) orbitals are not necessarily assumed to be doubly occupied. Since after rotation and partitioning the orbitals are no longer canonical, we separately diagonalize each of the five diagonal blocks of the averaged Fock matrix This partitioning is illustrated in Fig. 2.1.

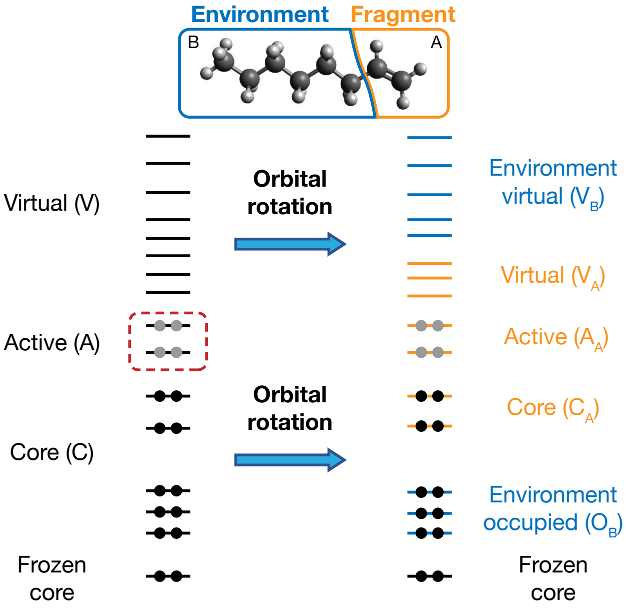


Figure 2.1 Illustration of the orbital partitioning employed in the active space embedding scheme (ASET(mf)).

After rotation and partitioning the orbitals we canonicalize the orbitals by separately diagonalizing each of the five diagonal blocks of the averaged Fock matrix

$$f_p^q = h_p^q + \sum_{rs} \langle pr \| qs \rangle \gamma_r^s, \tag{2.6}$$

where h_p^q are one-body integrals, $\langle pq \| rs \rangle$ are standard antisymmetrized two-electron integrals, and γ_r^s is the CASSCF one-body density matrix. Orbitals in each of the five subspaces are rotated according to the unitary transformation that diagonalizes each block of f_p^q. Frozen orbitals (both doubly occupied and virtual) are not projected onto the fragment but are semi-canonicalized independently. Our tests show that semi-canonicalization of frozen orbitals has no impact on the embedding energy, but could be important for the formulation of energy gradients.[106]

Here we show several examples to clarify this procedure. Firstly, we partition a minimal case with four hydrogen atoms step-by-step, using the first two H atoms as the fragment and the other two as the environment. The Hartree-Fock solution of H_4 on the STO-3G basis consists of four MOs:

MO matrix	MO 1	MO 2	MO 3	MO 4
κ_1	0.2469	0.5380	0.8625	0.8185
κ_2	0.4056	0.3939	-0.6014	-1.4914
κ_3	0.4056	-0.3939	-0.6014	1.4914
κ_4	0.2469	-0.5380	0.8625	-0.8185

The partition procedure needs to rotate and find 2 of those four orbitals that belong to fragment A. To evaluate the projector $\hat{P}^{\mathrm{A}} = \sum_{\mu,\nu\in\mathrm{A}} |\chi_\mu\rangle S_{\mu\nu}^{-1} \langle\chi_\nu|$, the first step it to compute $S_{\mu\nu}^{-1}$. In H_4, the AO overlap matrix is:

S	κ_1	κ_2	κ_3	κ_4
κ_1	1.0	0.6208	0.2221	0.0596
κ_2	0.6208	1.0	0.6208	0.2221
κ_3	0.2221	0.6208	1.0	0.6208
κ_4	0.0596	0.2221	0.6208	1.0

Since 1 and 2 are the fragment sites, the $S^{-1}_{\mu\nu}$ is:

$S^{-1}_{\mu\nu}$	κ_1	κ_2
κ_1	1.6271	-1.0101
κ_2	-1.0101	1.6271

Therefore, the projector $\hat{P}^{\mathrm{A}}$ is

$$\hat{P}^{\mathrm{A}} = 1.63\,|\kappa_1\rangle\,\langle\kappa_1| - 1.01\,|\kappa_1\rangle\,\langle\kappa_2| - 1.01\,|\kappa_2\rangle\,\langle\kappa_1| + 1.63\,|\kappa_2\rangle\,\langle\kappa_2|. \tag{2.7}$$

For all MOs (Eq. 2.1), the projector matrix elements will be:

$$P_{pq} = 1.63\sum_{i,j}^{A+B} c_i\,\langle\kappa_i|\,|\kappa_1\rangle\,\langle\kappa_1|\,|\kappa_j\rangle\,c_j \tag{2.8}$$

$$-\,1.01\sum_{i,j}^{A+B} c_i\,\langle\kappa_i|\,|\kappa_1\rangle\,\langle\kappa_2|\,|\kappa_j\rangle\,c_j \tag{2.9}$$

$$-\,1.01\sum_{i,j}^{A+B} c_i\,\langle\kappa_i|\,|\kappa_2\rangle\,\langle\kappa_1|\,|\kappa_j\rangle\,c_j \tag{2.10}$$

$$+\,1.63\sum_{i,j}^{A+B} c_i\,\langle\kappa_i|\,|\kappa_2\rangle\,\langle\kappa_2|\,|\kappa_j\rangle\,c_j. \tag{2.11}$$

A matrix representation of the procedure is:

$$\mathbf{P} = \mathbf{C}^T \mathbf{S}^{A+B,A} (\mathbf{S}^{A,A})^{-1} \mathbf{S}^{A,A+B} \mathbf{C} \tag{2.12}$$

The resulting **P** matrix for the H_4 embedding is:

P	ϕ_1	ϕ_2	ϕ_3	ϕ_4
ϕ_1	0.7561	0.3620	-0.1540	-0.1721
ϕ_2	0.3620	0.4432	0.3087	0.1431
ϕ_3	-0.1540	0.3087	0.5731	0.3545
ϕ_4	-0.1721	0.1431	0.3545	0.2275

After diagonalization, the eigenvalues of this matrix are:

Eigs	λ
$\phi_1^{'}$	0.9940
$\phi_2^{'}$	0.2053
$\phi_3^{'}$	0.7947
$\phi_4^{'}$	0.0060

If we choose a threshold $t = 0.5$, then $\phi_1^{'}$ and $\phi_3^{'}$ are the fragment (A) orbitals, and $\phi_2^{'}$ and $\phi_4^{'}$ are the environment (B) orbitals.

Next, we show a more sizable example containing an active space. In Fig. 2.2, we show the result of the partitioning procedure applied to 1-butene using a minimal basis set. In this example, the active orbitals are chosen to be the π/π^* orbitals, and the fragment contains the $HC_1{=}C_2$ group. The resulting $\mathbf{A}_\mathrm{A}$ space contains the π/π^* orbitals, while the $\mathbf{C}_\mathrm{A}$ and $\mathbf{V}_\mathrm{A}$ include core and valence orbitals of the $HC_1{=}C_2$ group, together with orbitals that extend slightly onto carbon C_3.

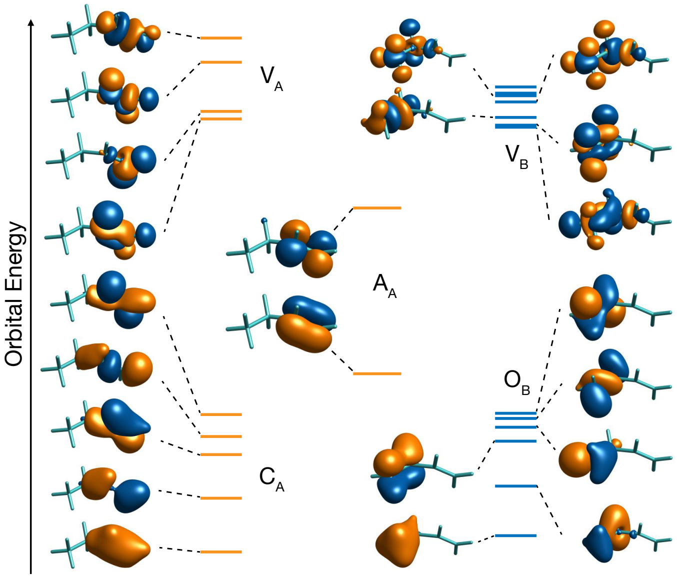


Figure 2.2 ASET(mf) embedding orbitals of 1-butene computed with $\tau = 0.5$ and the STO-3G basis. Orbitals with $\lambda_p \geq \tau$ are assigned to the fragment (A) while orbitals with $\lambda_p < \tau$ are assigned to the environment (B). See Fig. 2.1 for the definition of the orbital spaces.

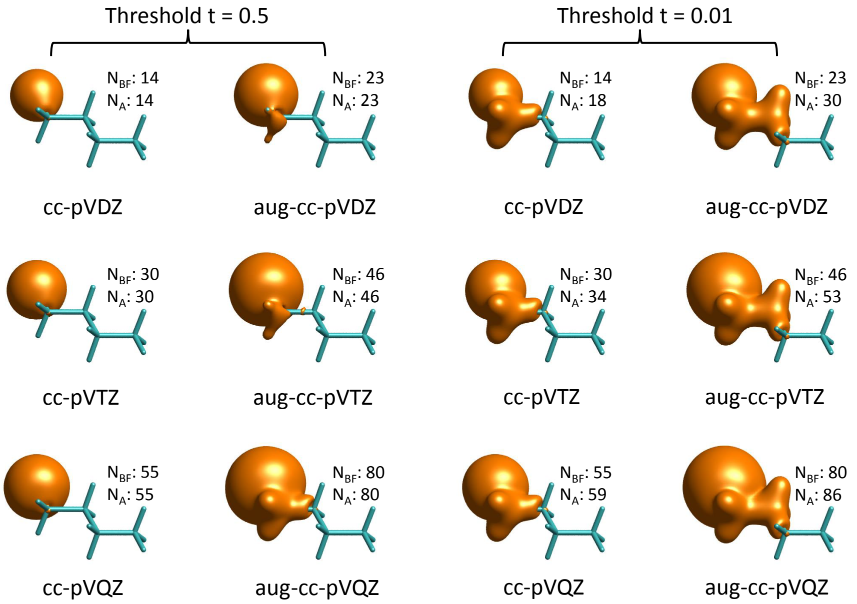
One potential concern of this partitioning procedure is that by optimizing the overlap of the fragment orbitals with the fragment atomic orbital basis, a dependence on the basis set is introduced. To investigate this aspect of the localization procedure, we consider the 1-fluorobutane molecule and include the F atom in the definition of the fragment. We then use the sum of all the fragment orbitals squared, $\rho^{\mathrm{A}}(\mathbf{r}) = \sum_p |\phi_p^{\mathrm{A}}(\mathbf{r})|^2$, as a proxy for the size of the fragment orbital space. In Fig. 2.3 we plot $\rho^{\mathrm{A}}(\mathbf{r})$ for a series of Dunning basis sets[189] using two threshold values for partitioning the orbitals: $t = 0.5$ (default) and 0.01. With the default choice of the threshold, the number of fragment orbitals (N_A) is identical in all cases to the number of basis functions on the fragment atoms (N_BF). However, we also observe that $\rho^\mathrm{A}(\mathbf{r})$ slightly extends to nearby environment atoms as the basis set increases, especially when diffuse functions are added. The smaller threshold value leads instead to selecting more fragment orbitals than those included in the fragment atomic basis, and $\rho^\mathrm{A}(\mathbf{r})$ across up to three bonds in the largest basis set. These results suggest that fragment selection with the default threshold is generally robust. One way to improve upon the current partitioning approach would be to truncate the virtual fragment space favoring orbitals that contribute the most to capturing the dynamical correlation of the fragments. In Sec. 2.4, the impact of different basis sets on the embedding computation results is presented and discussed further.

2.3 The ASET(mf) scheme

After the projection procedure, the final orbitals are partitioned into five sets: the core/active/virtual orbitals of the fragment ($\mathbf{C}_\mathrm{A}$, $\mathbf{A}_\mathrm{A}$, and $\mathbf{V}_\mathrm{A}$) and the occupied/virtual orbitals of the environment ($\mathbf{O}_\mathrm{B}$ and $\mathbf{V}_\mathrm{B}$). For the convenience of the following discussions, these five orbital spaces and the corresponding indices used to label MOs are shown in Fig. 2.4.

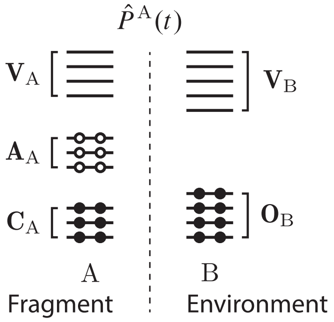
The zeroth-order picture adopted in ASET corresponds to the wave function ansatz

$$|\Psi_{\mathrm{A+B}}^{\mathrm{ASET(mf)}}\rangle = |\Psi_\mathrm{A}\rangle \otimes |\Phi_\mathrm{B}\rangle \tag{2.13}$$



N_{BF}: number of basis functions on the fragment
N_A: number of orbitals partitioned into the fragment (including frozen core)

Figure 2.3 Sum of the fragment orbitals squared, $\rho^{\mathrm{A}}(\mathbf{r}) = \sum_p |\phi_p^{\mathrm{A}}(\mathbf{r})|^2$, computed for 1-fluorobutane using various basis sets and two values of the partitioning threshold. All plots of $\rho^{\mathrm{A}}(\mathbf{r})$ use an isocontour value equal to 0.05.



Orbital Labels

A	u, v, w, x, y, z
$\mathbf{O}_{\mathrm{B}}$	m, n
$\mathbf{V}_{\mathrm{B}}$	e, f
$\mathrm{A} \cup \mathbf{O}_{\mathrm{B}}$	i, j, k, l
$\mathrm{A} \cup \mathbf{V}_{\mathrm{B}}$	a, b, c, d
$\mathrm{A} \cup \mathrm{B}$	p, q, r, s

Figure 2.4 Molecular orbital spaces and indices used in this dissertation.

where Ψ_{A} is the fragment wave function and Φ_{B} is the environment wave function, approximated with the Slater determinant

$$|\Phi_{\mathrm{B}}\rangle = \prod_{m}^{\mathbf{O}_{\mathrm{B}}} \hat{a}_{m}^{\dagger} |-\rangle \tag{2.14}$$

To obtain the ASET(mf) Hamiltonian, we start from the total Hamiltonian of the fragment plus environment ($\hat{H}^{\mathrm{A+B}}$) in the CASSCF MO basis

$$\hat{H}^{\mathrm{A+B}} = \sum_{pq} h_{p}^{q} \hat{a}_{q}^{p} + \frac{1}{4} \sum_{pqrs} \langle pq \| rs \rangle \, \hat{a}_{rs}^{pq} \tag{2.15}$$

where products of Fermionic creation ($\hat{a}_{p}^{\dagger}$) and annihilation ($\hat{a}_{p}$) operators are compactly expressed as $\hat{a}_{rs\cdots}^{pq\cdots} \equiv \hat{a}_{p}^{\dagger}\hat{a}_{q}^{\dagger}\cdots\hat{a}_{s}\hat{a}_{r}$. Normal-ordering the Hamiltonian with respect to Φ_{B} and keeping only the operator part that involves fragment orbitals, we arrive at the ASET(mf) Hamiltonian

$$\hat{G} = E_{\mathbf{O}_{\mathbf{E/B}}} + \sum_{pq}^{\mathrm{A}} \tilde{h}_{p}^{q} \{\hat{a}_{q}^{p}\}_{\Phi_{\mathrm{B}}} + \frac{1}{4} \sum_{pqrs}^{\mathrm{A}} \langle pq \| rs \rangle \, \{\hat{a}_{rs}^{pq}\}_{\Phi_{\mathrm{B}}} \tag{2.16}$$

Here the subscript "Φ_{B}" indicates normal ordering with respect to the state $|-_{\mathrm{A}}\rangle \otimes |\Phi_{\mathrm{B}}\rangle$, where $|-_{\mathrm{A}}\rangle$ is the vacuum fragment state. The quantity $E_{\mathbf{O}_{\mathbf{E/B}}}$ is the energy contribution due to the frozen and environment occupied orbitals and $\tilde{h}_{pq}$ are matrix elements of an effective one-body operator defined as

$$\tilde{h}_{p}^{q} = h_{p}^{q} + \sum_{m}^{\mathbf{O}_{\mathrm{B}}} \langle pm \| qm \rangle \tag{2.17}$$

This effective one-electron operator accounts for the interaction of the fragment electrons with the environment electrons frozen in the mean field state Φ_{B}. The total ASET(mf) energy is obtained by performing a high-level multireference computation using the Hamiltonian $\hat{G}$. The resulting energy may be expressed as the sum of the CASSCF energy ($E_0^{\mathrm{A+B}}$) and an energy correction from the high-level MR computation [$\delta E_{\mathrm{MR}}^{\mathrm{A}}(\hat{G})$]:

$$E_{\text{ASET(mf)-[MR]}} = E_0^{\mathrm{A+B}} + \delta E_{\mathrm{MR}}^{\mathrm{A}}(\hat{G}) \tag{2.18}$$

Note that the definition of the ASET(mf) Hamiltonian depends on the CASSCF solution used to define the orbital partitioning. Therefore, there are different ways to generalize this embedding approach to computations on multiple electronic states.

Lastly, we comment on the computational cost of ASET(mf). The projector P_{pq}^{A} itself can be computed and diagonalized efficiently with a cost that scales as $\sum_X N_X^3$ where X indicates one of the five orbital spaces resulting from the partitioning of the orbitals. This estimate excludes the cost of semi-canonicalizing the orbitals (dominated by a Fock matrix build) and integral transformation. Therefore, the overall computational cost of ASET(mf) is typically negligible compared to that of high-level multireference computations on the fragment.

2.4 Modelling weak interactions and bond-breaking using ASET(mf)

Firstly, we brief the computational settings used by all applications in this section. The ASET(mf) approach was implemented in Forte,[48] an open-source package that implements several multireference electronic structure methods. CASSCF orbitals and the corresponding one- and two-electrons integrals were obtained using Psi4[159] and imported into Forte. The embedding single-point energies in Sec. 2.4 and Sec. 2.5 are computed using the cc-pVDZ basis set,[189] unless otherwise specified. The equilibrium geometries of 1-octene, ethane, and pentyldiazene used in Sec. 2.4 are optimized with Psi4 at the B3LYP/cc-pVDZ level of theory.[12, 104] Geometries of *p*-benzyne and 9,10-anthracyne optimized in vacuum at the CASSCF(2,2)-DSRG-MRPT3/cc-pVDZ level of theory (using 3-point finite-difference gradients) are taken from Ref 109. All transition energies computed in this work refer to vertical excitation, whereby all molecules are kept fixed at their respective singlet geometries. Optimized structures for *p*-benzyne and 9,10-anthracyne adsorbed on NaCl were obtained by translating the gas-phase molecular geometries along the axis normal to the surface plane while constraining the geometry of the molecules, the NaCl atoms, and the relative orientation of the two. Initial orientations for *p*-benzyne and 9,10-anthracyne on the surface were selected manually. The

DSRG-MRPT2 computations are performed using Forte and selecting the flow parameter value $s = 0.5$. The p-benzyne and 9,10-anthracyne computations employed a density-fitted implementation of the DSRG-MRPT2[68] using the JK fitted and MP2 fitted cc-pVDZ auxiliary basis.[176, 70]

Firstly, we compare the accuracy of embedded computations using our approach that preserves the CASSCF active orbitals with the conventional DMET orbital partitioning based on a mean field reference. In the mean-field scheme, occupied and virtual orbitals are first separately rotated to fragment orbitals and then partitioned into core, active, and virtual sets. As a consequence, active orbitals loose their variational character. To compare these approaches, we compute the excitation energy for the lowest excited singlet state of 1-octene. We select an active space that includes the π/π^* orbitals, and fragment A is defined by the atoms $C_1{=}C_2$ and the hydrogens bonded to them (see Fig. 2.5).

As shown in Fig. 2.5, as the number of orbitals included in A increases, the DSRG-MRPT2 excitation energy in both embedded computations approach the value of a full calculation on the entire molecule using the respective reference orbitals. The ASET(mf) scheme shows both faster convergence and a less pronounced dependence of the energy with respect to the size of the fragment space. To investigate the different convergence behavior we examine the active orbitals generated by the two approaches, which are plotted on the right of Fig. 2.5. A comparison of the two partitioning schemes shows that the π/π^* CASSCF(2e,2o) orbitals extend slightly beyond the fragment boundary. In contrast, the orbitals based on a mean-field partitioning are strictly localized within the fragment boundary. This observation suggests that the mean-field orbital partitioning requires additional orbitals in order to span the space of the variationally optimal π/π^* CASSCF(2e,2o) orbitals.

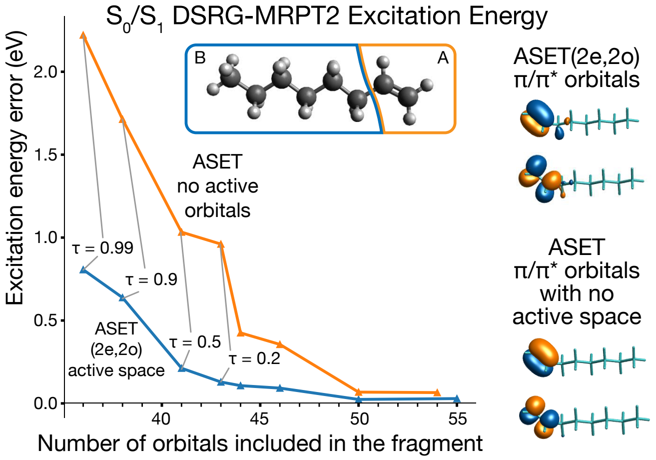


Figure 2.5 Excitation energy error of 1-octene computed with ASET(mf) and DSRG-MRPT2 (vertical axis, eV) computed with respect to the full DSRG-MRPT2 using the cc-pVDZ basis set. Comparison of ASET(mf) with an active space containing two electrons in two orbitals (2e,2o) and ASET(mf) with no active orbitals as a function of the fragment orbitals $[\dim(\mathbf{C}_\mathrm{A})+\dim(\mathbf{A}_\mathrm{A})+\dim(\mathbf{V}_\mathrm{A})]$ (horizontal axis)

.

Our second test of the ASET(mf) scheme focuses on bond-breaking processes. Key to the accurate description of bond breaking with embedding schemes is defining fragment orbitals that properly span the bonding/antibonding pairs relevant to a given process. In this section, we study two examples of bond breaking: a) C-C single-bond dissociation of an ethane molecule interacting weakly with a methane molecule, and b) N=N double-bond dissociation in pentyldiazene. These two examples test the accuracy of ASET(mf) in cases where the coupling of the fragment/environment is weak and strong, respectively. In all following computations, we manually select only the fragment atoms and select fragment orbitals using the projection described in Sec. 2.2 and the threshold $\tau = 0.5$.

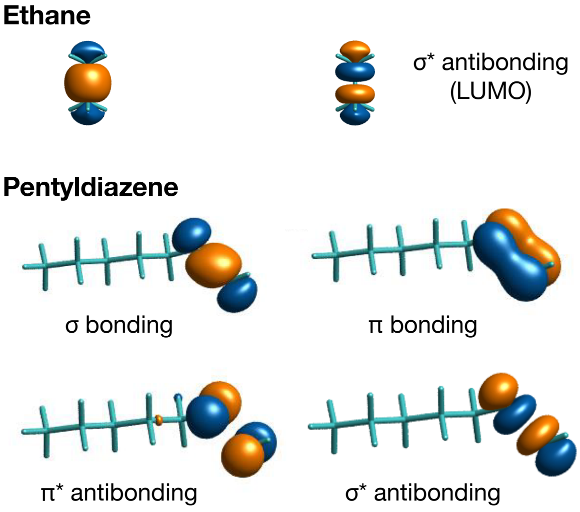


Figure 2.6 Active orbitals used for the ethane and pentyldiazene bond-breaking test cases. These orbitals are not localized and projected during the ASET(mf) computations.

In the first case, we consider a model system with one ethane molecule (fragment, A) and one methane molecule (environment, B) located 4.7 Å away from the closest ethane carbon atom. We study the dissociation of the ethane C−C bond using an active space composed of a pair of σ/σ^* C−C orbitals (shown in Fig. 2.6) along the entire dissociation path. Fig. 2.7

shows the difference between the potential energy curve computed with DSRG-MRPT2 embedded via ASET(mf) and the full DSRG-MRPT2 results. As shown in the inset of Fig. 2.7, the energy difference between these two curves is of the order of 0.2 E_h, due to the neglect of correlation on the environment in the ASET(mf) computation. However, after shifting, the differences between the two curves are found to be continuous, with a maximum deviation less than 0.03 kcal/mol. The excellent performance of ASET(mf) for this example is, as expected, due to the weak interaction between fragment and environment. We also verified that when all orbitals are included in the fragment, the embedded DSRG-MRPT2 results recover the full DSRG-MRPT2 description.

The second case we consider involves dissociation of the N=N bond in pentyldiazene ($CH_3(CH_2)_4$–N=NH), a test case similar to the one used to assess CAS-DMET.[139] This is a more challenging system since the boundary between fragment and environment cuts through a covalent C–C or N–C bond. The active space is also larger, requiring at least four orbitals: the C–C σ/σ^* and π/π^* pairs, which are plotted in Fig. 2.6. In Fig. 2.8, we show the dissociation curve computed using fragments that range from HN=N to the full pentyldiazene molecule, selecting the fragment orbitals using the ASET(mf) approach. As in the previous example, the dissociation curves for N=N are also continuous. In Table 2.1 we report the maximum absolute deviation (MAD) for each shifted curve. Even with the smallest fragment (HN=N), the MAD from the shifted full DSRG-MRPT2 curve that

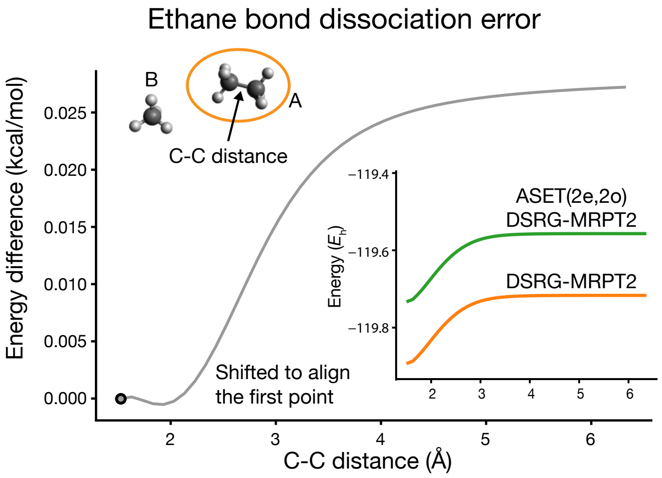


Figure 2.7 Ethane C-C dissociation curve in the presence of a weakly interacting methane molecule. Absolute energy and error for the full DSRG-MRPT2 and ASET(mf)/DSRG-MRPT2 (vertical axis, E_h for energy, kcal/mol for error) as a function of the ethane C-C bond length (horizontal axis). The energy difference curve is obtained after shifting the ASET(mf)/DSRG-MRPT2 curve so that the energy of the first point coincides with the full DSRG-MRPT2 energy. Computations use 50 geometries along the C−C bond vector and a 0.1 Å spacing.

is less than 1.5 kcal/mol. Inclusion of one and two CH_2 groups reduces this error to less than 0.4 and 0.1 kcal/mol, respectively.

In Table 2.1, we report the size of the ASET(mf) fragment orbital space for pentyldiazene together with relative timings for the corresponding DSRG-MRPT2 computations. Due to the scaling of DSRG-MRPT2 computations, the reduction in computational cost that follows from the embedding treatment is of the order of 3–10 times for the smallest three fragments. However, the advantage of the embedding will become even more significant in applications to larger molecules, or when a more expensive multireference method is employed.

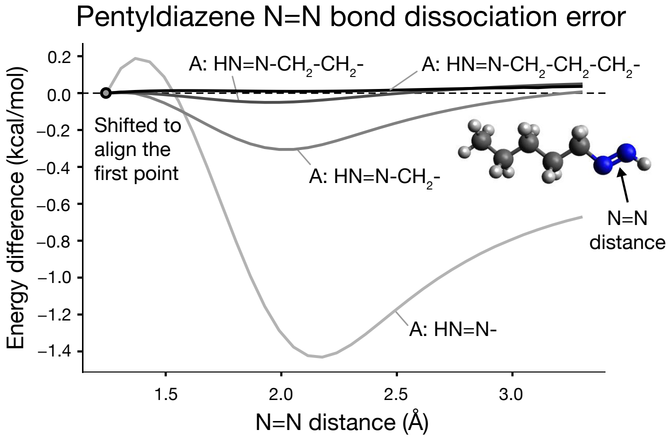


Figure 2.8 Pentyldiazene N=N dissociation curve. Absolute energy error for the full DSRG-MRPT2 and ASET(mf)/DSRG-MRPT2 (vertical axis, kcal/mol for error) as a function of the N=N bond length (horizontal axis) computed with the cc-pVDZ basis set. The energy difference curve is obtained after shifting the ASET(mf)/DSRG-MRPT2 curve so that the energy of the first point coincides with the full DSRG-MRPT2 energy. Computations use 35 geometries along the N=N bond vector.

Lastly, we discuss the accuracy of ASET(mf) and the size of the fragment orbital basis as a function of basis set size. To this end we have computed the N=N bond dissociation curve of pentyldiazene using the cc-pVXZ basis sets up to quadruple ζ quality ($X=$ D,T,Q). To reduce the cost of these computations we have employed a density-fitted implementation of the DSRG-MRPT2[68] together with the cc-pVXZ-JK and cc-pVXZ-RI auxiliary basis sets.[176, 70, 178] Potential energy curves for pentyldiazene computed with different basis sets for the HN=N and HN=N$-$C$_2$H$_4$ fragments are reported in Fig. 2.9. For the HN=N fragment, the maximum deviation from the full DSRG-MRPT2 curve (shifted) degrades slightly with the cc-pVTZ

Table 2.1 Number of ASET(mf) fragment orbitals, relative computational timings for the DSRG-MRPT2, and maximum absolute deviation (MAD, in kcal/mol) of shifted potential energy curves. The DSRG-MRPT2 timings exclude the integral transformation are computed with respect to computations using HN_2 as a fragment.

Fragment	Orbitals	Timing (relative)	MAD (kcal/mol)
HN_2	33	1.0	1.43
HN_2CH_2	57	3.3	0.31
$HN_2C_2H_4$	81	10.4	0.06
$HN_2C_3H_6$	105	50.8	0.05
$HN_2C_4H_8$	129	94.3	0.03
$HN_2C_5H_{11}$	158	212.4	0.00

basis set (1.86 kcal mol^{-1}) and becomes as large as 3.54 kcal mol^{-1} for the cc-pVQZ basis. However, as the fragment size is increased to include two extra methyl groups, the energy differences are smaller and the basis set dependence of the error is significantly reduced. The maximum error using the larger fragment is reduced to 0.24 and 0.39 kcal mol^{-1} using the cc-pVTZ and cc-pVQZ basis sets, respectively. As expected, an increase in fragment size is accompanied by an increases in the number of fragment virtual orbitals produced by the ASET(mf) procedure. For example, going from the HN=N fragment to $HN{=}N{-}C_2H_4$, the number of virtual orbitals obtained with the cc-pVTZ basis is 72 and 186, respectively, which is consistent with the number of basis functions centered on these two fragments.

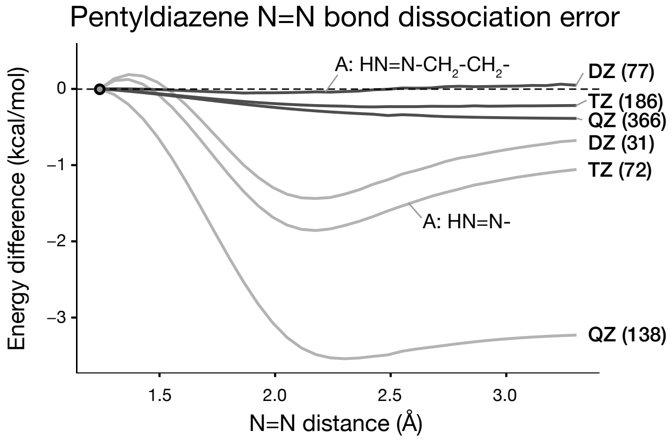


Figure 2.9 Pentyldiazene N=N dissociation curve. Comparison of the absolute energy error with respect to the full DSRG-MRPT2 for the ASET(mf)/DSRG-MRPT2 (vertical axis, kcal/mol for error) as a function of the N=N bond length (horizontal axis) and basis set. The energy difference curve is obtained after shifting the ASET(mf)/DSRG-MRPT2 curve so that the energy of the first point coincides with the full DSRG-MRPT2 energy. Computations use 35 geometries along the N=N bond vector. The labels XZ with X = D,T,Q, indicate the cc-pVXZ basis, while the corresponding number in parenthesis is the total fragment orbitals obtained by the ASET(mf) procedure for the first point in the curve.

2.5 A study of the products of the Bergmann reaction using ASET(mf)

In this section, we present an application of ASET(mf) to a realistic problem involving a multireference computation in the presence of a complex environment. We study the singlet–triplet splitting of two classical arynes, *p*-benzyne and 9,10-anthracyne, and examine the effect that an inert surface has on this property. Arynes have been the subject of several the-

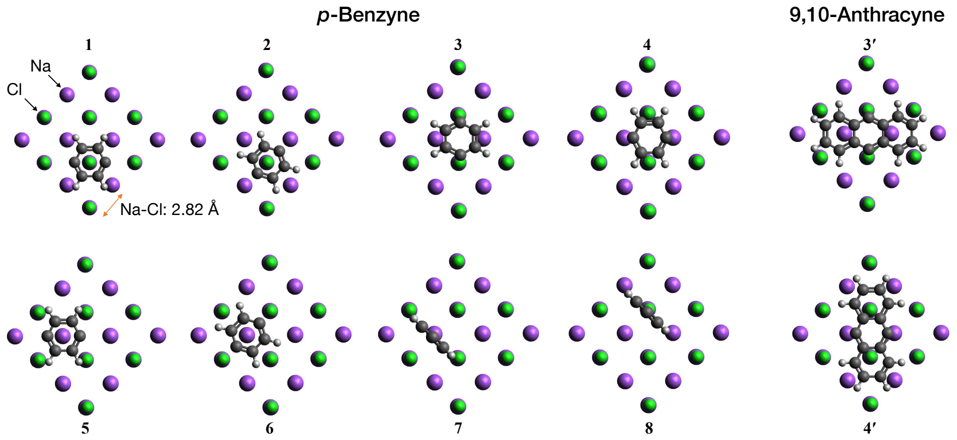


Figure 2.10 Geometrical arrangement of *p*-benzyne (**1**–**8**) and 9,10-anthracyne (**3′**,**4′**) on a model NaCl surface studied in this work. Sodium and chlorine atoms are represented with purple and green spheres, respectively.

oretical studies[84, 31, 182, 115, 127] due to their role in the Bergman cyclization reaction, the process that leads from a diyne to an aromatic diradical. Previous experimental studies found the gas-phase singlet–triplet gap of *p*-benzyne to be in the range 2–5 kcal/mol.[182, 115, 127] However, computed singlet–triplet gaps vary for different methods, and even the energy order of the singlet and triplet state can change depending on the theoretical treatment.[31, 127] Recent work has shown that the Bergman cyclization reaction that leads to 9,10-anthracyne (and its reverse) can be realized on a bilayer NaCl film on Cu(111) by manipulating the molecule with a CO-functionalized tip of an atomic force microscope.[150] Computational investigation of the resulting 9,10-anthracyne product suggested a triplet ground state.[150] This result is a contrast to gas-phase studies of 9,10-anthracyne using various multireference methods that predict a singlet ground state and a gap close to that of *p*-benzyne.[109] In this work, we employ ASET(mf) and DSRG-MRPT2 to study how a NaCl surface affects the singlet–triplet splitting of *p*-benzyne and 9,10-anthracyne.

To account for the interaction of *p*-benzyne and 9,10-anthracyne under the experimental conditions considered in Ref. 150, we begin by constructing a model system with two layers of a 4×4 NaCl lattice, with a fixed Na-Cl bond length equal to that of the crystal (2.82 Å).[100] In the case of *p*-benzyne, we study eight arrangements of the atoms on the NaCl model surface, six parallel and two perpendicular, as shown in Fig. 2.10. These include geometries in which the benzene ring is centered on a Cl atom (**1**, **2**), equally distant from Na and Cl atoms (**3**, **4**), centered on a Na atom (**5**, **6**), or perpendicular to the surface (**7**, **8**). For 9,10-anthracyne we consider only two parallel geometries (**3**$'$ and **4**$'$) analogous to the *p*-benzyne **3** and **4**. For each geometry, we compute the vertical (unrelaxed) singlet–triplet gap (ΔE_{ST}) as the difference between the triplet and singlet energy, neglecting corrections due to vibrational zero-point energy. The minimum active space required to describe these two electronic states contains two electrons in two orbitals, the bonding and antibonding σ radical orbitals. These are plotted for two geometries of *p*-benzyne and 9,10-anthracyne in Fig. 2.11. In our ASET(mf) treatment, the adsorbate is treated as an embedded fragment at the DSRG-MRPT2 level of theory, while the two 4×4 NaCl layers belong to the environment. Using the cc-pVDZ basis for *p*-benzyne, the ASET(mf) partition results into 104 fragment (A) orbitals (including

2 active orbitals) and 576 environment (B) orbitals; For 9,10-anthracyne, there are 236 fragment and 576 environment orbitals.

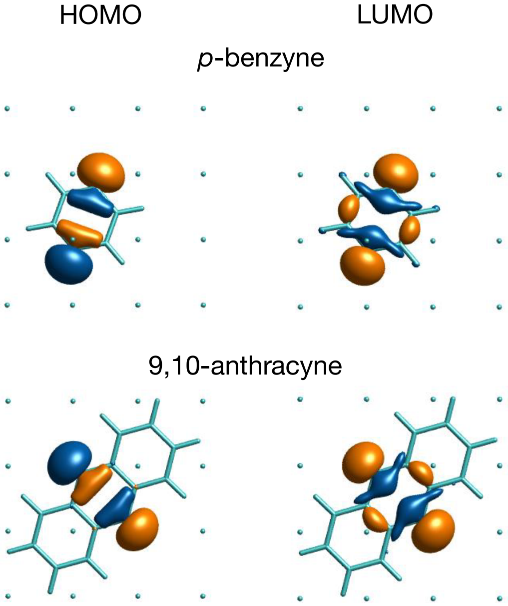


Figure 2.11 CASSCF(2e,2o) active orbitals for *p*-benzyne and 9,10-anthracyne diradicals for the lowest singlet state.

The relative energy and singlet–triplet splitting of *p*-benzyne in the vacuum and adsorbed on the NaCl surface at the optimum surface distance are shown in Table. 2.2. Among the molecular orientation considered in this study, those involving the interaction of the benzene ring with a Cl^- ion (**1** and **2**) or with the radical carbon atoms interacting with two Cl^- ions (**3**) are the most stable (in terms of the electron energy), with the latter being the minimum energy one. In comparison, interaction of the benzene ring with a Na^+ ion (**5** and **6**) or a radical carbon with two Na^+ ions (**4**)

Table 2.2 The relative ground state energy (in kcal/mol) and vertical singlet–triplet gaps (in kcal/mol) of different *p*-benzyne orientations on the NaCl surface computed with ASET(mf) and DSRG-MRPT2 using the cc-pVDZ basis.

Geometry	Energy (kcal/mol)	ΔE_{ST} (kcal/mol)
Vacuum	–	4.40
1	3.1	3.82
2	3.4	3.97
3	0.0	3.69
4	18.9	4.00
5	26.9	4.28
6	25.2	3.76
7	37.1	3.86
8	21.7	2.82

increases the energy by about 20 kcal/mol. The affinity of radical carbons for the Cl^- sites on the surface is also shown in perpendicular geometries **7** and **8**, where radical electron binding with Na^+ (**7**) has a singlet ground state 15.4 kcal/mol higher than **8**. For all adsorbed geometries, the singlet–triplet gap is smaller than the gas-phase value by 0.12–1.58 kcal/mol, with slight variation among the different orientations. This result implies that the NaCl double layer has the net effect of stabilizing the triplet state with respect to the singlet state. However, at the equilibrium geometry this stabilization is not sufficient to reverse the ordering of these two spin states.

In Fig. 2.12 we show ΔE_{ST} values computed for the lowest energy geometry (**3**) displaced along the direction normal to the surface (z axis). These results show that, in contrast to orientation, altering the distance of *p*-benzyne from the surface strongly impacts the singlet–triplet splitting. At

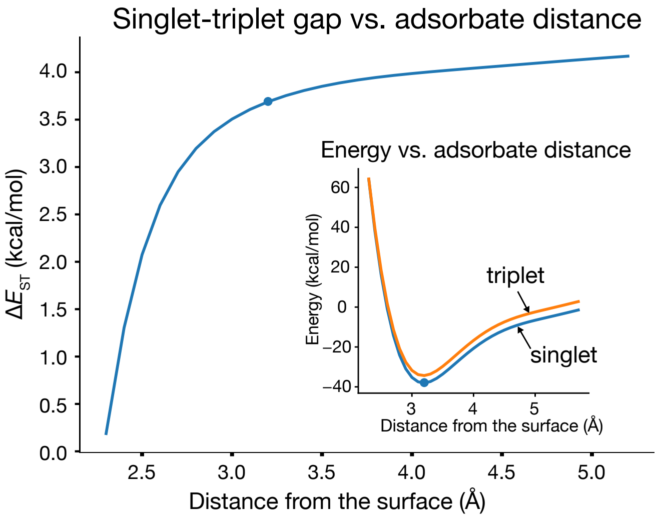


Figure 2.12 Distances to surface and vertical singlet–triplet gaps of *p*-benzyne (**3** geometry) on the NaCl surface computed with ASET(mf) and DSRG-MRPT2 using cc-pVDZ basis. The subplot shows the singlet and triplet ground state energies. 35 points are computed ranging from 2.3 Å to 5.7 Å. The lowest-energy adsorption geometry (**3**) is marked on both curves.

the equilibrium distance from the surface (3.2 Å), ΔE_{ST} is equal to ca. 3.7 kcal/mol. However, when the molecule is brought closer to the surface, ΔE_{ST} is reduced, and it can be made as small as 0.2 kcal/mol when the distance is 2.3 Å. When *p*-benzyne is displaced away from the surface, the singlet–triplet gap instead increases to over 4.2 kcal/mol (at $z = 5.5$ Å), approaching the ΔE_{ST} value in vacuum (4.4 kcal/mol).

Next, we proceed to improve the quality of the model NaCl bilayer with electrostatic embedding using classical charges. Specifically, we augment the ASET(mf) description by adding eight replicas of the NaCl 4×4 bilayer, as shown in Fig. 2.13. The classical charges on Na and Cl (+0.534 and

-0.534, respectively) were computed via a self-consistent procedure at the Hartree–Fock/def2-SVP level of theory.[177] Starting from a NaCl 4×4 bilayer, we obtain the average Mulliken charge on each atom type. We then repeat this computation embedding the bilayer in a classical charge field and obtain new average charges at the Hartree–Fock level. This procedure is iterated until achieving self-consistency of the classical charges. Although the point charge field has some impact on the embedding energy, it has a minor effect on the singlet–triplet gaps. For the most stable geometry of *p*-benzyne (**3**), we find that ΔE_{ST} is shifted by just -0.07 kcal/mol. A more significant shift (-0.15 kcal/mol) is observed for the **7** perpendicular geometry. Since the first nonzero term in the multipole expansion of benzyne is the quadrupole moment, long-range interactions with replicas of the NaCl unit cell beyond those included in this extended model decay as the inverse cube of the distance. Therefore, the sum of these contributions is expected to be nearly converged already in the extended model with eight replicas. We numerically estimated the magnitude of the interactions neglected in our model using classical point-charge electrostatic interactions from DFT and found them to be less than 0.01 kcal mol^{-1}.

Turning to 9,10-anthracyne, we consider two parallel orientations analogous to the most stable one for *p*-benzyne using, for convenience, the optimized distance from *p*-benzyne computations. The singlet–triplet gaps computed for these systems are reported in Table 2.3. In vacuum, the ΔE_{ST} of 9,10-anthracyne is lower by 0.95 kcal/mol with respect to that of *p*-

benzyne. When 9,10-anthracyne is embedded on a NaCl surface treated at the ASET(mf) level, the gaps for the **3**′ and **4**′ geometries are 0.94 and 0.66 kcal/mol lower than those in vacuum, respectively. In Table 2.3 we also report the relative ground state energies for the two orientations. Similar to *p*-benzyne, geometry **3**′ is more stable in terms of the total electron energy, in accordance with the preference for arrangements in which radical electrons interact with the Cl^- sites.

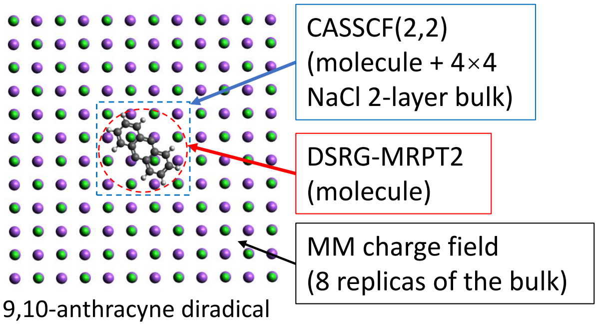


Figure 2.13 Definition of the fragment, ASET(mf) embedded region, and point-charges embedding potential used to model the interaction of 9,10-anthracyne diradicals with a NaCl bilayer surface. For each region we specify the level of theory. Sodium and chlorine atoms are represented with purple and green spheres, respectively.

In summary, our study suggests that the singlet–triplet gaps of *p*-benzyne and 9,10-anthracyne are not significantly perturbed when these molecules are adsorbed on a NaCl surface and that the singlet state is consistently the ground state for these systems. This result is in contrast with earlier DFT studies on 9,10-anthracyne,[150] which predicted a triplet ground state for the adsorbed molecule. This discrepancy is not surprising, espe-

Table 2.3 The relative ground state energy (in kcal/mol) and vertical singlet–triplet splitting (in kcal/mol) of vacuum and adsorbed 9,10-anthracyne computed with ASET(mf) and DSRG-MRPT2 using the cc-pVDZ basis.

Geometry	Energy (kcal/mol)	ΔE_{ST} (kcal/mol)
Vacuum	-	3.45
3′	0.0	2.51
4′	14.8	2.79

cially in light of the fact that singlet–triplet gaps of arynes vary significantly across different DFT functionals.[31, 127] Our study of the products of the Bergman reaction on a model NaCl surface using the ASET(mf) and the DSRG-MRPT2 shows the potential of such a combined approach in applications to large-scale problems that require a balanced description of strong correlation and environment effects.

2.6 Summary

This chapter introduces an active space embedding theory [ASET(mf)] to enable multireference computations on large-scale systems. This scheme combines mean-field frozen-core embedding with a simple fragment projection method that preserves the active space orbitals. This separation of active orbitals from the (localized) embedding orbitals is found to be crucial to accelerating the convergence of multireference computations as a function of the number of fragment orbitals.

Several examples are studies to examine the advantages of this approach. Our computation of the excitation energy of 1-octene highlights the importance of preserving variationally-optimized active orbitals from CASSCF in the embedding treatment. We also examine two bond-breaking problems involving weak (ethane + methane) and strong (pentyldiazene) coupling between the fragment and environment. In both cases, ASET(mf) yields curves that are highly parallel to those obtained from a multireference computation on the entire system. We also present a study of the single-triplet gap of *p*-benzyne and 9,10-anthracyne adsorbed on a NaCl surface using ASET(mf) augmented with a classical embedding potential. Contrary to a previous DFT study,[150] our results show that the ground state of *p*-benzyne and 9,10-anthracyne adsorbed on a NaCl surface is a diradical open-shell singlet.

In numerical experiments, we compute the N=N dissociation in pentyldiazene using ASET(mf) with DMET or PAOs virtual orbitals (employing the same number of core, active, and virtual orbitals); the two approaches are comparable in accuracy. Using a small basis set (cc-pVDZ), DMET virtuals lead to smaller relative errors than PAOs; however, in a larger basis (cc-pVTZ) more orbitals are included the description of the fragment and the accuracy of the two approaches becomes nearly identical (max deviation less of ca. 0.1 kcal mol^{-1}).

In summary, ASET(mf) is a simple and powerful embedding scheme that may significantly expand the domain of application of high-level multireference methods. Compared to other quantum embedding schemes, ASET(mf) intentionally avoids the self-consistent optimization of the embedding potential. This choice leads to a practical and economical one-step non-iterative embedding scheme that requires minimal user input. However, by ignoring the coupling between the fragment and environment, ASET(mf) is expected to be most accurate in the weak- to medium-coupling regime, especially in situations where electrostatic effects due to the environment play a dominant role. Additionally, like other local embedding schemes, ASET(mf) is best suited to describe systems in which electrons are localized. For example, in systems with delocalized electrons (e.g. graphene), orbital partitioning will introduce an artificial localization of the orbitals and shifts in their energies, which may introduce large errors in the computation of the correlation energy.

Using ASET(mf) as a baseline, further studies are conducted. In chapter 3, we will improve upon ASET via systematically-improvable treatments of the fragment/environment interaction based on approximate canonical transformations and second-order perturbation theory. In chapter 5, we will explore possible improvements to the partition procedure.

Chapter 3

Second-order active space embedding theory: ASET(2)

3.1 Extend ASET using canonical transformation

This chapter explores a systematic approach to improve the description of the fragment-environment interaction in embedding computations through a combination of canonical transformations and perturbation theory. This chapter is mainly based on a published work, with Dr. Chenyang Li's contribution to the code and implementation. See Ref. 72. (Publisher: American Chemical Society; Date: Mar 1, 2022; Copyright © 2022, American Chemical Society.)

As mentioned in the previous chapter, in embedding theory, the orbitals are localized either on the fragment (A) or environment (B), and consequently, the Hamiltonian for the fragment plus environment ($\hat{H}^{\mathrm{A+B}}$) is the sum of three terms

$$\hat{H}^{\mathrm{A+B}} = \hat{H}^{\mathrm{A}} + \hat{H}^{\mathrm{B}} + \hat{H}^{\mathrm{AB}} \tag{3.1}$$

where $\hat{H}^{\mathrm{A}}$ is the fragment Hamiltonian, $\hat{H}^{\mathrm{B}}$ is the environment Hamiltonian, and the fragment-environment interaction is encoded in the term $\hat{H}^{\mathrm{AB}}$. Note that in the absence of $\hat{H}^{\mathrm{AB}}$, diagonalization of $\hat{H}^{\mathrm{A}}$ alone would be sufficient to determine the energy levels of the fragment. This observation suggests eliminating $\hat{H}^{\mathrm{AB}}$ via a canonical transformation of the Hamiltonian. Following this transformation, the fragment is described by an effective Hamiltonian $\bar{H}^{\mathrm{A}}$, of size significantly smaller than the entire system. Canonical transformations find widespread use in physics and chemistry. They were first considered in the early works of Brandow,[19] Westhaus,[187] and Freed,[55] and later developed into practical quantum chemistry methods.[191, 8, 131, 118, 46, 123] Within the setting of the similarity renormalization group,[175] canonical transformations have been used to study quantum systems coupled to an environment.[86] Watson and Chan developed a perturbative canonical transformation scheme to reduce the size of the orbital space down to that of a minimum atomic-orbital basis while still retaining a qualitatively correct description of phenomena involving valence electrons.[174] Recently, research has been conducted to explore the application of canonical transformation and effective Hamiltonian techniques in quantum embedding.[119, 23]

This chapter proposes to formulate a systematically improvable embedding scheme based on the ASET scheme introduced in the previous chapter.[71] An essential feature of the ASET orbital partition procedure is the restriction of the orbital localization to non-active orbitals, which improves the accuracy of the resulting embedded Hamiltonian. In ASET(mf), the embedded Hamiltonian is the sum of the fragment Hamiltonian $\hat{H}^{\mathrm{A}}$ and frozen mean-field contributions from the environment, while the coupling $\hat{H}^{\mathrm{AB}}$ is neglected. To improve upon ASET(mf), we propose an embedding scheme that accounts for the fragment-environment interaction $\hat{H}^{\mathrm{AB}}$ via a unitary canonical transformation. This canonical transformation is realized in a computationally efficient and numerically robust way using the driven similarity renormalization group (DSRG).[43, 112] The DSRG has been implemented using both perturbative[107, 109, 111, 68] and iterative approximation schemes.[108, 195] Combining the DSRG with ASET(mf) leads to a hierarchy of post-MCSCF multireference embedding schemes that progressively better account for the fragment-environment interaction. This chapter focuses on the simplest combination: a one-shot second-order ASET [ASET(2)] scheme.

3.2 ASET(2) scheme

ASET perturbation theory. To improve upon ASET(mf), we use the framework of unitary effective Hamiltonian theory to improve the treatment of fragment-environment interactions. A canonical transformation of the

Hamiltonian via the exponential of an anti-Hermitian operator $\hat{A}$ leads to a transformed Hamiltonian ($\bar{H}$) that is the sum of operators that act on A, A + B, and B

$$\bar{H} = e^{-\hat{A}} \hat{H}^{\mathrm{A+B}} e^{\hat{A}} = \bar{H}^{\mathrm{A}} + \bar{H}^{\mathrm{AB}} + \bar{H}^{\mathrm{B}} \tag{3.2}$$

The purpose of this transformation should be twofold. First, is to diagonalize away terms that couple the fragment and environment

$$\bar{H}^{\mathrm{AB}} = 0 \tag{3.3}$$

When this condition is imposed, $\bar{H}$ does not contain a direct fragment-environment interaction term; however, the physical interaction between the fragment and the environment is still present and it is accounted by the terms $\bar{H}^{\mathrm{A}}$ and $\bar{H}^{\mathrm{B}}$ terms. The second purpose of the canonical transformation is to eliminate the coupling of the environment state with all fragment excited determinants (Φ'_{B}), that is

$$\langle \Phi'_{\mathrm{B}} | \bar{H}^{\mathrm{B}} | \Phi_{\mathrm{B}} \rangle = 0 \tag{3.4}$$

When these conditions are satisfied, then the energy of the system is the sum of the energy of the fragment Hamiltonian (obtained by diagonalizing $\bar{H}^{\mathrm{A}}$) plus the energy of the environment ($\langle \Phi_{\mathrm{B}} | \bar{H}^{\mathrm{B}} | \Phi_{\mathrm{B}} \rangle$). To achieve these two goals, the operator $\hat{A}$ must contain terms that couple A and B and terms that involve only second quantized operators acting on B

$$\hat{A} = \hat{A}^{\mathrm{AB}} + \hat{A}^{\mathrm{B}} \tag{3.5}$$

Like in the ASET(mf) scheme, we assume that the correlation within fragment (A) will be treated by a higher-order multireference method. Therefore, we always consider $\hat{A}^{\mathrm{A}}$ to be zero. In effective Hamiltonian theory, this condition is equivalent to excluding internal excitations.

It is important to point out that Eq. (3.3) corresponds to a *Fock-space condition*, that is, it is imposed directly on the normal-ordered operator components of $\bar{H}^{\mathrm{AB}}$. Such an approach is known to be prone to the intruder-state problem, and therefore, we employ the driven similarity renormalization group (DSRG) formalism,[107] which in practice solves a set of equation similar to Eq. (3.3):

$$\bar{H}^{\mathrm{AB}} = \hat{R}(s) \tag{3.6}$$

where $\hat{R}(s)$ is a term that regularizes the equations for diverging energy denominators and s is a parameter that controls the extent to which the elements of $\bar{H}^{\mathrm{AB}}$ are suppressed. An analogous set of equations also replaces Eq. (3.4). The use of the DSRG framework mainly affects the equations that determine the operator $\hat{A}$. Therefore, the scheme presented here is general in the sense that could be easily adopted to other regularization schemes or Hamiltonian partitionings. A thorough discussion of the DSRG is outside the scope of this work, and we refer the curious readers to a recent review.[107, 112]

In practice, Eqs. (3.3) and (3.4) can only be satisfied approximately and the costs of an embedding procedure should be minimized to ensure its applicability to large systems. Therefore, we propose to use low-order perturbation theory to perform the canonical transformation. The starting point is partitioning the full Hamiltonian $\hat{H}^{\mathrm{A+B}}$ into a sum of a zeroth-order [$\hat{H}^{(0)}$] and a first-order [$\hat{H}^{(1)}$] term

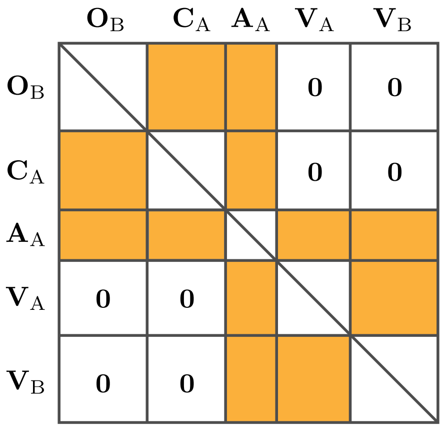
$$\hat{H}^{\mathrm{A+B}} = \hat{H}^{(0)} + \xi \hat{H}^{(1)} \tag{3.7}$$

where ξ is the perturbation ordering parameter. Both $\hat{H}^{(0)}$ and $\hat{H}^{(1)}$ can be further separated into contributions involving A, AB, and B, for example, $\hat{H}^{(0)} = \hat{H}^{\mathrm{A},(0)} + \hat{H}^{\mathrm{AB},(0)} + \hat{H}^{\mathrm{B},(0)}$. The zeroth-order ASET wave function is defined to be the CASSCF reference. Following a standard approach, perturbative corrections to the transformed Hamiltonian $\bar{H}$ may be obtained by expanding the operator $\hat{A}$ and energy E in a power series in ξ and truncating the ASET equations [Eq. (3.2) and (3.6)] to a finite order.

Second-order ASET: ASET(2). To develop a perturbative ASET scheme with $\bar{H}$ truncated to second order, we begin by partitioning the full Hamiltonian in a way consistent with the choice of reference. We first normal order $\hat{H}^{\mathrm{A+B}}$ with respect to the CASSCF wave function (Ψ_0) using the approach of Mukherjee and Kutzelnigg.[99] The resulting Hamiltonian may be written as

$$\hat{H}^{\mathrm{A+B}} = E_0^{\mathrm{A+B}} + \sum_{pq}^{\mathrm{A+B}} f_p^q \{\hat{a}_q^p\}_{\Psi_0} + \frac{1}{4} \sum_{pqrs}^{\mathrm{A+B}} \langle pq\|rs\rangle \, \{\hat{a}_{rs}^{pq}\}_{\Psi_0} \tag{3.8}$$

where the generalized Fock matrix elements f_p^q are given by Eq. (2.6) and $E_0^{\mathrm{A+B}}$ is the total CASSCF energy. Following the semicanonicalization step in the ASET procedure, the generalized Fock matrix has the block structure shown in Fig. 3.1, where all blocks on the diagonal are diagonal, blocks that couple virtual orbitals to the doubly occupied orbitals are zero due to the CASSCF stationarity condition, and the remaining blocks are nonzero.



Fock matrix in MO basis
after semi-canonicalization

Figure 3.1 Structure of the generalized Fock matrix in a semi-canonical basis after the ASET orbital localization procedure. Blocks corresponding to the frozen-core environment orbitals are not shown in this figure.

We use a diagonal Fock partitioning in which the ASET zeroth-order operator is given by

$$\hat{H}^{(0)} = E_0^{\mathrm{A+B}} + \sum_u^{\mathrm{A}} \varepsilon_u \{\hat{a}_u^u\}_{\Psi_0} + \sum_p^{\mathrm{B}} \varepsilon_p \{\hat{a}_p^p\}_{\Psi_0} \tag{3.9}$$

where $\varepsilon_p = f_p^p$. Note that $\hat{H}^{(0)}$ contains no terms that couple the fragment and the environment. The first-order Hamiltonian may be written as the sum of three terms $\hat{H}^{\mathrm{A},(1)}$, $\hat{H}^{\mathrm{B},(1)}$, and $\hat{H}^{\mathrm{AB},(1)}$ involving operators that act on A, B, and A + B, respectively.

We consider orbital relaxation and correlation effects up to two-body terms, and express $\hat{A}$ in terms of single and double substitution operators $\hat{T}_1$ and $\hat{T}_2$ as:

$$\hat{A} = \hat{A}_1 + \hat{A}_2 = \hat{T}_1 - \hat{T}_1^\dagger + \hat{T}_2 - \hat{T}_2^\dagger \tag{3.10}$$

The excitation operators contain only those components that couple A and B ($\hat{T}^{\mathrm{AB}}$), and excitations exclusively on B ($\hat{T}^{\mathrm{B}}$). For example, in the case of single excitations, we define these two blocks as

$$\hat{T}_1^{\mathrm{AB}} = \sum_u^{\mathrm{A}} \sum_e^{\mathbf{V}_{\mathrm{B}}} t_e^u \{\hat{a}_u^e\}_{\Psi_0} + \sum_m^{\mathbf{O}_{\mathrm{B}}} \sum_u^{\mathrm{A}} t_u^m \{\hat{a}_m^u\}_{\Psi_0} \tag{3.11}$$

and

$$\hat{T}_1^{\mathrm{B}} = \sum_m^{\mathbf{O}_{\mathrm{B}}} \sum_e^{\mathbf{V}_{\mathrm{B}}} t_e^m \{\hat{a}_m^e\}_{\Psi_0} \tag{3.12}$$

The two terms that enter $\hat{T}_1^{\mathrm{AB}}$ capture orbital relaxation effects involving A and B, and physically correspond to charge-transfer excitations that move electrons between the fragment and environment. Instead, $\hat{T}_1^{\mathrm{B}}$ describes orbital rotations within the environment. The double excitation operator $\hat{T}_2$ can be similarly partitioned into contributions from A and B, and B alone. Physically, $\hat{T}_2^{\mathrm{AB}}$ accounts for charge-transfer excitations coupled to a single excitation on A or B and the product of one single excitation on A and one on B. The remaining term, $\hat{T}_2^{\mathrm{B}}$, instead captures correlation effects in the environment.

It is easy to see that by taking the sum of the zeroth- and first-order transformed Hamiltonians one obtains an embedding approach equivalent to mean-field ASET. The first nontrivial correction to ASET(mf) is found at second-order in perturbation theory. When the zeroth- through second-order terms are combined, we obtain the following Hamiltonian

$$\begin{aligned}\bar{H}^{[2]} &= \hat{H} + [\hat{H}, \hat{A}^{(1)}] + \frac{1}{2}[[\hat{H}^{(0)}, \hat{A}^{(1)}], \hat{A}^{(1)}] \\ &= \bar{H}_0^{[2]} + \sum_{pq}^{\mathrm{A+B}} \bar{H}_p^{q,[2]} \{\hat{a}_q^p\}_{\Psi_0} + \frac{1}{4} \sum_{pqrs}^{\mathrm{A+B}} \bar{H}_{pq}^{rs,[2]} \{\hat{a}_{rs}^{pq}\}_{\Psi_0} \end{aligned} \tag{3.13}$$

This expression ignores three-body contributions that arise already in single-commutator terms. Note also, that although the second-order operator $\hat{A}^{(2)}$ also enters Eq. (3.13), this operator does not contribute to the fragment portion of $\bar{H}^{[2]}$ and, therefore, it is not shown.

In the second line of Eq. (3.13), we express $\bar{H}^{[2]}$ using second-quantized operators normal-ordered with respect to the CASSCF reference. The second-order energy ($\bar{H}_0^{[2]}$) contains several fully connected terms that enter also in second-order DSRG multireference perturbation theory (DSRG-MRPT2).[107] However, in our standard (and most efficient) implementation of ASET(2), we neglect the contributions from the three-body density cumulant (a rank six tensor with indices running over all fragment orbitals) to all terms in $\bar{H}^{[2]}$. This approximation does not introduce artifactual intruder states,[107] and as shown in the Results section, its impact on the accuracy of ASET(2) is negligible. The one- and two-body components of $\bar{H}^{[2]}$ labeled by fragment orbital indices may be written as the sum of the bare integrals plus corrections as

$$\bar{H}_u^{v,[2]} = f_u^v + \frac{1}{2}\left(\bar{C}_u^{v,(2)} + \bar{C}_v^{u,(2)}\right) \tag{3.14}$$

$$\bar{H}_{uv}^{xy,[2]} = \langle uv \| xy \rangle + \frac{1}{2}\left(\bar{C}_{uv}^{xy,(2)} + \bar{C}_{xy}^{uv,(2)}\right) \tag{3.15}$$

where detailed expressions for $\bar{C}_u^{v,(2)}$ and $\bar{C}_{uv}^{xy,(2)}$ are reported in Sec. 3.3. These quantities depend on the first-order amplitudes, given by

$$t_a^{i,(1)} = \left[f_a^i + \sum_{ux}^{\mathrm{A}} \Delta_u^x t_{ax}^{iu,(1)} \gamma_u^x\right] \frac{1 - e^{-s(\Delta_a^i)^2}}{\Delta_a^i} \tag{3.16}$$

$$t_{ab}^{ij,(1)} = \langle ab \| ij \rangle \frac{1 - e^{-s(\Delta_{ab}^{ij})^2}}{\Delta_{ab}^{ij}} \tag{3.17}$$

where the denominators are defined as $\Delta_a^i = \varepsilon_i - \varepsilon_a$ and $\Delta_{ab}^{ij} = \varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b$. In the implementation section (Sec.3.3), we provide the detailed equation for those terms.

Any high-level multireference computation on the fragment requires only the component of the ASET(2) Hamiltonian spanning the fragment orbitals. Moreover, to include the effect of the (frozen) environment electrons, it is also necessary to normal order $\bar{H}^{(2)}$ respect to the state $|-_{\mathrm{A}}\rangle \otimes |\Phi_{\mathrm{B}}\rangle$. The resulting dressed Hamiltonian ($\bar{G}$) is expressed as

$$\bar{G} = E^{\mathrm{A+B},[2]} + \sum_{uv}^{\mathrm{A}} \bar{h}_u^{v,[2]} \hat{a}_v^u + \frac{1}{4} \sum_{uvxy}^{\mathrm{A}} \bar{h}_{uv}^{xy,[2]} \hat{a}_{xy}^{uv} \tag{3.18}$$

where $E^{\mathrm{A+B},[2]}$ contains the environment correlation energy

$$E^{\mathrm{A+B},[2]} = \bar{H}_0^{[2]} - \sum_{xy}^{\mathrm{A}} \bar{H}_x^{y,[2]} \gamma_y^x - \sum_{uvxy}^{\mathrm{A}} \bar{H}_{uv}^{xy,[2]} \left(\frac{1}{4}\gamma_{xy}^{uv} - \gamma_x^u \gamma_y^v\right) \tag{3.19}$$

and the dressed integrals $\bar{h}_u^{v,[2]}$ and $\bar{h}_{uv}^{xy,[2]}$ are defined as

$$\bar{h}_u^{v,[2]} = \bar{H}_u^{v,[2]} - \sum_{xy}^{\mathrm{A}} \bar{H}_{ux}^{vy,[2]} \gamma_y^x \tag{3.20}$$

$$\bar{h}_{uv}^{xy,[2]} = \bar{H}_{uv}^{xy,[2]} \tag{3.21}$$

The energy expression for ASET(2)-[MR] can be written as:

$$E_{\mathrm{ASET(2)}} = E_0^{\mathrm{A+B}} + \delta E_{\mathrm{c,ASET(2)}}^{\mathrm{AB}} + \delta E_{\mathrm{c,ASET(2)}}^{\mathrm{B}} + \delta E_{\mathrm{MR}}^{\mathrm{A}}(\bar{G}) \tag{3.22}$$

Where $\delta E_{\mathrm{c,ASET(2)}}^{\mathrm{X}}$ is the ASET(2) correlation energy contribution from the fragment-environment or environment (X = AB or B), while $\delta E^{\mathrm{A}}_{\mathrm{MR}}(\bar{G})$ is the correlation energy correction from the high-level MR computation using the dressed Hamiltonian. Explicit expressions for $\delta E^{\mathrm{B}}_{\mathrm{c,ASET(2)}}$ and $\delta E^{\mathrm{AB}}_{\mathrm{c,ASET(2)}}$ are presented in Sec. 3.3. In comparison to ASET(mf), ASET(2) contains additional contributions from the fragment-environment and environment correlation $[\delta E^{\mathrm{AB}}_{\mathrm{c,ASET(2)}} + \delta E^{\mathrm{B}}_{\mathrm{c,ASET(2)}}]$ and it uses the dressed Hamiltonian $\bar{G}$ in the high-level MR computation instead of the frozen Hamiltonian $\hat{G}$.

When all fragment orbitals are taken to be active and the fragment Hamiltonian is diagonalized at the full configuration interaction level, ASET(2) becomes equivalent to a second-order multireference perturbation theory, specifically, the partially-relaxed variant of DSRG-MRPT2.[107] In this limit it is possible to compare ASET(2) with other multireference perturbation theories. A previous study[109, 110] found that in comparison to CASPT2[3, 4, 24] and partially-contracted NEVPT2[5, 6] DSRG-MRPT2 is slightly less accurate. For example, the nonparallelism error along the dissociation energy curves of F_2, H_2O_2, C_2H_6, and N_2 computed with the DSRG-MRPT2 has a mean value of 5.3 kcal mol^{-1} vs. 2.5 and 3.2 kcal mol^{-1} for CASPT2 and NEVPT2, respectively.

Generalization to excited states and reference relaxation. The ASET(2) procedure may be readily generalized to treat excited states. In the simplest approach, we employ the same ASET orbital partitioning for all electronic states, and use ground-state or state-averaged CASSCF orbitals to define the fragment and the environment. Then, for each electronic state, the corresponding fragment RDMs are generated from CASCI solutions within the active space, and the Fock matrix is semi-canonicalized differently for each state. The ASET(2) correction in turn uses the density and Fock matrix for each specific CI root, and the $\delta E^{\mathrm{AB}}_{\mathrm{c,ASET(2)}}$ and $\delta E^{\mathrm{B}}_{\mathrm{c,ASET(2)}}$ corrections will acquire state-specific character. Similarly, the dressed Hamiltonian $\bar{G}$ will be different for different states.

Alternatively, a more accurate representation of excited states may be obtained by defining a different orbital partitioning for each state using state-specific CASSCF solutions. This alternative procedure leads to more accurate results but increases the cost of the ASET(2) procedure. Another possible approach is to enforce self-consistency in ASET(2) embedding in high-level MR theories that account for reference relaxation effects. This approach consists in taking the updated CASCI coefficients, recomputing the RDMs, and re-evaluating the ASET(2) energy and dressed Hamiltonian (without change the fragment orbitals). This procedure could then be iterated until the ASET(2) energy does not change from one cycle to the next. In this paper, we test both approaches to orbital partitioning but we do not investigate the self-consistent version of ASET(2).

Analysis of the ASET(2) Hamiltonian. It is instructive to analyze some of the contributions to the ASET(2) Hamiltonian and understand the role of the exponential terms in Eqs. (3.16) and (3.17). Consider for example, the effective two-electron interaction, $\bar{H}^{xy,(2)}_{uv}$. The first contribution to this quantity is given by the contraction of a modified two-body integral

$(\tilde{v}^{ey}_{uv} = \langle uv\|ey\rangle\,[1+e^{-s(\Delta^{uv}_{ey})^2}])$ with a single excitation from the fragment to the environment virtual orbitals (t^x_e)

$$\bar{H}^{xy,(2)}_{uv} \leftarrow \sum_{e}^{\mathbf{V}_\mathrm{B}} \tilde{v}^{ey}_{uv} t^x_e, \quad u,v,x,y \in \mathrm{A} \tag{3.23}$$

This correction can be interpreted physically as the combination of a charge-transfer excitation ($\phi_x \rightarrow \phi_e$, from t^x_e) combined with a two-electron scattering interaction that brings the electron back to the fragment ($\phi_e\phi_y \rightarrow \phi_u\phi_v$, from $\tilde{v}^{ey}_{uv}$). Because the tensor t^x_e involves indices on both the fragment and the environment, this may be also understood as an orbital relaxation (re-hybridization) effect. Figure 3.2 represents diagrammatically the effective interaction in terms of the bare interaction plus corrections, where the first correction shown corresponds to the contribution from Eq. (3.23). Note that if the fragment (ϕ_x) and environment (ϕ_e) orbitals become degener-

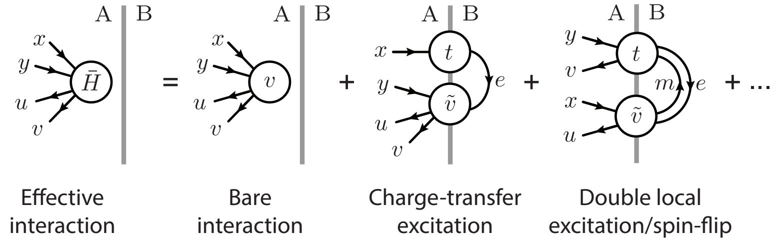


Figure 3.2 Diagrammatic representation of the fragment effective two-electron interaction $\bar{H}$. The first term is the bare Coulomb interaction (v), while the second term corresponds to the contributions from Eq. (3.23). The vertical gray line separates the fragment (A) from the environment (B). Tensors are indicated with circles, while tensor indices are represented with incoming (upper indices) and outgoing (lower indices) lines.

ate, the exponential term in Eq. (3.16) will suppress this excitation. This is an essential feature of ASET(2), because in general there is no guarantee that the fragment and environment orbitals will be non-degenerate.

Note that the modified interactions for the fragment orbitals are, in a sense, independent of orbital occupation. For example, the contribution to $\bar{H}^{xy,[2]}_{uv}$ considered above exists even if orbital ϕ_x is unoccupied in the CASSCF reference. Even though these excitations do not contribute to the reference energy (as it can be seen by analyzing the contributions to the expectation value of $\bar{H}^{[2]}$ with respect to the CASSCF reference), they still play an important role since they modify interactions between electrons if orbital ϕ_x is occupied.

3.3 ASET(2) implementation

Hamiltonian terms. we provide detailed equations to evaluate the ASET(2) Hamiltonian. Introducing the modified first-order integrals

$$\tilde{f}^a_i = f^a_i + \left[f^a_i + \sum_{ux}^{\mathrm{A}} \Delta^x_u t^{iu}_{ax} \gamma^x_u \right] e^{-s(\Delta^i_a)^2} \tag{3.24}$$

$$\tilde{v}^{ab}_{ij} = \langle ij \| ab \rangle \left[1 + e^{-s(\Delta^{ij}_{ab})^2} \right] \tag{3.25}$$

the second-order energies under the embedding partition are given by

$$\begin{aligned}
\delta E^{\mathrm{B}}_{\mathrm{c,ASET(2)}} &= \tilde{f}^{e}_{m} t^{m}_{e} + \frac{1}{4} \tilde{v}^{mn}_{ef} t^{ef}_{mn} \qquad (3.26) \\
\delta E^{\mathrm{AB}}_{\mathrm{c,ASET(2)}} &= \tilde{f}^{e}_{u} t^{v}_{e} \gamma^{u}_{v} + \tilde{f}^{v}_{m} t^{m}_{u} \eta^{u}_{v} \\
&+ \frac{1}{2} \lambda^{xy}_{uv} [\tilde{f}^{e}_{x} t^{uv}_{ey} - \tilde{f}^{v}_{m} t^{um}_{xy} + \tilde{v}^{ev}_{xy} t^{u}_{e} - \tilde{v}^{uv}_{my} t^{m}_{x}] \\
&+ \frac{1}{2} \tilde{v}^{ef}_{mu} t^{mv}_{ef} \gamma^{u}_{v} + \frac{1}{2} \tilde{v}^{ev}_{mn} t^{mn}_{eu} \eta^{u}_{v} + \frac{1}{4} \tilde{v}^{ef}_{xu} t^{yv}_{ef} \gamma^{x}_{y} \gamma^{u}_{v} \\
&+ \frac{1}{4} \tilde{v}^{vy}_{mn} t^{mn}_{ux} \eta^{u}_{v} \eta^{x}_{y} + \tilde{v}^{ve}_{mx} t^{my}_{ue} \gamma^{x}_{y} \eta^{u}_{v} \\
&+ \frac{1}{2} \tilde{v}^{ve}_{yz} t^{wx}_{ue} \gamma^{y}_{w} \gamma^{z}_{x} \eta^{u}_{v} + \frac{1}{2} \tilde{v}^{vy}_{mw} t^{mz}_{ux} \gamma^{w}_{z} \eta^{u}_{v} \eta^{x}_{y} \\
&+ \frac{1}{8} \tilde{v}^{cd}_{xy} t^{uv}_{ab} \eta^{a}_{c} \eta^{b}_{d} \lambda^{xy}_{uv} + \frac{1}{8} \tilde{v}^{uv}_{kl} t^{ij}_{xy} \gamma^{k}_{i} \gamma^{l}_{j} \lambda^{xy}_{uv} \\
&+ \tilde{v}^{bu}_{jx} t^{iv}_{ay} \gamma^{j}_{i} \eta^{a}_{b} \lambda^{xy}_{uv} \qquad (3.27)
\end{aligned}$$

Here, γ^{u}_{v}, $\eta^{u}_{v} = \delta^{u}_{v} - \gamma^{u}_{v}$, and λ^{uv}_{xy} are the elements of the one-particle density matrix, one-hole density matrix, and the two-body density cumulant introduced in Ref. 99, respectively. In Eqs. (3.26) and (3.27), we have adopted Einstein's convention for summation over repeated indices and the orbital labelings are shown in Fig. 2.4. Note that in the expression for $\delta E^{\mathrm{AB}}_{\mathrm{c,ASET(2)}}$ [Eq. (3.27)] each term involves amplitude tensors with at least one fragment index.

The intermediates $\bar{C}^{v,(2)}_{u}$ and $\bar{C}^{xy,(2)}_{uv}$ that enter in the definitions of $\bar{H}^{v,[2]}_{u}$

and $\bar{H}_{uv}^{xy,[2]}$ [see Eq. (3.14)] are

$$\begin{aligned}
\bar{C}_{u}^{v,[2]} = {} & \tilde{f}_{u}^{e} t_{e}^{v} - \tilde{f}_{m}^{v} t_{u}^{m} + \tilde{v}_{uj}^{va} t_{a}^{i} \gamma_{i}^{j} - \tilde{v}_{um}^{vy} t_{x}^{m} \gamma_{y}^{x} \\
& + \tilde{f}_{j}^{a} t_{au}^{iv} \gamma_{i}^{j} - \tilde{f}_{m}^{y} t_{xu}^{mv} \gamma_{y}^{x} \\
& + \frac{1}{2} \tilde{v}_{ui}^{ab} t_{ab}^{vj} \gamma_{j}^{i} - \tilde{v}_{ui}^{yb} t_{xb}^{vj} \gamma_{j}^{i} \gamma_{y}^{x} \\
& + \tilde{v}_{ux}^{ew} t_{ey}^{vz} \lambda_{wz}^{xy} + \frac{1}{4} \tilde{v}_{um}^{wz} t_{xy}^{vm} \lambda_{wz}^{xy} + \frac{1}{2} \tilde{v}_{um}^{zy} t_{wx}^{vm} \gamma_{z}^{w} \gamma_{y}^{x} \\
& - \frac{1}{2} \tilde{v}_{ij}^{va} t_{ub}^{ij} \eta_{a}^{b} + \tilde{v}_{xj}^{va} t_{ub}^{yj} \eta_{a}^{b} \eta_{y}^{x} \\
& - \tilde{v}_{mx}^{vw} t_{uy}^{mz} \lambda_{wz}^{xy} - \frac{1}{4} \tilde{v}_{xy}^{ve} t_{ue}^{wz} \lambda_{wz}^{xy} - \frac{1}{2} \tilde{v}_{wx}^{ve} t_{ue}^{zy} \eta_{z}^{w} \eta_{y}^{x} \\
& + \frac{1}{2} \tilde{v}_{xy}^{ez} t_{eu}^{wv} \lambda_{wz}^{xy} - \frac{1}{2} \tilde{v}_{my}^{wz} t_{xu}^{mv} \lambda_{wz}^{xy} \\
& + \frac{1}{2} \tilde{v}_{xu}^{ev} t_{ey}^{wz} \lambda_{wz}^{xy} - \frac{1}{2} \tilde{v}_{mu}^{wv} t_{xy}^{mz} \lambda_{wz}^{xy} \qquad (3.28) \\
\bar{C}_{uv}^{xy,[2]} = {} & \mathscr{P}_{-}(x,y) \tilde{v}_{uv}^{ey} t_{e}^{x} - \mathscr{P}_{-}(u,v) \tilde{v}_{mv}^{xy} t_{u}^{m} \\
& + \mathscr{P}_{-}(u,v) \tilde{f}_{u}^{e} t_{ev}^{xy} - \mathscr{P}_{-}(x,y) \tilde{f}_{m}^{x} t_{uv}^{my} \\
& + \frac{1}{2} \tilde{v}_{uv}^{ab} t_{ab}^{xy} - \tilde{v}_{uv}^{ze} t_{we}^{xy} \gamma_{z}^{w} \\
& + \frac{1}{2} \tilde{v}_{ij}^{xy} t_{uv}^{ij} - \tilde{v}_{wm}^{xy} t_{uv}^{zm} \eta_{z}^{w} \\
& + \mathscr{P}_{-}(u,v) \mathscr{P}_{-}(x,y) (\tilde{v}_{iu}^{ax} t_{av}^{jy} \gamma_{j}^{i} - \tilde{v}_{mu}^{zx} t_{wv}^{my} \gamma_{z}^{w}) \qquad (3.29)
\end{aligned}$$

The operator $\mathscr{P}_{-}(p,q)$ acting on a function $f(p,q)$ that depends on indices p and q yields $\mathscr{P}_{-}(p,q) f(p,q) = f(p,q) - f(q,p)$. We also point out that in Eqs. (3.27)–(3.29) the excitations within the fragment are ignored: $t_{u}^{v} = 0$ and $t_{uv}^{xy} = 0, \forall\, u,v,x,y \in \mathrm{A}$. Moreover, since the operator $\hat{H}^{(0)}$ is diagonal, the contribution from the commutators $[\hat{H}^{(0)}, \hat{A}^{(k)}], k = 1,2$ to $\bar{H}_{u}^{v,[2]}$ and $\bar{H}_{uv}^{xy,[2]}$ is null.

Implementation details. Our implementation of ASET(2) takes advantage of our DSRG-MRPT2 code to compute the tensor elements of $\bar{H}^{[2]}$. Any higher-order multireference method ([MR]) can then be used to diagonalize the $\bar{H}$ built by ASET(2) downfolding. We denote this combination of approach as ASET(2)-[MR].

The ASET(2) procedure consists of the following steps:

1. Select the active orbitals and run a CASSCF computation on A + B. This computation returns the reference energy $E_{\text{CASSCF}}^{\text{A+B}}$, and the CASSCF wave function. Optionally, one may decide to use a mean-field wave function instead of CASSCF (e.g., Hartree–Fock or Kohn–Sham DFT).

2. Form the embedding orbitals using the fragment projector $\hat{P}^{\text{A}}$. This partitioning leaves the original active orbitals ($\mathbf{A}_{\text{A}}$) fixed.

3. Compute the singles and doubles amplitudes by solving Eqs. (3.16) and (3.17).

4. From the amplitudes compute the scalar second-order energy ($\bar{H}_0^{[2]}$) and the matrix elements of $\bar{H}^{[2]}$. Transform the matrix elements of $\bar{H}^{[2]}$ into the one- and two-body dressed integrals. The resulting dressed integrals ($\bar{h}$) span only the spaces $\{\mathbf{A}_{\text{A}}, \mathbf{C}_{\text{A}}, \mathbf{V}_{\text{A}}\}$.

5. Run a high-level [MR] computation using the dressed integrals ($\bar{h}$)

with the space $\mathbf{A}_\mathrm{A}$ treated as active, while $\mathbf{C}_\mathrm{A}$ and $\mathbf{V}_\mathrm{A}$ are closed and virtual sets, respectively.

The first two steps of the ASET(2) procedure are common to ASET(mf). Note that step **3** is equivalent to running a DSRG-MRPT2 computation with core, active, and virtual orbital sets chosen to be $\{\mathbf{O}_\mathrm{B}\}$, $\{\mathbf{A}_\mathrm{A}, \mathbf{C}_\mathrm{A}, \mathbf{V}_\mathrm{A}\}$, and $\{\mathbf{V}_\mathrm{B}\}$, respectively. Therefore, our implementation takes advantage of existing optimized routines to evaluate the most expensive steps of the ASET(2) procedure.

In step **4**, the RDMs of the fragment are required to compute the ASET(2) Hamiltonian. We employ the RDMs obtained from a CASCI computation in the fragment active space ($\mathbf{A}_\mathrm{A}$), which for the ground state is equivalent to the CASSCF RDMs. The one-body RDMs in the fragment contains only two contributions from the core and active blocks

$$\begin{aligned} &\gamma_u^v = \delta_{uv} && u,v \in \mathbf{C}_\mathrm{A} \\ &\gamma_u^v = \gamma_u^v(\text{CASCI}) && u,v \in \mathbf{A}_\mathrm{A} \end{aligned} \tag{3.30}$$

The two-body RDM (and the corresponding cumulant) is similarly reconstructed from the CASCI 2-RDM as

$$\begin{aligned} &\gamma_{uv}^{xy} = \delta_{ux}\delta_{vy} - \delta_{uy}\delta_{vx} && u,v,x,y \in \mathbf{C}_\mathrm{A} \\ &\gamma_{xu}^{yv} = -\gamma_{xu}^{vy} = \ldots = \gamma_x^y \delta_{uv} && u,v \in \mathbf{C}_\mathrm{A}, x,y \in \mathbf{A}_\mathrm{A} \\ &\gamma_{xy}^{zw} = \gamma_{xy}^{zw}(\text{CASCI}) && w,x,y,z \in \mathbf{A}_\mathrm{A} \end{aligned} \tag{3.31}$$

As mentioned above, we neglect contributions that arise from the three-body density cumulant as these are expensive to evaluate and lead to minor improvements to the ASET(2) energy. It is important to note that these contributions only amount to a shift in the correlation energy and do not change the renormalized interaction produced by the canonical transformation. More specifically, the three-body cumulant (λ^{xyz}_{uvw}, see Ref. 99) contributes to the $\delta E^{\mathrm{AB}}_{\mathrm{c,ASET(2)}}$ term and may be expressed in terms of the doubles amplitudes and modified two-electron integrals ($\tilde{v}$, defined in Sec. 3.3) as:

$$\delta E^{\mathrm{AB}}_{\mathrm{c,ASET(2)}} \leftarrow \frac{1}{4} \sum_{uvwxyz}^{\mathrm{A}} \left(\sum_{m}^{\mathbf{O}_{\mathrm{B}}} \tilde{v}^{uv}_{mz} t^{mw}_{xy} + \sum_{e}^{\mathbf{V}_{\mathrm{B}}} \tilde{v}^{we}_{xy} t^{uv}_{ez} \right) \lambda^{xyz}_{uvw} \tag{3.32}$$

The effect of this approximation is easy to analyze for the excitation energy of a molecule at a fixed geometry. In this case, the neglected term introduces a shift in the excitation energy that is dominated by a term linear in the difference in the three-body cumulant of the two electronic states. We numerically examine the effect of this approximation in our discussion of the excitation energy of 1-octene.

Computational cost and truncation of environment. In addition to the cost of the high-level MR computation, in a conventional four-center integral implementation of ASET(2), the most expensive step of the embedding procedure is step **4**, where an integral transformation is required. The cost for the transformation scales as N^5 while the evaluation of the ASET(2) amplitudes and the MP2-like energy contribution $\delta E^{\mathrm{B}}_{\mathrm{c,ASET(2)}}$ is dominated by a step that scales as $N_{\mathbf{O}_\mathrm{B}}^2 N_{\mathbf{V}_\mathrm{B}}^2$ when the number of fragment orbitals is small compared to the environment. The highest-scaling term with respect to the number of fragment orbitals has a computational cost proportional to $N_\mathrm{A}^4 N_{\mathbf{V}_\mathrm{B}}^2$, with N_A being the number of fragment orbitals. This analysis accounts for the savings introduced by neglecting the three-body density cumulant, which scale as $N_\mathrm{A}^6 N_{\mathbf{V}_\mathrm{B}}$. Another approximation available in our implementation is neglecting the expensive contributions due to excitations in the environment, corresponding to the term $\delta E_{\mathrm{c,ASET(2)}}^{\mathrm{B}}$ in Eq. (3.22). Since the term $\delta E_{\mathrm{c,ASET(2)}}^{\mathrm{B}}$ shows only a weak dependence on the electronic state of the fragment (this dependence entering via the one-body density matrix in the denominators), this approximation is especially accurate when computing the energy difference between electronic states at a fixed geometry. However, we find that when describing bond-breaking processes and modeling weak interactions, $\delta E_{\mathrm{c,ASET(2)}}^{\mathrm{B}}$ may be significant and cannot be neglected. Similar ideas have been explored in a recent study by Kowalski on Subsystem Embedding Subalgebra Coupled Cluster (SES-CC) theory.[10] They describe the amplitude as different groups or subalgebra and use this classification to develop an efficient subsystem embedding scheme.

3.4 Application examples

Computational details. Firstly, we brief the computational details of this section. The ASET(2) embedding was implemented in the open-source package FORTE.[48] The one- and two-electrons integrals necessary to run the CASSCF computations were obtained from PSI4.[159] All DSRG computations were performed using FORTE. The zeroth-order symmetry adapted perturbation theory (SAPT0) computations in Sec. 3.4 are computed using PSI4.[77, 62, 159] Guess active orbitals used in CASSCF optimizations were selected using the atomic valence active space (AVAS) technique. [149] Unless otherwise specified, all computations adopted the cc-pVDZ basis set[189] and the DSRG flow parameter in both ASET and DSRG-methods was set to $s = 0.5\ E_{\mathrm{h}}^{-2}$. For each example, the ASET(2)-[MR-LDSRG(2)] computations employed the density-fitting (DF) implementations using the cc-pVDZ-JKFIT and -RI auxiliary basis sets for CASSCF and environment DSRG computations, respectively.[68, 195, 176, 70] For all pentyldiazene, octene and O_2–benzene examples, the core 1s orbitals of C, N, and O are frozen in the CASSCF computations and in the ASET(2) procedure. The equilibrium geometries of 1-octene and pentyldiazene were optimized at the B3LYP/cc-pVDZ level of theory[12, 104] using PSI4. The benzene geometry is taken from CCCBDB without further optimization.[83]

N-N bond dissociation in pentyldiazene. We begin by investigating the performance of the ASET(2) method on bond-breaking processes by computing the nitrogen double-bond dissociation of pentyldiazene.[139, 76, 71] For this example, we select an active space that spans four N=N bonding/antibonding orbitals $(\sigma, \sigma^*, \pi, \pi^*)$ obtained from a CASSCF(4e,4o) computation.

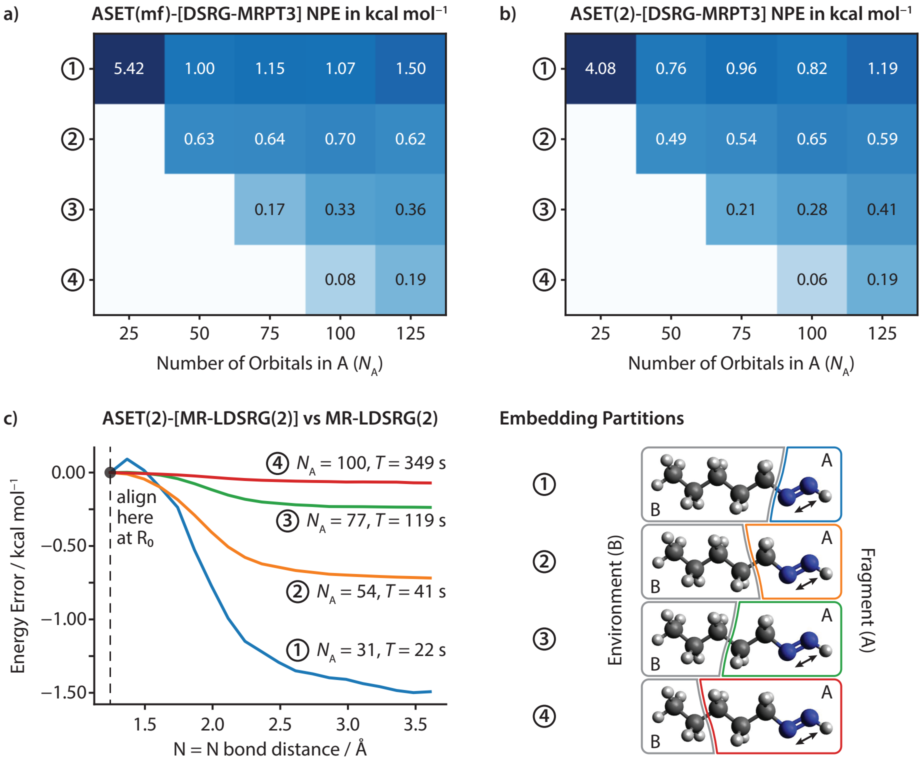


Figure 3.3 Pentyldiazene bond dissociation computed with ASET(2)-[DSRG-MRPT3] and ASET(2)-[MR-LDSRG(2)]. a) The Non-Parallelism Error (NPE) of ASET(mf)-[DSRG-MRPT3] computations to full DSRG-MRPT3 using different number of orbitals in A (N_A) and different fragment definitions. b) The corresponding ASET(2)-[DSRG-MRPT3] results. c) The error curves of ASET(2)-[MR-LDSRG(2)] against full MR-LDSRG(2); results plotted against N=N bond length (Å). Each curve shows a different partition of the pentyldiazene molecule and the partition threshold is fixed at $t = 0.5$. The curves are shifted to align at the R$_0$ geometry (the gray circle). Timings for a single update of the MR-LDSRG(2) amplitudes are shown by the T values (in seconds). All computations here use DSRG flow parameter $s = 0.5\ E_\mathrm{h}^{-2}$ and cc-pVDZ basis.

The standard approach for selecting the fragment orbitals includes all those orbitals that when projected onto the fragment basis have a norm greater than a fixed threshold t. Four different fragment definitions are investigated: 1) $-N_2H$, 2) $-CH_2N_2H$, 3) $-(CH_2)_2N_2H$, and 4) $-(CH_2)_3N_2H$. In this test, we use DSRG-MRPT3 as the fragment solver, and quantify the relative errors using the non-parallelism error (NPE). The NPE measures the degree to which a potential energy curve is parallel to a reference curve, independently of any constant energy shift, and it is defined as

$$\text{NPE} = \max_r \Delta E(r) - \min_r \Delta E(r) \tag{3.33}$$

where $\Delta E(r)$ is the deviation of the energy from it reference value at a N=N bond distance r (obtained by translating all atoms accordingly).

To compare the performance of ASET(mf) and ASET(2), we build two NPE heatmaps (Fig. 3.3a and Fig. 3.3b) using DSRG-MRPT3 as a fragment solver and different number of fragment orbitals (N_A = 25, 50, 75, 100, 125). The heatmap entries are shown only when the number of fragment orbitals is greater than or equal to the number of basis functions on the fragment. A comparison of the two heatmaps shows that although ASET(2)-[DSRG-MRPT3] generally yields a smaller NPE than ASET(mf)-[DSRG-MRPT3]—with improvements as large as ca. 1.5 kcal mol^{-1}—the gains in accuracy are limited and not guaranteed to be systematic. For example, using the $-(CH_2)_2N_2H$ fragment, ASET(mf)-[DSRG-MRPT3] has a slightly smaller NPE (by at most 0.05 kcal mol^{-1}) than ASET(2)-[DSRG-MRPT3] when N_A = 75 and 125.

The approach followed in ASET to select the fragment orbitals is based on applying a fixed threshold to the eigenvalues of the fragment projector [Eq. (2.2)]. With the standard threshold ($t = 0.5$), the fragments for the four embedding partitions (1–4) contain 31, 54, 77, and 100 orbitals, and the ASET(2)-[DSRG-MRPT3] NPE is equal to 1.28, 0.36, 0.14, and 0.06 kcal mol^{-1}, respectively. Hence, the fragment $-\mathrm{CH_2N_2H}$ is sufficient to achieve a chemically accurate NPE (error less than 1 kcal/mol) using ASET(2).

A typical scenario when running embedded computations is that of having only a fixed amount of computational resources available, which imposes limits on the number of fragment orbitals that can be treated with a high-level computations. In this case one may ask: How should the fragment be chosen to minimize the NPE? To answer this question, we consider the error trends in the heatmaps moving along the vertical axis (corresponding to a fixed value of N_A). For both versions of ASET, we find that NPE decreases as the fragment includes more atoms, suggesting that a good alternative to the standard threshold-based partitioning may be to fix N_A and maximize the size of the fragment (while ensuring that N_A is greater than the number of basis functions on the selected fragment). However, when the fragment size is fixed and more orbitals are included (moving along the horizontal axis in the top plots in Fig. 3.3) we observe that the accuracy of a computation does not always increase. This non-monotonic behavior is likely caused by the fact that the orbitals are ordered according to their overlap with the projector and not their contribution to their energetic contribution.

Next, we consider the cost-saving enabled by ASET(2) when using a more expensive fragment solver. Here we consider the MR-LDSRG(2) method, which requires the iterative solutions of equations that scale proportionally to the sixth power of the number of fragment orbitals and that can quickly become a computational bottleneck in applications to large molecules. Dissociation curves as a function of the N=N bond length are shown in Fig. 3.3c. The left panel of this figure shows the energy difference between the ASET(2)-[MR-LDSRG(2)] and full MR-LDSRG(2) energy for all fragment partitionings using a cc-pVDZ basis. In this case, fragments are selected using the partition threshold $t = 0.5$ at all geometries. The dissociation curves computed with ASET(2)-[MR-LDSRG(2)] are smooth and continuous for all partitions and rapidly converge to the full MR-LDSRG(2) result. The NPE for the $-N_2H$ fragment is 1.56 kcal mol^{-1}, and it is further reduced to 0.72, 0.24, and 0.07 kcal mol^{-1} when one, two, or three CH_2 groups are included into the fragment, respectively. Compared to the full MR-LDSRG(2) dissociation energy (113.8 kcal mol^{-1}), the error introduced by embedding is around 1% of the dissociation energy with $-N_2H$ fragment, and it is only 0.1% with the larger $-(CH_2)_3N_2H$ fragment. These results demonstrate a robust and consistent converging behavior for ASET(2), which is similar to the one seen for frozen mean-field ASET(mf) embedding.[71]

The computational timings for different partitions are also presented by the T values in Fig. 3.3c. Because MR-LDSRG(2) is an iterative scheme, we report both the average time for an iteration and the average number of iterations required for each fragment choice (obtained from computations of the entire dissociation curve). The single-step cost for the $-\mathrm{N_2H}$, $-\mathrm{CH_2N_2H}$, $-(\mathrm{CH_2})_2\mathrm{N_2H}$, and $-(\mathrm{CH_2})_3\mathrm{N_2H}$ fragment grows rapidly, from about 22 seconds to 349 seconds. The respective average number of MR-LDSRG(2) iterations are 10, 9, 10, and 12, showing that different partitionings do not significantly affect the convergence rate of the MR-LDSRG(2). Compared to the cost of the embedded MR-LDSRG(2) computation, the cost of preparing the ASET(2) dressed Hamiltonian is minor. These tests demonstrate the potential of using ASET(2) to reduce the computational cost of multireference computations, especially when the fragment solver is a computationally expensive method such as FCI or MR-LDSRG(2).

Lastly, we discuss the contribution of the environment correlation $[\delta E^{\mathrm{B}}_{\mathrm{c,ASET(2)}}]$ term. The differences in NPE from ignoring $\delta E^{\mathrm{B}}_{\mathrm{c,ASET(2)}}$ are -0.15, -0.37, -0.15, -0.05 kcal/mol for the $-\mathrm{N_2H}$ (31 N_{A}), $-\mathrm{CH_2N_2H}$ (54 N_{A}), $-(\mathrm{CH_2})_2\mathrm{N_2H}$ (77 N_{A}), and $-(\mathrm{CH_2})_3\mathrm{N_2H}$ (100 N_{A}) fragments, respectively. It is interesting that for the N=N dissociation path of pentyldiazene, ignoring the $\delta E^{\mathrm{B}}_{\mathrm{c,ASET(2)}}$ term actually reduces the NPE and improves the results significantly.

$\pi \to \pi^*$ Excitation energy of 1-octene. Our next example is an application of ASET(2) to study the $\mathrm{S_0}$ to $\mathrm{S_1}$ ($\pi \to \pi^*$) local-excitation of 1-octene, a system used in several previous embedding studies.[120, 71] Our computations treat the terminal $-\mathrm{C{=}CH_2}$ as the fragment and the rest of molecule

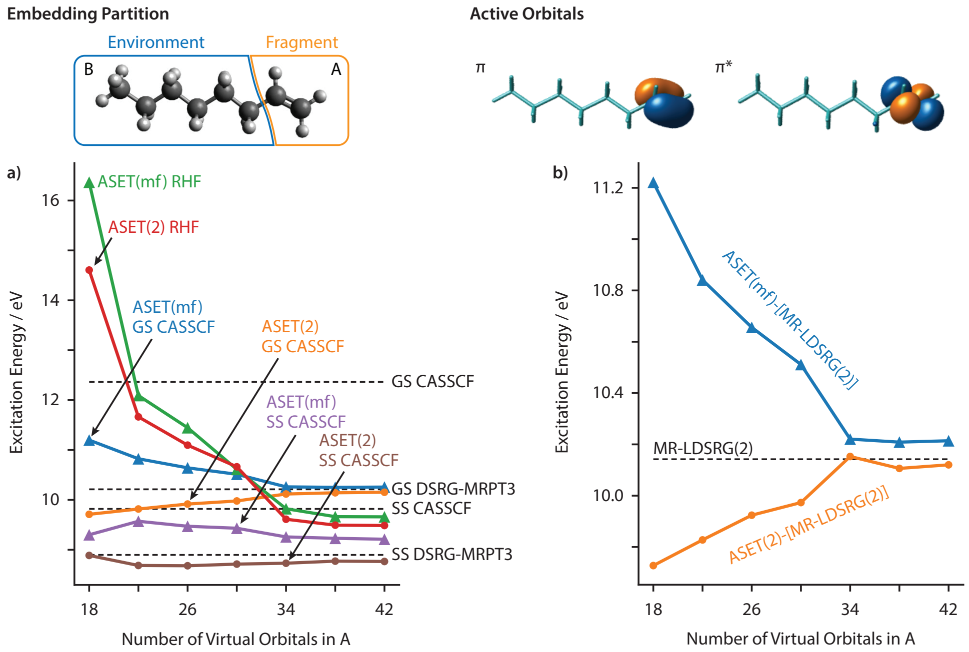


Figure 3.4 S_0 to S_1 vertical excitation (in eV) of 1-octene computed with ASET(mf)-[DSRG-MRPT3], ASET(2)-[DSRG-MRPT3], ASET(mf)-[MR-LDSRG(2)], and ASET(2)-[MR-LDSRG(2)]. a) Excitation energy results of ASET(mf)-[DSRG-MRPT3] and ASET(2)-[DSRG-MRPT3] plotted against different numbers of virtual orbitals included in the fragment (A), computed using three different reference wavefunction: RHF, CASSCF ground state wavefunction (GS), and CASSCF state-specific wavefunction (SS) for both S_0 and S_1. Full CASSCF and DSRG-MRPT3 energies are shown using dashed horizontal lines. b) A comparison of ASET(mf)-[MR-LDSRG(2)] and ASET(2)-[MR-LDSRG(2)] excitation energy versus different numbers of virtual orbitals included in the fragment. Full MR-LDSRG(2) results are shown by the dashed line. Virtual orbitals are selected according to eigenvalues of projector in Eq. (2.2) over all MOs. ASET(2) computations are performed with all external blocks zeroed. All computations use cc-pVDZ basis and DSRG flow parameter $s = 0.5\ E_{\mathrm{h}}^{-2}$.

as the environment. We use this partitioning to accentuate the differences between the methods compared here.

This example illustrates the complexity of treating multiple electronic states with an embedding theory, and in particular the important role played by orbital relaxation effects. Excitation energies computed with various combinations of reference orbitals, embedding scheme, and fragment solvers as a function of the ASET selection threshold, are shown in Fig. 3.4a. Together with these data, we also report the excitation energy computed with CASSCF and the DSRG-MRPT3 using ground-state (GS) or state-specific (SS) orbitals. The simplest embedding approach we consider uses ground state RHF orbitals as a starting point for ASET. As shown in Fig. 3.4a, with GS RHF orbitals both ASET(mf) (green curve) and ASET(2) (red curve) show large errors when the fragment space is small, and slightly underestimate the DSRG-MRPT3 (GS) excitation energy.

Employing ground-state CASSCF orbitals, leads to more accurate excitation energies, especially in the case of ASET(2) (orange curve) even for a small number of fragment orbitals. As the fragment orbital space increases in size, the ASET results based on RHF and ground-state CASSCF orbitals tend to cluster around the DSRG-MRPT3 results computed with ground-state CASSCF orbitals. This clustering reflects the fact that in both cases the active orbitals are not optimized to describe the $\pi \rightarrow \pi^{*}$ singlet state. The last set of data that we analyze is based on a state-specific CASSCF treatment of both the ground and excited state. These results show very fast convergence for ASET(2) (brown curve), while a somewhat erratic behavior for ASET(mf) (purple curve). Among all of these combinations, both the ASET(2) based on ground-state or state-specific CASSCF orbitals offer well-converged excitation energies with modest fragment spaces and improve upon their mean-field analogs.

Next, we compare ASET(mf) and ASET(2) combined with a more accurate high-level MR treatment based on the MR-LDSRG(2). Fig. 3.4b shows the ASET(mf)-[MR-LDSRG(2)] and ASET(2)-[MR-LDSRG(2)] excitation energy versus the number of fragment virtual orbitals, and compares them to the full MR-LDSRG(2) excitation energy. We see again that one advantage of ASET(2) over ASET(mf) is the more rapid convergence of the excitation energy with number of fragment orbitals. For example, with 18 fragment virtual orbitals, ASET(mf)-[MR-LDSRG(2)] predicts the excitation energy to be 11.22 eV. The corresponding ASET(2)-[DSRG-MRPT2] prediction (9.73 eV) is much closer to the full MR-LDSRG(2) result (10.14 eV). To reach an excitation energy with similar accuracy, it is necessary to include at least 30 virtual orbitals in the ASET(mf)-[MR-LDSRG(2)] computation (10.51 eV).

Lastly, we investigate the errors introduced by neglecting the contributions from the three-body cumulants (λ_3) to the ASET(2) energy. To this end, we have computed the change in excitation energy when including λ_3 in the ASET(2)-[DSRG-MRPT2] approach. Computations using 18, 22, and 26 fragment virtual orbitals show that the shift in excitation energy is of the order of 0.001 eV or less. We have also tested the effect of neglecting the environment correlation energy [$\delta E^{\mathrm{B}}_{\mathrm{c,ASET(2)}}$] and found its effect on the excitation energy to be of the order of 0.001 eV for 18–26 fragment orbitals and gradually becoming smaller as more virtual orbitals are included in the fragment.

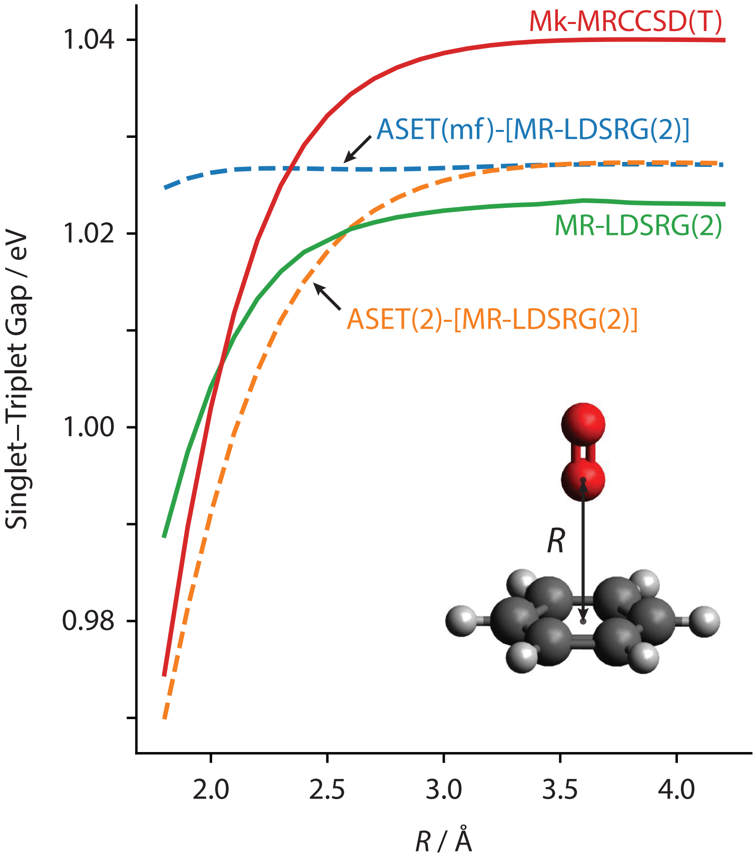


Figure 3.5 Singlet–triplet gaps of the perpendicular O_2–benzene complexes computed using ASET(mf)-[MR-LDSRG(2)], ASET(2)-[MR-LDSRG(2)], Mk-MRCCSD(T), and the full MR-LDSRG(2). The variable R is defined as the distances from benzene center of mass to the closest oxygen atom of O_2. The singlet–triplet gap is computed as $E_{\delta \mathrm{ST}} = E_{\mathrm{S}} - E_{\mathrm{T}}$. All DSRG computations use the cc-pVDZ basis and the DSRG flow parameter $s = 0.5\ E_{\mathrm{h}}^{-2}$.

Interaction energy and singlet–triplet gap of the O_2–benzene complex. In the examples consider so far, a mean-field treatment of embedding already captures the dominant effects of the environment on the fragment, and adding second-order corrections via ASET(2) improves the accuracy, especially in computations of excited states. However, there are situations where a perturbative treatment beyond the mean-field level is necessary to achieve a qualitatively correct description of molecular properties. In this example, we study the O_2–benzene complexes, a weakly interacting system used in early DFT studies to benchmark the importance of dispersion corrections.[65, 186]

Our first test focuses on the singlet-triplet gaps of O_2 for both parallel and perpendicular geometries as a function of the distance from the benzene molecule. The active space for this system is comprised of the valence orbitals of the two oxygen atoms, with starting guesses selected using the AVAS technique.[149] We employ the ASET(mf) and ASET(2) schemes to compute the singlet–triplet gap of O_2 (fragment, A) interacting with benzene (environment, B), using MR-LDSRG(2) as the higher-level multireference theory. The singlet–triplet gap, defined as $\Delta E_{\mathrm{ST}}(R) = E_{\mathrm{S}}(R) - E_{\mathrm{T}}(R)$, for the perpendicular configuration as a function of the O_2–benzene distance (R) is shown in Fig. 3.5. Both O_2 and benzene geometries were kept fixed as R varies. For this example, reference interaction energies were computed using Mukherjee's multireference coupled cluster with singles, doubles, and perturbative triples [Mk-MRCCSD(T)].[47] In the Mk-MRCCSD(T) computations, the reference wave function was defined by an active space containing two π^* orbitals of O_2. Using this active space, the triplet state Mk-MRCCSD(T) potential energy curve of the O_2–benzene system is found to be in excellent agreement with the one from single-reference CCSD(T).

Figure 3.5 shows that ASET(mf) predicts almost no change in the gap as a function of R, while ASET(2) predicts a relative stabilization of the singlet state as large as 0.05 eV in the range of R values between 1.8 Å and 3.5 Å. In the range of geometries considered, the full MR-LDSRG(2) computations show a 0.03 eV stabilization of the singlet state relative to the triplet ground state, while Mk-MRCCSD(T) predicts this value to be slightly larger (0.06 eV). Compared to these reference results, only the embedding computations based on ASET(2) shows a good quantitative agreement. This example shows that in certain cases, the contribution of weak correlation terms from the environment cannot be captured thoroughly by a simple frozen-core embedding scheme based on a mean-field or a CASSCF reference. In this example, low-order perturbative corrections that account for weak interactions with the environment seem to capture the bulk of the differential contributions in the singlet–triplet gap.

The ASET(2) scheme not only improves computations of local properties of the fragment, but can also produce more accurate fragment-environment interaction potentials. To illustrate this point we compute the ground-state interaction energy, defined as $E_{\text{int}} = E_{\text{T}}(R) - E_{\text{T}}(R = 1000\,\text{Å})$, for the parallel configuration of O_2–benzene, since it is more stable than the perpendicular one. The experimental interaction energy of triplet O_2 with benzene is around 1 kcal/mol,[186] and it is dominated by weak Van der Waals interactions that cannot be modeled by mean-field treatments without the inclusion of dispersion corrections.

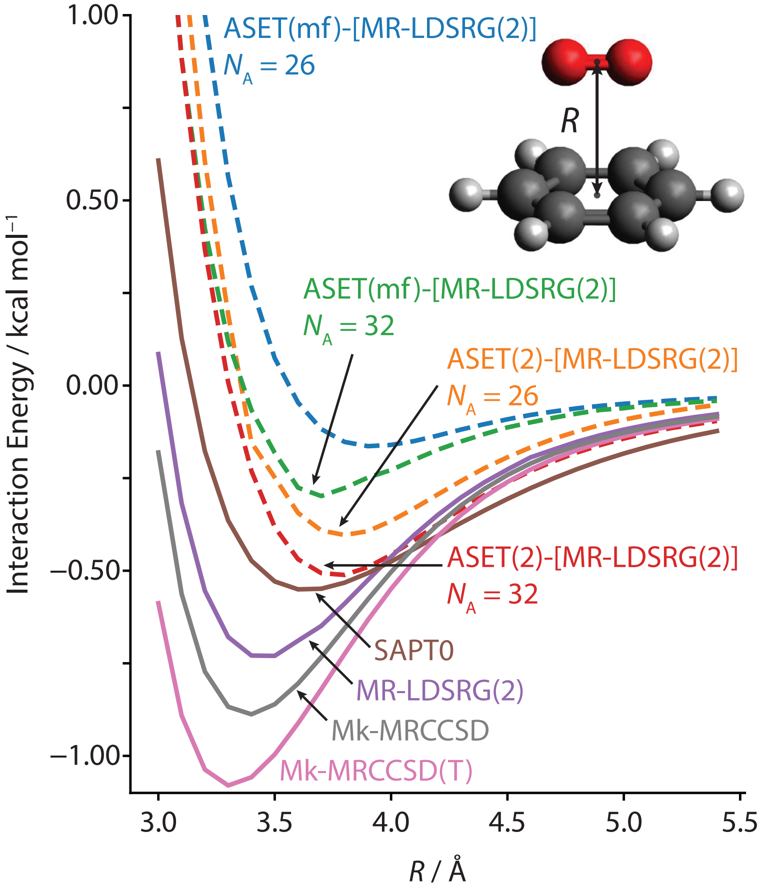


Figure 3.6 Interaction energy of parallel O_2–benzene complexes computed using ASET(mf)-[MR-LDSRG(2)], ASET(2)-[MR-LDSRG(2)], full MR-LDSRG(2), SAPT0, Mk-MRCCSD, and Mk-MRCCSD(T). The variable R is defined as the distances from benzene center to the geometrical center of O_2. The interaction energy is $E_{\text{int}} = E_{\text{T}}(R) - E_{\text{T}}(R = 1000\,\text{Å})$. All computations use cc-pVDZ basis. In DSRG computations the flow parameter is set to $s = 0.5\ E_{\text{h}}^{-2}$.

Following the previous example, we treat O_2 as the fragment and benzene as the environment. Figure 3.6 shows the interaction energy of triplet O_2 and benzene at their respective fixed geometries computed using ASET(mf) and ASET(2) as the embedding method and MR-LDSRG(2) as the high-level multireference theory. For both ASET approaches, we report computations with 26 fragment orbitals (obtained using a threshold $t = 0.5$) and with 6 additional benzene π/π^* orbitals. We also provide reference curves computed with Mk-MRCCSD(T) and an open-shell version of symmetry-adapted perturbation theory (SAPT0).[77, 62]

Using the default threshold, the mean-field ASET(mf) captures only a fraction of the dispersion interaction, resulting in very shallow potential for the O_2–benzene complex. The ASET(2)-[MR-LDSRG(2)] curve shows a modest improvement over the corresponding mean-field treatment, capturing qualitatively the profile of the full MR-LDSRG(2) curve, with this agreement becoming more accurate at long distances ($R > 4.5$ Å). However, including the additional benzene π/π^* orbitals into the fragment leads to significant improvements in the ASET(mf) and ASET(2) curves. In particular, the ASET(2) curve (red) predicts an equilibrium interaction energy similar to that of SAPT0 (ca. -0.5 kcal mol^{-1}) and is indistinguishable from the Mk-MRCCSD(T) one when $R > 4.5$ Å. These results show that the combination of a higher-level embedding treatment of fragment-environment interactions with a more targeted choice of the fragment orbital space may provide a path to performing accurate computations of interaction energies.

3.5 Summary

In this chapter, We developed a one-shot second-order active-space embedding theory [ASET(2)] that accounts for the interaction of the fragment and environment beyond the level of frozen-core embedding methods. In ASET(2), the fragment-environment interaction is included via a canonical transformation truncated to second-order that approximately decouples (block diagonalizes) the full Hamiltonian (fragment + environment). The ASET(2) procedure produces a downfolded Hamiltonian for the fragment space that accounts for fragment-environment interactions and electron correlation in the environment treated at the second order.

We test the ASET(2) procedure on three systems to assess the improvement brought to computations of bond dissociation curves, excited states, and weakly interacting open-shell systems. Our first example—the N=N bond dissociation of pentyldiazene—shows that when changes in the electronic structure are localized on the fragment, ASET(2) brings only modest improvements over the ASET(mf) frozen mean-field embedding in the computed dissociation curves. In our second example—the excitation energy of 1-octene—the response of the environment to changes in the fragment electronic structure plays an important role. In this case, ASET(2) brings consistent and systematic improvements over ASET(mf), and yields converged excitation energies even with relatively small fragment spaces. In our last example—the O_2–benzene system—we demonstrate that ASET(2) can capture important fragment-environment weak interactions in systems where dispersion plays an important role.

In summary, the ASET(2) scheme offers a one-shot approach to improving frozen mean-field embedding theories via unitary canonical transformations. An unexpected and interesting outcome of our study is the observation that improving the description of fragment-environment interaction does not always lead to significant improvements in accuracy. This observation, in turn, suggests that mean-field frozen embedding schemes like ASET(mf) may already be sufficiently accurate for a wide range of applications. The remaining major source of errors in embedding approaches originates from partitioning the system into fragment plus environment, suggesting the need for improved partitioning schemes. As shown in our analysis of pentyldiazene, an optimal partitioning strategy should prioritize identifying the largest fragment compatible with a given number of fragment orbitals. Chapter 5 discusses possible improvements to the partitioning strategy using an AO-label-based projector and clustering techniques. We will also explore whether the ASET scheme can be used for basis set extrapolation, where, instead of using atoms to define a projector, we select the fragment (A) as a subset of the larger basis sets and put the rest of basis functions to the environment (B). An interesting avenue of future inquiry is the extension of ASET to nonperturbative schemes, like the random-phase approximation, which could be useful to describe fragments coupled to extended metallic systems. ASET(2) also shares similar features with downfolding schemes recently introduced in the context of quantum computing to reduce the qubits requirements.[11, 9, 126, 38] Therefore, future work could also explore the application of ASET to bridge real-world chemistry problems with quantum computers.

Chapter 4

Studying CO inversion on NaCl surface using ASET

4.1 Introduction to CO inversion

In this chapter, we will describe an application of ASET to study CO inversion on NaCl(100) surface. This chapter is mainly based on an ongoing paper, which Dr. Meng Huang contributes to the vibrational analysis sections. Firstly, we brief the background of this problem. The transformation between molecules and their isomeric forms is an insufficiently studied area since this process involves the switching of quantum states.[37] Recently, the isomerization of CO molecule on a NaCl surface is recently observed in experiments.[103] From the vibrational emission spectra, signs of configuration switch are clearly observed between C-down and O-down isomer.

The experiment also indicates that the isomerization happens when both isomers are in high-lying vibrationally excited states. Several theoretical studies have been conducted to simulate and verify the experimental observation. Chen et al. computed Potential energy surfaces (PESs) of CO on NaCl(100) under different coverage, using a periodic model and Density Functional Theory (DFT).[25] They discovered that the energy order of the C-down and O-down isomer changes at high vibrationally-excited states ($v \approx 40$). Sinha and Saalfrank verified their discovery and computed the 2-D and 3-D anharmonic vibrational eigenstates.[157] Nandi et al. performs a dynamics study for this isomerization using a finite CO-NaCl cluster model.[129] They highlight that the isomerization may happen at larger CO-NaCl distances than the conventional isomerization saddle point.

Those studies explain the experiment well; however, they all computed the potential energy surfaces using single-reference theories, while CO bond stretch involves significant multideterminantal effects.[34, 188] Single-reference theory can only predict the qualitatively correct C-O stretch (r) PES around the equilibrium distance r_0, thus limiting their ability to study high-lying excited states explicitly. This problem can be solved using a multireference theory for the PES computations. Using multireference computations, an accurate description of the CO stretch will be available for a wider range of r, including the extreme r values.

In multireference theory, the orbitals are partitioned into active and inactive orbitals, and in the small active orbitals, the full Hamiltonian can be built and diagonalized, while the inactive orbitals will be treated approximately.[117] Studying sizable systems like CO-NaCl adsorption with multireference theory is challenging. In this case, the CO-CO interactions and CO-NaCl(100) interactions are dominated by weak van der Waals (vdW) interactions; using higher-level computations on CO alone is insufficient to describe the small variations of the potential energy surface (PES). However, the newly developed quantum embedding theories are possible solutions. As mentioned in previous chapters, in quantum embedding, a system is partitioned into the fragment and environment, and the orbital spaces after partition can be treated with high-level and low-level theory, respectively. Various quantum embedding schemes have been developed and successfully applied to large chemical systems.[185, 64, 105, 95, 94, 190, 18, 51, 132, 20, 32, 179, 2, 16, 184, 133, 21, 98, 61, 82, 57, 196] The Active Space Embedding Theory (ASET) is a quantum embedding scheme that incorporates both the active space partition and quantum embedding partition.[71] It is shown to reduce the cost of multireference computations significantly while still keeping the results close to the full multireference computations.

This chapter will apply the ASET scheme to study the CO isomerization, using the multireference second-order perturbative Driven Similarity Renormalization Group (DSRG-MRPT2) theory as the fragment solver.[107, 109] DSRG is a series of intruder-free multireference theory, guaranteeing its robustness. This chapter demonstrates how the combination of multireference theory and quantum embedding can generate large-scale, highly accurate potential energy surfaces.[43, 111, 68, 112] This procedure would allow us to study highly vibrationally excited state explicitly. It will provide a verification of previous studies and new insights on the CO isomerization system.

This chapter is organized as follows: In Sec. 4.2, we briefly review the ASET embedding scheme and presented the computational details for quantum-mechanical computation, classical potential, and vibrational analysis. In Sec. 4.3, we tested different fragments for the embedding computation. In Sec. 4.3.1, we use the optimal model to perform a potential-energy-surface (PES) scan for both single-cO model and a 1/1 full coverage model, and discuss C-down O-down transition and desorption. In Sec. 4.3.2, we analyze the vibrationally excited states based on the PES. In Sec. 4.4, we summarize the results and raise insights.

4.2 Theory Review

4.2.1 Embedding model for CO absorbed on NaCl

In this section we describe the multilevel embedding model used to compute the PES of CO adsorbed on a NaCl surface. Since modeling the CO molecule over a broad bond length range is key to understanding the isomerization process, we start from a complete-active-space self-consistent field (CASSCF) calculation of CO and a small NaCl cluster. Following Boese and Saalfrank,[17] we consider a cluster containing two 3×3 NaCl layers, with a Na central atom in the surface layer closest to the CO. The active space is chosen to include eight valence orbitals of CO that maximize the overlap with the atomic 2s and 2p orbitals of C and O, using the atomic valence active space (AVAS).[148] When applied to an isolated CO molecule, this procedure yields the active space CASSCF orbitals shown in Fig. 4.1.

These orbitals are always included in the high-level embedding treatment.

After generating CASSCF orbitals, we apply the mean-field version of our active space embedding theory (ASET).[71] As described in Chapter 2, the ASET scheme is used to localize and separate orbitals into two sets: fragment (A) and environment (B) orbitals. This partitioning employs the atomic basis to localize and partition orbitals, and requires the user to specify a list of atoms assigned to the fragment. To accurately reproduce the interaction of CO with the surface, the orbitals that belong to the CO molecule and the closest Na and Cl should be included in the fragment space (A). In Sec. 4.3, we test several embedding models that include the CO molecule and differ in the extent to which the Na and Cl atoms of the surface are included in the fragment atom list.

The CASSCF active orbitals (**A**) are by default included in the fragment orbital space. The core (**C**) and virtual (**V**) virtual orbitals are instead separately rotated and partitioned into subspaces belonging to the fragment and the environment. This rotation aims to maximize the overlap

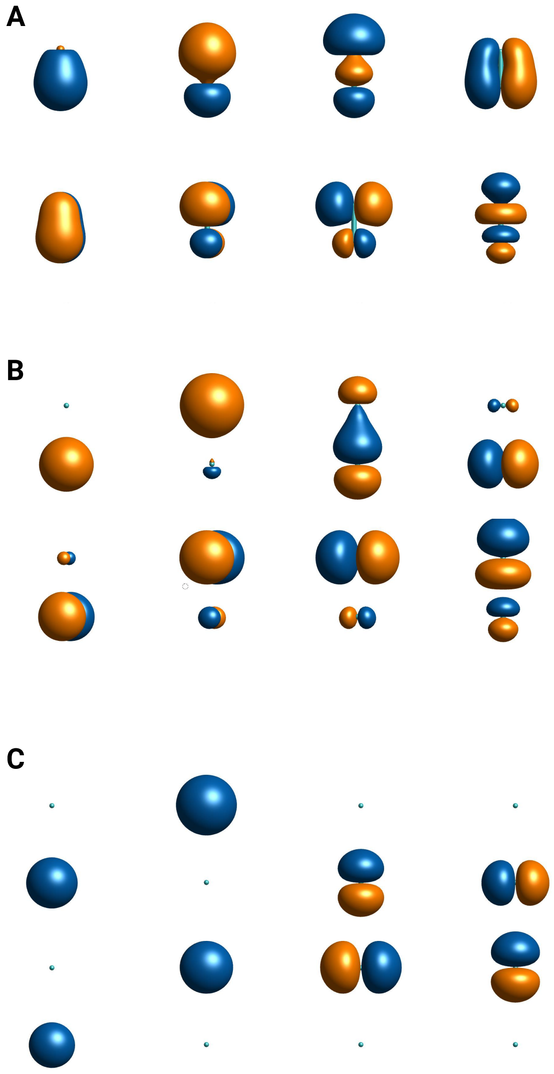


Figure 4.1 The active orbitals of CO around the equilibrium bond distance (A, 1.15 Å), in the recoupling region (B, 1.95 Å), and in the dissociation limit (C, 2.75 Å). The starting orbitals are selected using the AVAS procedure.

of the fragment orbitals with the basis functions located on the fragment atoms. A localized orbital is assigned to the fragment if, after projection onto the orbital basis, its norm is greater than a threshold t. This criterion can be interpreted as selecting orbitals for which the probability of finding an electron on the fragment is greater than t. Note that the active orbitals are not mixed in this transformation. Due to the size of the NaCl surface, we optionally freeze some of the fragment-core and environment-occupied orbitals, in which case these are excluded from the partitioning procedure. A detailed description of this procedure may be found in Refs. 71 and 72.

After partitioning and localization, ASET produces an effective Hamiltonian for the fragment, which provides the input to a high-level multireference treatment. The mean-field version of ASET accounts for the fragment-environment interaction with a static effective one-electron potential. Although this treatment neglects instantaneous fragment-environment fluctuations, a study that introduce this missing effect at the second-order level in perturbation theory found that the improvement in the energetics is small.[72]

Our high-level multireference treatment is based on the driven-similarity renormalization group (DSRG) second-order perturbation theory (DSRG-MRPT2). The DSRG-MRPT2 is derived assuming a diagonal Fock operator. One feature that characterizes the DSRG, is a controllable diagonalization of the Hamiltonian that helps avoid intruder-state problems and depends on the so-called flow parameter s.[42] Compared to other multireference perturbation theories like CASPT2 and NEVPT2,[4, 5] DSRG-MRPT2 has the advantage of being robust to small denominators and requiring fewer computational resources as the energy depends only on the three-body density matrix of the reference.[107, 4, 5] Benchmarks show that these advantages come at the expense of slight loss of accuracy compared to CASPT2 and NEVPT2.[107, 111, 68]

4.2.2 Classical model for external potentials

To simulate the long-range interaction between CO and NaCl both along the surface and deep into the bulk, we include the electrostatic (external) potential due to surrounding Na^+ and Cl^- ions treated as point charges. The quantum mechanics (QM) cell we used in our computations includes a NaCl bulk Na_9Cl_9. As shown in Fig. 4.2, the QM cell is not a classical unit cell since it has positive (+) and negative (-) phases. This choice of the cluster guarantees the symmetry of the QM cluster, reducing the number of points necessary to construct a Potential Energy Surface (PES) significantly. As previous research suggested, the electrostatic terms dominate the CO-NaCl and CO-CO interactions.[103, 129] Therefore, we construct our external electrostatic potential using point-charge, which would take the long-range interactions into account. The placements of the point charge are expanded layer-by-layer; each new layer extends both in 100 plane (XY) and in (-1, 0, 0) direction (-Z), as shown in Fig. 4.2. The electrostatic energy contribution of the Nth layer to the CO inversion is estimated through the charge-dipole interaction:

$$E_{\mathrm{L}_N} = -\sum_{j \in \mathrm{L}_N} \frac{q_j \mu \cos\theta}{{R_e}^2}, \tag{4.1}$$

where q_j is the point charges in L_N; we use +0.536 for Na^+ and -0.536 for Cl^-, as optimized in the previous chapter (Chapter 2).[71] μ is the CO dipole, the dipole is evaluated at different r using CASSCF density matrices. The dipoles are 0.1269 a.u. (0.3226 D) and -0.3885 a.u. (-0.9874 D) at $r = 1.132$ and $r = 1.832$, respectively. R_e is the distance between the point charge and CO center of charge, and θ is the angle between 1) the dipole vector and 2) the vector from q_j to CO center-of-charge. Note that the CO center-of-charge has a different trajectory with center-of-mass during the rotation. The dispersion interactions also plays an important role in this system, especially for the first MM layer. Since all short-range dispersion will be handled in the embedding fragment using DSRG-MRPT2, the only term we need to correct is the long-range attractive potential. Based on Ref. 125, we use:

$$U_{\mathrm{disp}} = -\sum_{i \in CO} \sum_{j \in \mathrm{Env}} \frac{C_{ij}}{r_{ij}^6}, \tag{4.2}$$

where C_{ij} are 383.3, 3935.9, 256.6, and 2633.0 for C...Na^+, C...Cl^-, O...Na^+, and C...Cl^-. The unit is KJ/mol*Å^6. The set Env includes ASET(mf) environment (8 Na^+ and 5 Cl^- ions) plus all added MM layers. This formula is similar to a D2 correction in DFT.[66] Here we test how the number of layers affect the C-down O-down differences. The energy contributions are also shown in Fig. 4.2.

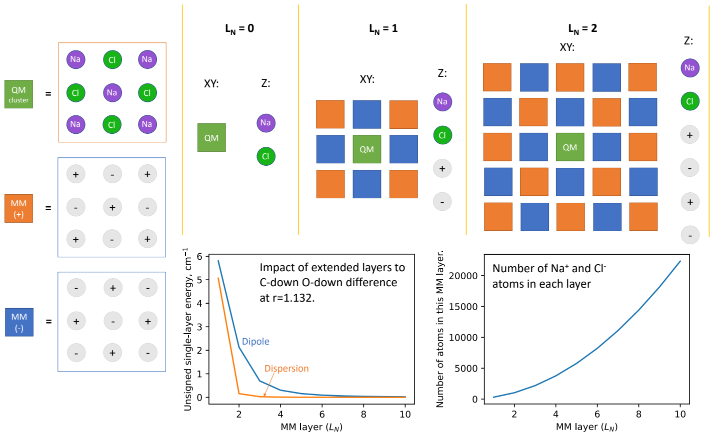


Figure 4.2 An illustration of the QM and MM model. The left plot labels the QM cell and two corresponding MM cells. The upper-right figure illustrates our definition of layers. The diagram below them shows how much the dipole interactions and dispersion interactions each layer contributes. The sizes of each layer (number of atoms) are also shown.

It can be seen that the dipole contributions are less than 1 cm^{-1}for $L_N > 3$, and a converging behavior is observed. The dispersion contributions decay faster; it is under 1 cm^{-1}after the first layer. Therefore, in the main-text computations, we will use L$_N = 1$ to construct the potential. Additional corrections are considered for the adsorption energy analysis, as discussed in Sec. 4.3.1. Here we estimate the maximum error we can make with this finite-cluster approximation. When r = 1.832 Å, the CASSCF dipole reaches 0.3882 a.u., according to Eq. 4.1, the maximum error introduced by ignoring faraway ($L_N \geq 2$) Na$^+$ and Cl$^-$ is 7.9 cm^{-1}, which is not significant enough to affect any analysis. The QM-MM Hamiltonian will be:

$$h_i^{\text{QM-MM}} = h_i^{QM} - \sum_{j \in \text{L}_1} \frac{q_j}{\mathbf{r_i} - \mathbf{R_{e,j}}} \tag{4.3}$$

As shown in Fig. 4.3, in the following computations, this classical external potential accounts for 17 replicas of the finite NaCl cluster included in the quantum mechanical computation (153 Na^+ and 153 Cl^- atoms in total).

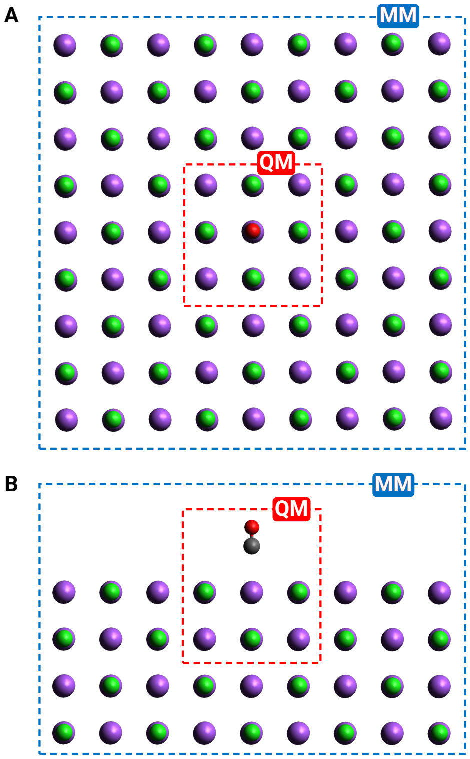


Figure 4.3 Top (A) and side (B) view of the QM/MM model for CO-NaCl(100). Carbon and oxygen atoms are represented with gray and red, while Na^+ and Cl^- ions are colored with purple and green.

4.2.3 Vibration models

In order to investigate the inversion of highly vibrationally excited CO, we computed the vibrational eigenstates using three reduced-dimensionality models. Keeping the position of the Na and Cl atom fixed, there are six degrees of freedom for the C and O atoms; in this chapter, we decompose the coordinates of CO as a center-of-mass (CM) plus rotations and C-O bond stretching. The six coordinates we employ are shown in Fig. 4.4.

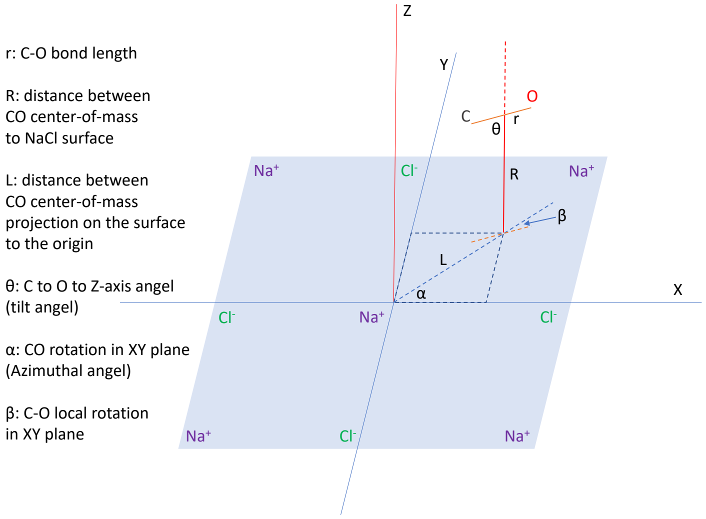


Figure 4.4 Definition of all 6 coordinates that encompass all possible CO movements on the NaCl surface.

The simplest model accounts only for the CO stretching mode along the CO distance (r). The corresponding vibrational Hamiltonian is simply

$$\hat{H}_{\mathrm{1D}} = -\frac{\hbar^2}{2\mu}\frac{\partial^2}{\partial r^2} + V(r) \tag{4.4}$$

where μ is the reduced mass of CO and $V(r)$ is the potential computed at fixed values of the other five coordinates.

The other two models account for the couplings of the CO stretching mode with other degrees of freedom. They are taken from a previous study[157] on CO-NaCl based on a DFT potential. The first one involves the CO distance and the CO tilt angle (θ), where $\theta = 0$ corresponds to CO oriented with the C atom pointing to the surface. The corresponding model Hamiltonian, $\hat{H}_{2\mathrm{D}(r,\theta)}$ can be written as

$$\hat{H}_{2\mathrm{D}(r,\theta)} = -\frac{\hbar^2}{2\mu}\frac{\partial^2}{\partial r^2} - \frac{\hbar^2}{2\mu r^2}\frac{\partial^2}{\partial \theta^2} + V(r,\theta) \tag{4.5}$$

Here the kinetic energy coupling between the two vibrations is represented by the dependence of the reduced mass on the bond length of the CO molecule. The second model involves the CO stretching mode and Z-axis traslation of the CO center of mass (R). The corresponding model Hamiltonian, $\hat{H}_{2\mathrm{D}(r,R)}$ can be written as

$$\hat{H}_{2\mathrm{D}(r,R)} = -\frac{\hbar^2}{2\mu}\frac{\partial^2}{\partial r^2} - \frac{\hbar^2}{2M}\frac{\partial^2}{\partial R^2} + V(r,R) \tag{4.6}$$

where M is the total mass of the CO molecule.

To diagonalize these three model Hamiltonians, we employ the one-dimensional and two-dimensional Discrete Variable Representation (DVR) method, using the representation of Colbert and Miller.[29] We impose periodic boundary conditions on the tilt angle ($\theta \in [-\pi, \pi]$) and use a grid of evenly spaced points. The number of grid points is 201 for the 1D model, 59×101 for the 2D(r,θ) model, and 59×59 for the 2D(r,R) model. The values of potential energy at each DVR grid point are evaluated using cubic spline interpolations of *ab initio* energies. The grids used provide a compromise between the ability to fully diagonalize the Hamiltonian matrix and the accuracy of the energy levels, which are converged to less than 0.1 $\mathrm{cm}^{-1}$for highly vibrationally excited states of CO up to quantum number 25.

4.2.4 Computational details

All ASET and DSRG-MRPT2 computations reported in this chapter were performed with FORTE,[48] using an efficient implementation based on density-fitted integrals.[111] All computations used integrals, reference orbitals, and electrostatic potentials obtained from PSI4.[159] For CASSCF and DSRG-MRPT2 computations, the density fitting basis sets for C and O atoms are cc-pVTZ-JKFIT (CASSCF) and cc-pVTZ-RI (DSRG-MRPT2), while for Na and Cl atoms the smaller cc-pVDZ-JKFIT and cc-pVDZ-RI bases were selected.[189, 176, 178] For the geometry comparison in Sec. 4.3.1, we use def2-SVP basis. The core orbitals of C, O, Na and Cl are frozen in all computations.

4.3 Model benchmarking and geometry optimization

We begin by first benchmarking the accuracy of the ground-state potential energy curve of CO computed with the DSRG-MRPT methods and compare it to other electronic structure methods used in previous studies. In Fig. 4.5, we show the potential curve for a single CO and bond distances ranging from 0.83 to 2.03 Å using two DFT functionals (PBE, B3LYP),[138, 13, 163] restricted MP2, and restricted CCSD(T), and the DSRG-MRPT2/3. Note that the PBE and B3LYP computations cannot be converged beyond the recoupling region (r values larger than ca. 1.61 Å and 1.73 Å).

The curve in Fig. 4.5 shows no significant difference around the equilibrium distance; however, significant deviations start to appear in the recoupling region ($r >$ ca. 1.5 Å). In particular, B3LYP and PBE computations cannot be converged beyond a certain point, while MP2 and CCSD(T) start to show major deviations after 1.62 Å and 1.75 Å, respectively. From the potential energy curves, we performed DVR computation to obtain vibrational constants and an estimate of the maximum vibrational level (v_{max}) that each method allows us to calculate. The results are shown in Table 4.1. Due to convergence issues, the potentials generated from DFT cannot be used to compute the threshold for CO inversion on NaCl(100), which is believed to be around vibrational states $v = 25$. MP2 and CCSD(T) gen-

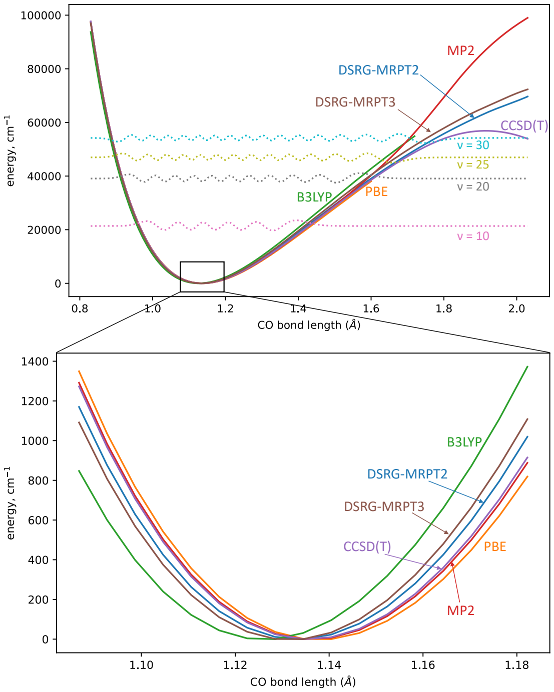


Figure 4.5 Dissociation curve of a single CO in the gas phase computed using different methods. The curves are aligned with their respective minimum energy. The scan is done from 0.83 Åto 2.03 Å, using 50 points in total (more points around 1.13 Åto capture the minimum). Due to convergence difficulties and lost of continuity, PBE data beyond 1.61 Å and B3LYP data beyond 1.73 Å are not presented. The vibrational wavefunctions (10,20,25,30) computed using DSRG-MRPT2 and DVR are shown in dotted lines. More detailed curves are enlarged and shown in the bottom subplot.

erally underestimate vibrational constants respectively by ca. 33 and 21 cm^{-1}compared to the experimental value. However, both DSRG-MRPT2 and DSRG-MRPT3, show small deviations from experimental frequencies (ca. 2 and 17 cm^{-1}, respectively). This analysis highlights the necessity of using multireference methods to accurately compute the potential energy

Table 4.1 Stretching frequencies of gas-phase CO computed using different methods. The quantity v_{max} indicates the maximum vibrational level that can be computed with each method. classical potential and empirical dispersion are not included.

Model	v_{max}	ω_e (cm^{-1})	$\omega_e x_e$ (cm^{-1})
PBE	20	2129.8	13.5
B3LYP	25-30	2208.4	12.4
MP2	>30	2136.8	16.0
CCSD(T)	30	2148.6	12.6
DSRG-MRPT2	>30	2167.7	12.8
DSRG-MRPT3	>30	2186.4	12.5
Exp.^{a}		2169.8	13.3

a: Experimental value taken from Ref. 103.

curve of CO, especially in the high-energy region relevant to the inversion of CO.

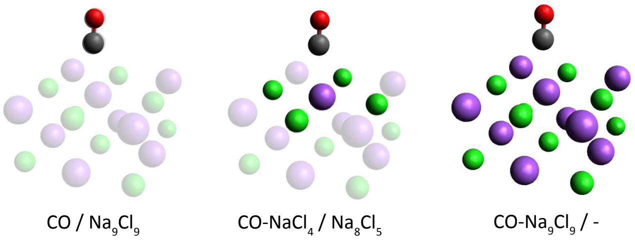


Figure 4.6 Different cluster models tested. Purple ball indicates Na$^+$ and green ball indicates Cl$^-$. The transparent atoms are considered as ASET environment. The blue circle marks a classical electrostatic field as indicated in Fig. 4.2.

In the next step, we identify a model for the CO-NaCl(100) system that is sufficiently accurate and computationally feasible. In Figure 4.6 we show the two fragment/environment partitionings of a Na_9Cl_9 cluster used in the embedding approach and the full Na_9Cl_9 cluster. From these we generate

Table 4.2 Comparison of the embedding models shown in Fig. 4.6. The R is optimized using one-dimensional scan where all other coordinates are fixed. The ν_0 (C-d) and ν_0 (O-d) are computed using a 1-D DVR along r using 100 points between $r = 0.83$ Å and $r = 1.53$ Å.

Model	Fragment	MM Charges	R (C-d)	R (O-d)	ν_0 (C-d)	ν_0 (O-d)	$\Delta\nu_0$
CO	-	-	-	-	2147.6	2147.6	0.0
1	CO	No	3.61	3.52	2142.5	2134.5	8.0
2	$CO-NaCl_4$	No	3.39	3.18	2146.1	2133.7	12.4
3	$CO-Na_9Cl_9$	No	3.37	3.14	2147.2	2133.4	13.8
4	CO	Yes	3.71	No adsorption	2140.7	-	-
5	$CO-NaCl_4$	Yes	3.44	3.19	2144.3	2132.7	11.6
6	$CO-Na_9Cl_9$	Yes	3.41	3.18	2145.8	2131.9	13.9
DFT[a]	-	-	3.33	3.11	2136.1	2121.6	14.5
Exp.[b]	-	-	-	-	2149.9	2136.9	13.0

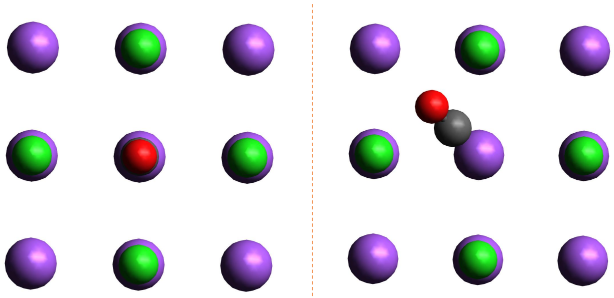
a: from Ref. 25, taking the 1/8 PBE results converted to $C^{12}O^{16}$ using reduced mass.
b: from Ref. 103, taking the extrapolated results for ν_{0-1} in the supporting info.

six models and compare their properties with previous experimental and DFT results.[103, 25] Models **1** and **2** are based on ASET and include in the definition of the fragment CO and $CO-NaCl_4$, respectively. Model **3** contains the entire Na_9Cl_9 cluster and it is equivalent to a full correlated computation without embedding. Models **4**–**6** correspond to models **1**–**3** combined with a classical external charge field.

In Table 4.2 we report the definition of the six models and corresponding properties for a single CO molecule adsorbed on NaCl, together with results for bare CO, previous DFT results, and experimental data.[103, 25] For each model we compute the equilibrium center-of-mass distance from the surface for C-down adsorption [R(C-d)] and O-down adsorption [R(O-d)], the fundamental CO stretching frequency for C-down [ν_0(C-d)] and O-down adsorbed [ν_0(O-d)] CO, and the difference between the two fundamental CO stretching frequencies, $\Delta\nu_0 = \nu_0(\text{C-d}) - \nu_0(\text{O-d})$. The equilibrium R value is determined by a potential energy scan with 0.01 Å spacing and r fixed at 1.13 Å. The corresponding fundamental CO stretching frequency is computed via a 1-dimensional DVR based on a 8-point potential computed in the range $r_e - 0.3$ Å to $r_e + 0.4$ Å at the equilibrium R value.

As can be seen from Table 4.2, the properties obtained from the most elaborate model (**6**) are in excellent agreement with previous DFT results and experimental values, with R(C-d) differing only by 0.08 Å (with respect to DFT) and fundamental frequencies differing from extrapolated experimental values by less than 5.0 cm^{-1}. The ASET-based model **4** treats only the fragment [CO-NaCl$_4$]$^{3-}$ at the DSRG-MRPT2 level and accurately reproduces the features of model **6**. Interestingly, model **4** fails to show O-down absorption for CO, indicating that an explicit treatment of the interaction of CO with the surface atoms in necessary to recover this feature. When comparing the cost of models **5** and **6** we find that computations with the latter are around 10 times more expensive than the former, due to the larger number of correlated MOs in the latter (123 and 292, respectively). The influence of other NaCl cells are approximated using a classical charge field of 153 Na$^+$ and Cl$^-$ ions. The addition of the MM potential makes the model more realistic, especially in terms of adsorption energy, which will be discussed in Sec. 4.3.1. The details of how the ionic charges for this MM potential were derived are reported in the Appendix. Therefore, due to its favorable accuracy/cost ratio, we will employ model **5** for our full PES computations.

Using the embedding model **5**, we optimize the geometry of both C-down and O-down configurations with a grid-search strategy combined with polynomial interpolation, and found that the global minimum in both cases is a perpendicular geometry with $r = 1.1325$ Å, $R = 3.4309$ Å, for C-down ($\theta = 0°$) and $r = 1.1333$ Å, $R = 3.2198$ Å, for O-down ($\theta = 180°$), with $L = 0$ and α and β arbitrary in both cases. This result is in agreement with a previous theoretical study by Meredith and Stone,[125] which predicted that an isolated CO molecule adopts a perpendicular geometry with the C atom above a Na^+ ion and the more recent work of Boese and Saalfrank.[17]. In contrast, in a monolayer both experiments[75] and theory[125, 172, 25, 129] suggest that at low temperatures (below 35 K) the CO molecules assumes a tilted geometry with two possible minima (1×1 and 2×1). Furthermore, at temperatures above 35 K, the CO molecules adopt a 1×1 structure with all molecules perpendicularly aligned to the surface. This structure is estimated by theory to lay only 32 cm^{-1}above the most stable monolayer structure.



Perpendicular:
R = 3.359 Å; r = 1.141 Å; L = 0.0 Å;
θ = 0.0°; α = 0.0° ; ψ = 0.0°.

Tilted:
R = 3.024 Å; r = 1.143 Å; L = 1.463 Å;
θ = 44.8° ; α = 135.0° ; ψ = 0.0°.

Figure 4.7 The perpendicular and tilted geometries optimized using PBE/def2-SVP. The coordinates used are defined in Fig 4.4.

We have found that it is rather difficult for *ab initio* theories to accurately capture the energy difference between the perpendicular and tilted geometry of a single CO molecule. To illustrate this point, we consider two stationary points on the PBE/def2-SVP potential energy surface for the CO$-$Na$_9$Cl$_9$ cluster: the perpendicular and tilted geometries shown in Fig. 4.7. For both geometries we then computed the relative energy, $\Delta E = E_{\text{tilted}} - E_{\text{perpendicular}}$, using various DFT functionals (PBE, SCAN, B3LYP, PBE0)[138, 163, 13, 1, 164] optionally including Grimme's D3 dispersion corrections,[66] MP2, ASET(mf)-[DSRG-MRPT2] and ASET(mf)-[DSRG-MRPT3] based on model **5**,[71, 107, 109] and full DSRG-MRPT2 and DSRG-MRPT3 computations. As shown in Table 4.3, there is a wide spread in the compute relative energies. The PBE, SCAN, B3LYP functionals predict that the tilted geometry is lower in energy, while PBE0 slightly favors the perpendicular one. Adding the empirical dispersion correction (D3) leads to predict the tilted geometry to be more stable for all functionals; however, the magnitude of the D3 correction ranges from a slight destabilization (by 5 cm^{-1}for SCAN) to stabilization up to 288 cm^{-1}(B3LYP) of the tilted geometry. In contrast, MP2 and the ASET-DSRG models are all in agreement with the general consensus between theory and experiment that for a single CO molecule the perpendicular geometry is more stable, predicting that the tilted geometries to lie 176-217 cm^{-1}above the perpendicular one.[75, 172] We have also analyzed the long-range dispersion introduced by MM cells, which is a 20.8 cm^{-1}stabilization of the tilted geometry for the

Table 4.3 The energy difference between C-down perpendicular and tilted geometries ($E_{\text{tilted}} - E_{\text{perpendicular}}$) computed using different methods. The "+MM" indicates a classical potential consisting of 153 Na^{+} and 153 Cl^{-} ions, as discussed in the Appendix.

Method	$E_{\text{tilted}} - E_{\text{perpendicular}}$ (cm^{-1})
PBE + MM	-120.4
PBE-D3 + MM	-291.6
SCAN + MM	-562.5
SCAN-D3 + MM	-557.4
B3LYP + MM	-205.4
B3LYP-D3 + MM	-493.0
PBE0 + MM	36.6
PBE0-D3 + MM	-118.2
MP2 + MM	217.2
ASET(mf)-[DSRG-MRPT2] + MM	213.7
ASET(mf)-[DSRG-MRPT3] + MM	176.3
Full DSRG-MRPT2 + MM	288.2

first MM layer and less than 0.4 $\text{cm}^{-1}$for the second and further layers of NaCl cells. Details of MM cells and dispersion corrections can be found in Sec. 4.2.2. In the following analysis of the CO inversion based on reduced dimensionality models, we will therefore consider the potential expanded around the perpendicular geometry minimum.

4.3.1 Potential energy surface for CO inversion

In this section we investigate the potential energy surface (PES) of CO adsorbed on NaCl(100). We begin by discussing the inversion of an isolated CO on NaCl(100) and then extend our analysis to a CO monolayer. The starting point of our investigation of the single CO is the perpendicular minimum geometry of model **5**. We first investigate the dependence of the energy as a function of the azimuthal angle (α) for three values of the tilt angle (θ = 30, 90, 150 degrees). In scanning the energy, we keep the projection of the CO center-of-mass fixed onto the central Na^+ ion by imposing $L = 0$ and maintain the CO center-of-mass distance from the surface fixed at the optimum distance for the C-down configuration. Fig. 4.8 shows the results of this analysis. We generally note that for all three tilted configurations there are four equivalent minima corresponding to diagonal orientations of the CO molecule ($\alpha = 45^\circ + k90^\circ$, with $k = 0, 1, 2, 3$).

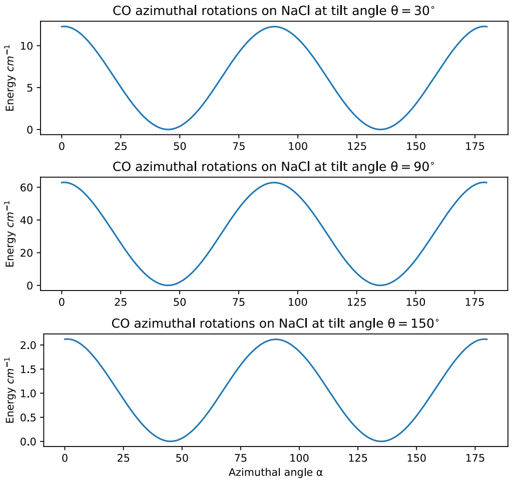


Figure 4.8 CO energy surfaces along azimuthal rotations computed using ASET(mf)-[DSRG-MRPT2] and cc-pVTZ basis.

Based on these considerations, we build a reduced dimensionality potential energy surfaces involving the CO bond length (r), the distance of the CO center of mass from the NaCl(100) surface (R), and the tilt angle (θ) defined as the angle between C-O vector and the NaCl(100) normal vector. In scanning these variables we fix $\alpha + \beta = 45^\circ$ and $L = 0$, which corresponds to a CO oriented diagonally. Our potential energy surface covers r from 0.832 to 1.832 (Å), R from 2.83 to 4.93 (Å), and θ from 0° to 180°. All data are computed using ASET(mf)-[DSRG-MRPT2], with empirical dispersion and external charge fields included. A total of 3072 points are computed for the PES, and the entire scan takes less than 40 hours using 288 CPU cores.

The PES is fitted using Scikit-learn's polynomial regression module with cross-validation (0.8:0.2 split, 5 folds) to prevent over-fitting. The final model consists of 24 parameters. We ensure through cross-validation that all R^2-scores are higher than 0.996, and the root-mean-square errors (RMSE) for all folds are smaller than 0.01 cm^{-1}. The minima and transition states are found in this PES using L-BFGS optimization in SciPy, with boundaries enforced using a exponential penalty beyond the R and r data range.[137, 171] Fig. 4.9 shows four slices of the PES at CO bond lengths $r = 0.872$, 1.132, 1.432, and 1.732 Å using contour lines at 100 cm^{-1} energy intervals.

Both C-down ($\theta = 0^\circ$) and O-down ($\theta = 180^\circ$) CO geometries correspond to minima, with the former being the global one at $r = 1.132$ Å and $R = 3.42$ Å. There are several important observations to make. Firstly, the equilibrium center of mass distance, R, varies significantly with respect to both the bond length and the orientation of the CO. At $r = 0.832$ Å, the C-down and O-down CO display minima at $R = 3.18$ and $R = 3.06$ Å. When the r is increased to 1.132, 1.432, and 1.732 Å, the optimal value of R for C-down CO increase to 3.42, 3.68, and 3.97 Å, respectively. In contrast,

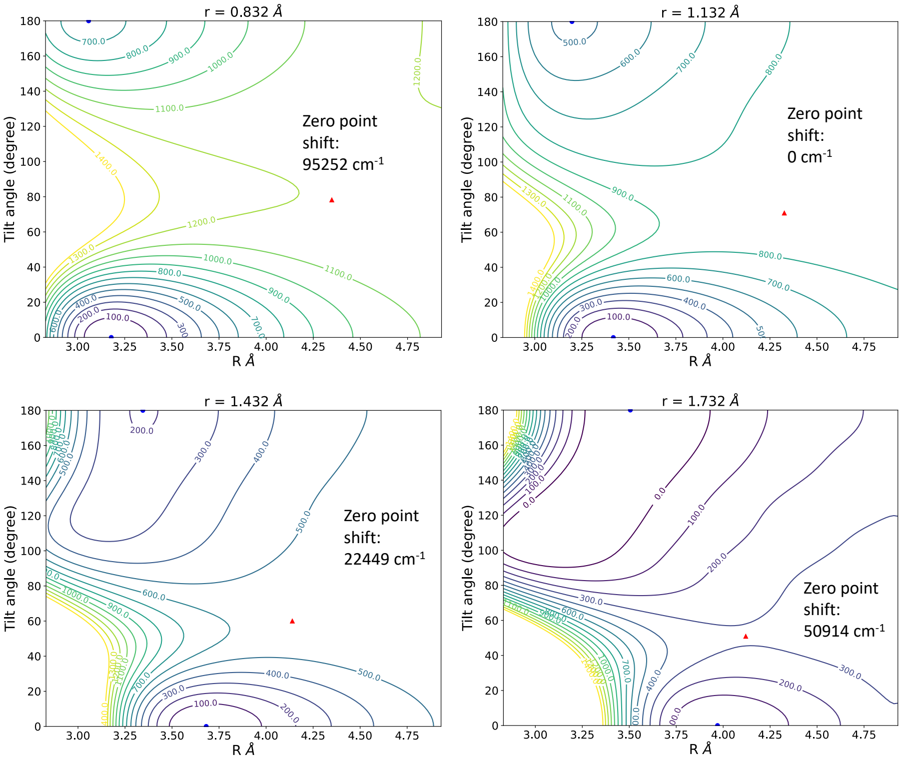


Figure 4.9 Potential energy surfaces computed using the single CO model using ASET(mf)-[DSRG-MRPT2]. The 2-D slices of the PES at $r = 0.832$, 1.132, 1.432, and 1.732 Å are presented. The x-axis is R, and the y-axis is the tilt angle. Each contour line indicates 100 cm^{-1}. Blue balls mark the minima and the red triangle marks the transition state. The shifts from the global minimum for each figure are marked.

the optimal R value for O-down CO does not display such large variations, though its stable adsorption basin becomes flatter with larger r values.

A second important observation concerns the energy difference between the C-down and O-down equilibrium geometries, which is crucial in determining the energetics CO inversion on the NaCl surface. At $r = 0.832$, 1.132, 1.432 Å, the C-down geometries are energetically more stable, but the energy difference goes from 653.7 cm^{-1}down to 479.6 $\text{cm}^{-1}$and 192.6 cm^{-1}, respectively. When $r = 1.732$ Å, the O-down geometry becomes 110.9 cm^{-1}lower in energy than the C-down one. Similar observations are also observed in previous research, where C-down O-down inversion is observed at $r = 1.59$ using PBE and periodic models.[25]

Finally, the height of the barrier between the C-down and O-down configurations along the θ coordinate decreases as r increases. When $r = 0.832$ Å, the barrier height from C-down to O-down is 1196.8 cm^{-1}, while it decreases to 843.5 cm^{-1}, 569.7 cm^{-1}, and 305.7 $\text{cm}^{-1}$when $r = 1.132$, 1.432, and 1.732 Å, respectively. Therefore, the inversion between C-down and O-down geometries is facilitated by a lowering of the barrier when the CO bond length is stretched to 1.7–1.8 Å, in agreement with the dipole-driven inversion mechanism discussed in the literature.[103, 25, 157, 129] Another interesting observation is that the transition states appear at larger R values compared to the equilibrium R value for both the C- and O-down configurations. This barrier change indicates that translation of the CO along the R direction may be coupled to the other degrees of freedom during the inversion process. In the next section, we will dive further into this transition.

The experiment of CO inversion happens with very high CO coverage, where all Na^{+} sites are saturated, forming a monolayer or even bi-layer adsorption.[103] This section will also examine the potential energy surfaces at high coverage, using a slightly modified model. As shown in

Fig. 4.10, four additional CO adsorbed at $R = 3.43$ Å are added to the QM cells. These CO will be assigned as QM environment and fixed at the equilibrium adsorption distances, making the fragment size the same as the previous embedding computations. Using this model, the potential energy surfaces of CO inversion under monolayer adsorption (1/1 coverage) can be simulated. It is worth mentioning that environmental CO makes it more difficult to generate smooth PES without using state-average computations, especially for tilted geometries with r longer than 1.5 Å. Therefore, here we only compute the PES for $r = 1.132$ and $r = 1.432$ Å, together with the r-R PES where θ is fixed at 0° and 180°.

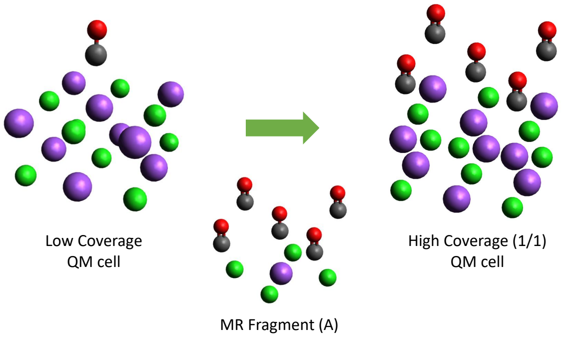


Figure 4.10 The model used to study 1/1 coverage of CO on NaCl surface. All Na^+ are saturated by C-down CO at $R = 3.436$, while only the center CO is allowed to move and other CO are fixed. The fragment is still $CONaCl_4$; the other 4 CO are assigned as the embedding environment.

The noticeable differences between 1/1 coverage PESs and the previous computations (Fig. 4.9) include several aspects: 1) the energy differences between C-down and O-down configurations increase slightly. At $r = 1.432$

Å, the C-down O-down differences are around 192.6 $\mathrm{cm}^{-1}$for single-CO model and 275.8 $\mathrm{cm}^{-1}$for 1/1 high-coverage model. This increase indicates that thermochemically, C-down O-down flip may be harder, and the vibrational excitation level required for this transition is likely to be higher than the single-CO model. 2) The barrier between the two configurations is also slightly higher than the single-CO model, especially with stretched r. 868.8 $\mathrm{cm}^{-1}$and 758.5 cm^{-1}energy are required for the inversion to happen at $r = 1.132$ Å and $r = 1.432$ Å, respectively, comparing to the 843.5 $\mathrm{cm}^{-1}$and 569.7 $\mathrm{cm}^{-1}$for the single CO model. Therefore, dynamically, the CO inversion may be more difficult with high coverage than that for the low coverage.

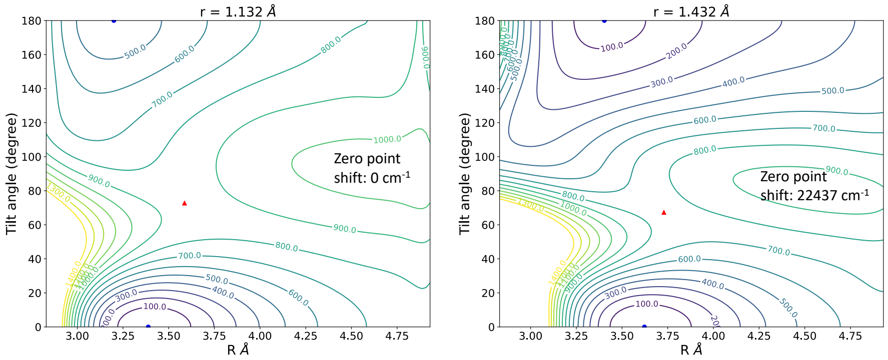


Figure 4.11 The potential energy surfaces for 1/1 coverage CO on NaCl computed using ASET(mf)-[DSRG-MRPT2]. Two slices at $r = 1.132$ Å and $r = 1.432$ Å are presented. Each contour line represents 100 cm^{-1}. Blue balls mark the minima and the red triangle marks the transition state. The shifts from the global minimum are marked.

An interesting question to ask here is facing this barrier, whether CO will desorb instead of flipping to the O-down configuration? In Table 4.4, we computed all adsorption energies using a 1-D scan along R for different r. For the equilibrium CO distance ($r = 1.132$ Å), the C-down adsorption energy is 1262.8 $\mathrm{cm}^{-1}$and 1442.3 $\mathrm{cm}^{-1}$for single and 1/1 model using 1-layer MM cells. However, to compare the desorption with the inversion path, we need the adsorption energy at extended r. Therefore, we investigate the desorption curve for both single CO and 1/1 coverage CO at $r = 1.432$ Å and $r = 1.732$ Å using our model. The results are shown in Fig. 4.12. For $r = 1.432$ Å, the C-down is still preferred for both single and 1/1 CO; the adsorption energy for C-down are 1028.4 $\mathrm{cm}^{-1}$and 1356.2 cm^{-1}, respectively. While for the O-down configuration, the adsorption energy is 849.4 $\mathrm{cm}^{-1}$and 1067.2 $\mathrm{cm}^{-1}$for single and 1/1 CO, respectively. Referring to Fig. 4.11, under 1/1 coverage, the barrier from C-down to O-down is 758.5 cm^{-1}, and the barrier from O-down to C-down is 485.6 cm^{-1}. These barriers are significantly lower than the desorption energy. Note that the inversion barrier presented in our research is an up-limit estimation (since we don't follow the MEP), the actual barrier should be even lower. When we further stretch r to 1.732 Å, the C-down O-down adsorption energy becomes reversed; however, both C-down and O-down adsorptions are still strong. The adsorption energies are 807.7 $\mathrm{cm}^{-1}$and 1218.1 $\mathrm{cm}^{-1}$for single and 1/1 C-down, and 914.6 $\mathrm{cm}^{-1}$and 1236.6 $\mathrm{cm}^{-1}$for single and 1/1 O-down, respectively. Therefore, CO inversions are very likely preferred over desorption in both low and high coverage situations. All above discussions are based on 1 layer of external empirical corrections (1 Disp/MM). One missing factor may be that our model is a finite cluster, which may miss the interactions from the extended NaCl surfaces ($L_N \geq 2$). Using the dipole interaction and empirical dispersion approaches in Appendix, we evaluate the total electrostatic and dispersion impact of $L_N = 2$ to $L_N = 9$. From Table 4.4, the extended cells contribute less than 12 cm^{-1}. It is obvious that the both interactions from the extended NaCl cells do not impact the adsorption energy significantly. Therefore, the CO desorption from NaCl surface is not favorable in our model, even with very high-level vibrational excitations.

Table 4.4 The computed CO adsorption energy. The bare computation indicates model **2** with no empirical corrections; the "$L_N = 1$ Disp/MM" indicates empirical dispersion corrections including the embedding environment and the 1st MM layer and the charge-dipole interactions involving the 1st MM layer; the "$L_N = 9$ Disp/MM" indicates an additional 8 layers (added to "$L_N = 1$ Disp/MM" model) of classical corrections including dispersion and charge-dipole interactions. Definition of those corrections can be found in Appendix.

Configuration	r (Å)	bare	$L_N = 1$ Disp/MM (cm^{-1})	$L_N = 9$ Disp/MM (cm^{-1})
C-down,single	1.132	1140.4	1262.8	1271.3
O-down,single	1.132	713.3	824.3	829.2
C-down,1/1	1.132	1312.7	1442.3	1451.0
O-down,1/1	1.132	898.9	1009.9	1014.8
C-down,single	1.432	917.0	1028.4	1032.4
O-down,single	1.432	742.5	849.4	858.6
C-down,1/1	1.432	1244.8	1356.2	1360.2
O-down,1/1	1.432	964.3	1067.2	1076.4
C-down,single	1.732	703.0	807.7	809.5
O-down,single	1.732	812.0	914.6	925.8
C-down,1/1	1.732	1109.9	1218.1	1219.8
O-down,1/1	1.732	1141.5	1236.6	1247.7

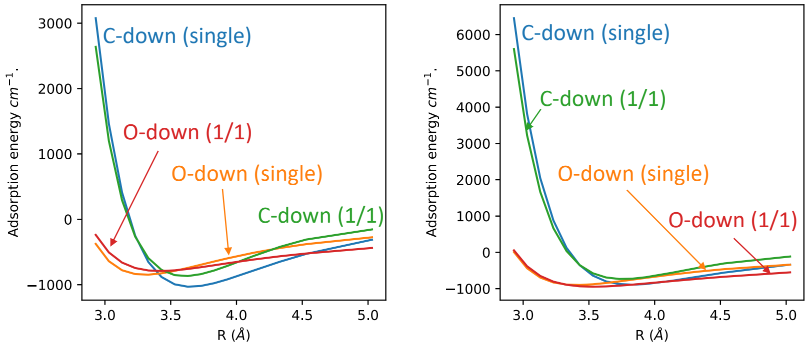


Figure 4.12 The desorption curve for both single CO and 1/1 coverage CO at $r = 1.432$ Å(left) and $r = 1.732$ Å(right). For the 1/1 coverage CO, only the center CO desorb, while other CO are kept fixed. The data are from "$L_N = 1$ Disp/MM" model. All curves are aligned with respect to their electron energies computed at $R = 200$ Å.

4.3.2 Vibrational analysis (single-CO model)

In this section we perform vibrational analysis based on aforementioned CO-NaCl(100) potential for an isolated molecule using different vibration models introduced in Sec. 4.2.3. The eigenvalues of each vibrational model corresponding to the states with stretching quantum number v_r in the range 0–25 are calculated and fitted with a cubic polynomial. The vibrational constants extracted from this fit are listed in Table 4.5. For CO molecule in the gas phase, the vibrational constants ω_e and $\omega_e x_e$ for $^{13}C^{18}O$ from our model (2067.1 cm^{-1}, 11.78 cm^{-1}) are in perfect agreement with the experimentally determined values (2067.8 cm^{-1}, 12.07 cm^{-1});[103] however, our predicted value of $\omega_e y_e$ (4.76×10^{-3} cm^{-1}) is only about half of the experimentally determined one (9.1×10^{-3} cm^{-1}). This suggests that our potential leads to sub cm^{-1}accuracy for the prediction of CO $0 \rightarrow 2$ transition frequency , but can deviate by 11 $\text{cm}^{-1}$for the prediction of CO $23 \rightarrow 25$ transition frequency from the experiment.

The simple 1D model based on the PES slice of C-Down geometry (at R = 3.43 Å with θ = 0°) and O-down geometry (at R = 3.23 Å with θ = 180°) already provides a preliminary description to account for the CO flipping mechanism, where it successfully predicts the red-shift of the C-down CO vibration and the blue-shift of the O-down CO vibration, which has been reported in previous theoretical studies [25, 157]. The energy difference between the O-down CO and the C-down CO, ΔE_0, is 474 cm^{-1} at their vibrational ground state, and this difference will be negative when v_r reaches 25 which means the O-down CO will be thermodynamically more favorable.

The inclusion of the other two degrees of freedom leads to an in-depth understanding of the CO-NaCl system. The 2D(r, θ) model, which involves both the CO stretch and CO inversion based on the PES slice at R = 3.43 Å, shows no significant difference in the CO stretching frequency constants compared to simple 1D models at R = 3.43 Å. This can be rationalized by the localization of the wavefunctions of the states with no vibration excitation on the CO inversion, even though the effective inversion barrier decreases when the CO stretch are highly excited. These wavefunctions are plotted in Figure S1 of the Supporting Information. In this model ΔE_0 is 484 cm^{-1}, where the corresponding ΔE_0 for the two 1D model at R =

3.43 Å is 525 cm^{-1}which indicates a roughly -41 cm^{-1}contribution from the zero-point energy difference in the CO inversion mode at C-down and O-down geometry.

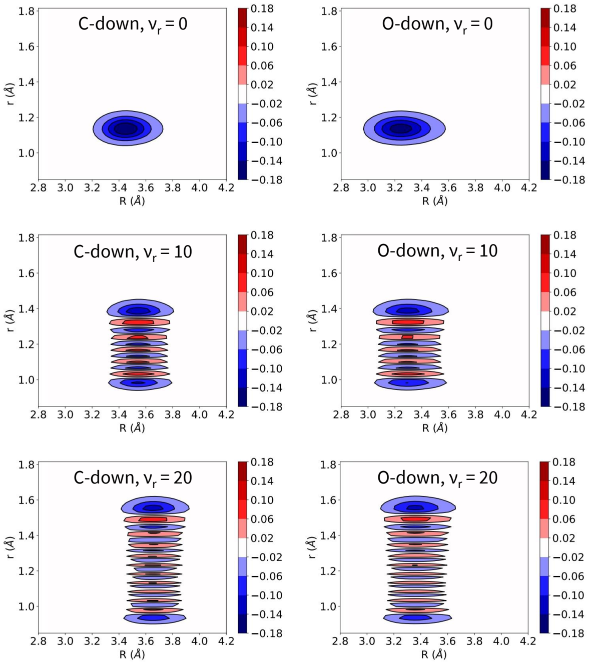


Figure 4.13 Vibrational wavefunctions of $v_r = 0$, 10 and 20 states at C-down and O-down geometries calculated using a two-dimensional mode involving CO stretch and CO translation with respect to the NaCl surface.

On the contrary, the 2D(r,R) model based on PES slice at $\theta = 0$ and 180° lead to a more significant difference. The ω_e constants of the O-down and C-down CO increases for -0.7 cm^{-1}and 0.4 cm^{-1}, which is still relatively close to the 1D value. More importantly, the $\omega_e y_e$ constant of the CO molecules at C-down geometry is only 2.73×10^{-3} cm^{-1}, which is much smaller than the one from 1D model. This is due to a slight desorption of the CO molecules at the C-down geometry when CO stretch mode is excited, which can be seen from Figure 4.13 where the wavefunctions of states with different v_r calculated using the 2D(r,R) model are plotted. The expectation value of R for the wavefunction of C-down CO is 3.45 Å for $v_r = 0$ state, but increased to 3.66 Å when $v_r = 20$. The wavefunction for O-down CO has relatively less desorption, where the expectation value of R increases from 3.21 Å to 3.33 Å as v_r increases from 0 to 20, which is consistent with its less deviated $\omega_e y_e$ vibration constants from the 1D value. This desorption is consistent a simple point charge model, where the repulsion between surface and CO increases as the dipole moment of CO increases. Under this model, the ΔE_0 is 462 cm^{-1} and we are unable to directly observe any CO states at O-down geometry have a lower energy than the states at C-down geometry with the same v_r in our DVR calculation. The constants from the fit shown in Table 4.5 lead to an estimate of $v_r = 37$ for the "flipping" to happen.

4.4 Summary

In summary, we build a quantum embedding model to study the CO inversion PES on NaCl(100) surface using multireference computations. The computational model uses $\mathrm{CONaCl_4}$ as the fragment, $\mathrm{Na_8Cl_5}$ as the QM environment, and $\mathrm{Na_{153}Cl_{153}}$ as the MM environment. We found that the optimal geometry for both C-down and O-down configurations is vertical, i.e., not tilted. As a result, the potential energy surfaces involving three degree-

Table 4.5 The experiment and theoretical vibrational constants for the $^{13}C^{18}O$ vibration. The theoretical constants are computed using different model Hamiltonians based on the DSRG-MRPT2 potential.

Model	ω_e (cm^{-1})	$\omega_e x_e$ (cm^{-1})	$\omega_e y_e$ (10^{-3}cm^{-1})
Gas Phase	2067.1 ± 0.8	11.78 ± 0.07	4.76 ± 1.86
C-down, 1D	2075.8 ± 0.7	11.55 ± 0.07	9.62 ± 1.68
O-down, 1D($R = 3.23$Å)	2063.8 ± 0.8	11.71 ± 0.07	4.49 ± 1.80
O-down, 1D($R = 3.43$Å)	2062.6 ± 0.8	11.74 ± 0.07	3.80 ± 1.80
C-down, 2D(r, θ)	2075.6 ± 0.8	11.56 ± 0.07	9.95 ± 1.76
O-down, 2D(r, θ)	2062.4 ± 0.8	11.74 ± 0.07	3.84 ± 1.80
C-down, 2D(r, R)	2075.1 ± 0.8	11.65 ± 0.07	2.73 ± 1.80
O-down, 2D(r, R)	2064.2 ± 0.8	11.72 ± 0.07	3.15 ± 1.81

of-freedom are generated. We confirm that the C-down configuration is preferable at smaller C-O bond distances from the single-CO PES. In contrast, the O-down configuration will be lower in energy for stretched C-O bond distances (>1.6 Å). The PES also varies significantly with the adsorption distances. Vibrational analysis shows that the vibration-driven flip does happen at high vibrationally-excited states. The exact vibrational level for this flip to happen is around $v = 25$, according to a 2-D DVR analysis using r and θ, or $v = 37$, according to a 2-D DVR analysis using r and R. At high coverage, the energy differences between C-down and O-down geometries are smaller, making the inversion thermomechanically easier. However, the barrier between the two configurations increases, which means the inversion would be dynamically slower. Our model also finds that the CO desorption energies are generally higher than the inversion barriers even with stretched r. This observation indicates that desorption is not likely during the CO inversion process.

The model used in this chapter still has some limitations. Our results match the experiment well for C-down configurations but less accurate for O-down configurations. The reason is mainly due to the O-down configurations being flexible in the experiment; their adsorption is not precisely centered on Na^+ but probably involves translations and azimuthal rotations in an extensive range. Those tilted tendencies may be coming from the CO-CO interactions, which are not computed explicitly in our model. Advanced geometry optimization or dynamic simulation will provide more insights into this problem. With the development of analytical gradient for both quantum embedding and DSRG theory,[106, 173] we hope that this problem can be simulated better in the future.

Chapter 5

Improve orbital partition using modified projectors

5.1 AO basis projector

This chapter will briefly discuss two ongoing works on the possible improvement of ASET. The contents and data in this chapter have not been published.

As discussed in chapter 2 and chapter 3, one main problem of the ASET localization and projection procedure is that it assigns orbitals based on the orbital overlap. However, the overlap is not a good indicator of the correlation energy physically. For example, in the computation of O_2-benzene in Chapter 3, the orbitals before localization and partition are shown in Fig. 5.1. Since we choose O_2 as the embedding fragment, it is thus natural to partition all orbitals centered on O_2 (after localization) to the fragment. However, what would happen if we want to include slightly more orbitals to describe the O_2-benzene interactions better? Chemically, one would guess that adding some of the π orbitals of the benzene would be a good idea. Therefore, we circled six orbitals that are good candidates to add in Fig. 5.1. However, this fragment partition cannot be achieved successfully simply using the threshold-based procedure described in Chapter 2. If we modify the threshold until six more orbitals are included, the orbitals that entered the fragment are circled in Fig. 5.2. It is evident that the partition of the occupied orbitals is as expected (red circle). However, the virtual orbitals are much more problematic: some high-lying virtual orbitals are partitioned into the fragment instead of the π^* orbitals we ex-

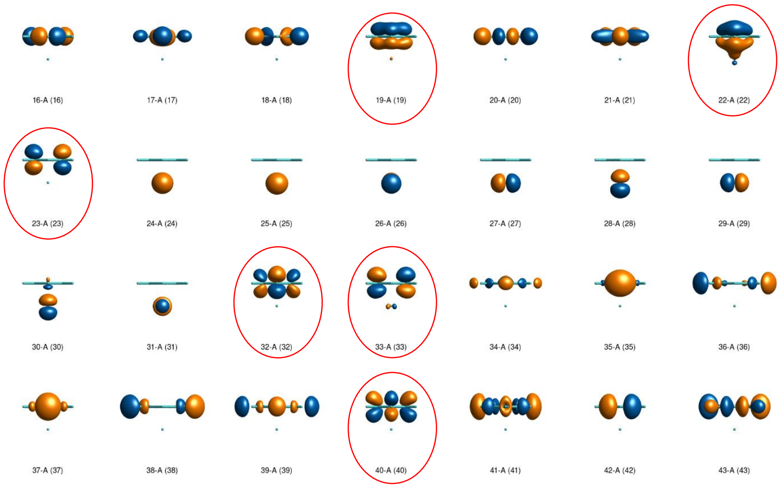


Figure 5.1 Molecular orbitals of O_2-benzene before localization. Red circles indicate the expected choice of additional fragment orbitals.

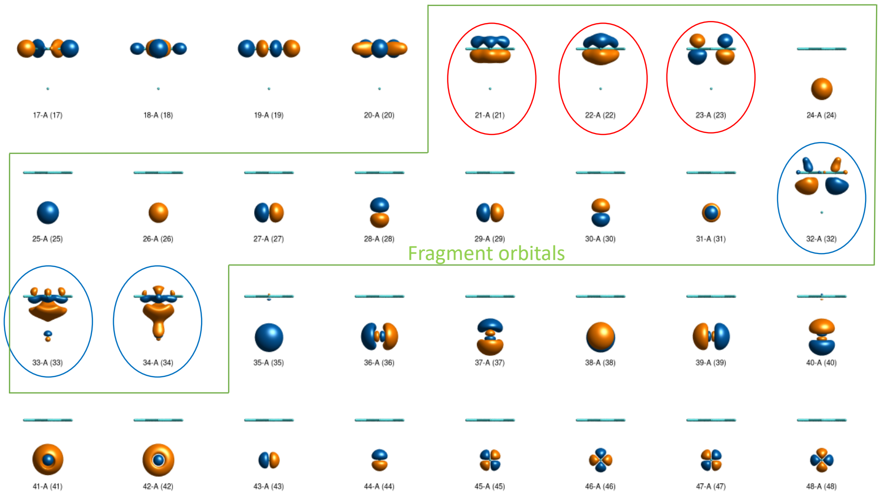


Figure 5.2 Molecular orbitals of O_2-benzene after the localization and partition. Red circles indicate the additional fragment orbitals that are correctly partitioned as expected. Blue circles are orbitals that enter the fragment partition unexpectedly.

pected. This problem is because the partition procedure is based on the maximization of the overlap, which is not a good indicator of orbital energy or importance in many cases. One can manually tweak and put the expected orbitals into the fragment, as we do in Chapter 3. However, a more general solution is to implement an improved projector scheme that allows more flexible and controllable partition. In previous Chapters, we mentioned an atomic valence active space (AVAS) technique, which selects active orbitals automatically using an AO subspace. Inspired by this approach, we propose that the embedding computation can be done similarly, using the AO subspace instead of the atom for the fragment definition. In terms of implementation, the scheme will be the combination of:

- S_a: a set of atoms in the system (C1, H2, O4, ...).
- S_b: a set of AO basis function labels [C2(2pz), H3(1s), ...].

The projector is constructed similarly as

$$\hat{P}^{\mathrm{S_a,S_b}} = \sum_{\mu,\nu \in \mathrm{S_a} \bigcup \mathrm{S_b}} |\chi_\mu\rangle \, S^{-1}_{\mu\nu} \, \langle\chi_\nu| \, . \tag{5.1}$$

For the O_2-benzene system, the selection of S_a and S_b are ["C1", "O1"] and ["C2(2pz)", "C3(2pz)", "C4(2pz)", "C5(2pz)", "C6(2pz)", "C7(2pz)"], respectively. Using this setting and selecting 32 fragment orbitals, we successfully obtained the same result as manually picked orbitals (see Fig. 3.6).

The new projector allows higher flexibility for embedding. We tried two ideas based on this technique: one of them focused on automating the embedding process using unsupervised learning algorithms. The other focused on using ASET as a basis set extrapolation algorithm.

5.2 Overlap-distance-based basis selection and clustering

The first project focused on using clustering, an unsupervised learning algorithm, to create molecular fragments automatically. The goal is to automate the embedding computations to be used directly onto sizable systems without manually setting fragments and tweaking orbitals. Based on the projector proposed in the previous section, we use the AO basis as the label; the clustering process should find several exclusive AO subspaces that will serve as our embedding fragments.

A distance metric is needed to perform clustering. Consider two composites consisting of AO basis sets S_i and S_j; here are many ways to define the distance between them. For test purposes, we define:

$$d_{ij} = -\log(\frac{\sum_{\mu}^{S_i}\sum_{\nu}^{S_j}|S_{\mu\nu}|}{N_{S_i}N_{S_j}}), \tag{5.2}$$

where N_{S_i} and N_{S_j} are the cardinal number of S_i and S_j, and $S_{\mu\nu}$ is the corresponding AO overlap matrix element. This distance metric creates an order so that higher total overlap results in shorter distances and vice versa. There are also flaws in this definition. Firstly, if we consider the distance of an AO basis function with another basis function in the same atom, the distance should be 0.0 physically, but in this definition, it is not. However, this problem is not crucial since we will only use a clustering algorithm that relies on the relative order of the distances for the following discussions. In addition to the AO overlap matrix S, one can also add the AO Fock matrix (F) to the distance matric, with a weight parameter α:

$$d_{ij} = -\log(\frac{\sum_{\mu}^{S_i}\sum_{\nu}^{S_j}(S_{\mu\nu}+\alpha F_{\mu\nu})}{N_{S_i}N_{S_j}}). \tag{5.3}$$

This definition involves some level of orbital energy indications in the process. One can adjust α to change the AO Fock components' importance.

Here we propose one algorithm to cluster the AO subspace based on Kruskal's algorithm.[96] The steps are:

- 1. Convert input molecule into a graph G, where each vertex V represents a basis function or a composite of basis functions. The weight of edges $d_i j$ are the distances defined in Eq. 5.2. We ignore small overlaps (large or infinite d_{ij}).
- 2. Sort all edge weights $d_i j$.
- 3. Select the shortest edges and merge the edge's two vertices.
- 4. Check the number of sets or other criteria to decide whether we terminate the process.
- 5. If the criteria are not met, go back to step 3.

The criteria for termination can be various. The simplest choice is to set a number-of-fragments N_{frag}, when the number of vertices is equal to N_{frag}, we terminate the algorithm. The other approach is to use cluster inertia.

$$I_V = \frac{\sum_{i,j \in V} d_{ij}}{N_V}, \tag{5.4}$$

where $\sum_{i,j \in V}$ is the sum of all edges merged into one vertex, while N_V is the number of vertices merged into this vertex. This metric checks how "tight" a cluster is. If the system has some naturally defined boundaries, I_V will change rapidly around that range. One can use the "elbow" method to determine when it is a good time to terminate the clustering process. There are also situations where the user knows how large a fragment they need. This assertation is especially important when doing multireference computations, where one should estimate how many basis functions are feasible based on the CPU and memory of the device. For example, from personal experiences, without density-fitting, the MR-LDSRG(2) can only compute systems up to 150 basis functions effectively. In this situation, we may need the algorithm to give us fragments whose sizes are smaller than 100 basis functions. Therefore, the algorithm terminates when the number of basis functions in the largest cluster reaches the limitation. We provide three examples to test different scenarios below.

Firstly, we tested the automatic fragmentation procedure on 1-octene and pentyldiazene; both are examples we have seen in Chapters 2 and 3. In this case, we use both the N_{frag} criteria and the basis set function criteria. To simplify the problem, we enforce the AO basis functions that belong to the same atom to stay in the same cluster. As shown in Fig. 5.3, it is obvious that the automatic procedure successfully generates fragments that are well-defined and ready to compute.

Secondly, we tested the situation where a natural boundary exists in the system. First, a $(H_2O)_13$ cluster is placed randomly; then, we perform the clustering procedure of the AO subspace on the whole system. In this case, we scan over different N_{frag} and focus on the inertia of the single cluster. As shown in Fig. 5.4, there is a sudden change of the inertia increasing rate

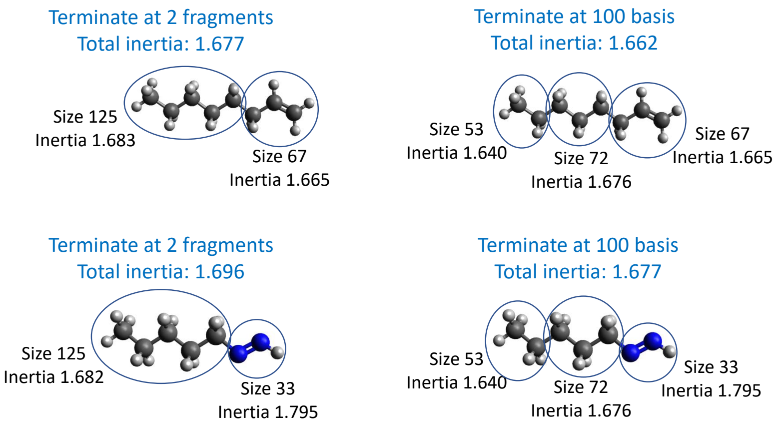


Figure 5.3 Automatically generated fragments for 1-octene and pentyldiazene using both N_{frag} and basis functions limit criteria. The fragment inertia and total inertia are marked.

when $N_{\mathrm{frag}} = 13$. Since there are 13 $\mathrm{H_2O}$ in the cluster, it is very natural to create 13 fragments each includes one $\mathrm{H_2O}$.

Finally, we tested the idea of finding a fragment under a certain budget. Here we investigate the d-d spin gap of $[\mathrm{Fe}(\mathrm{H_2O})_6]_2{}^+$ (Fig. 5.5). Assume our fragment high-level [MR] method can only treat a maximum of 100 basis functions; we set the max cluster size to 100. Another restriction is that the five 3d orbitals must be within the largest cluster. This restriction can be imposed by limiting Kruskal's algorithm to merge edges only when one of the vertices is in the same cluster with the five 3d orbitals.

The spin gaps computed using different schemes are shown in Table. 5.1. It is clear that the clustering procedure is more robust during the partition

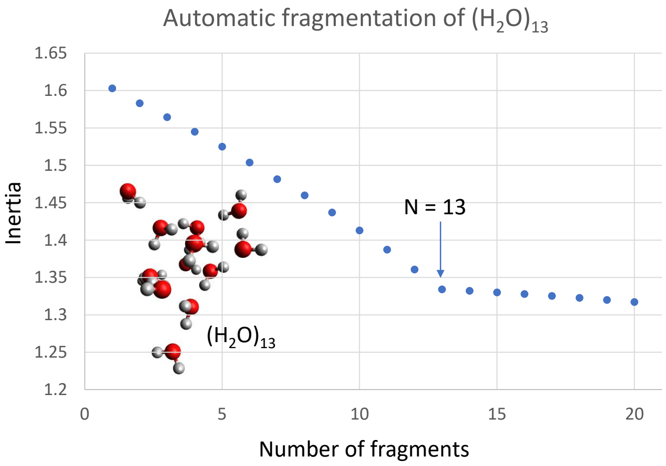


Figure 5.4 Automatic fragmentation of $(H_2O)_{13}$ using clustering. Cluster inertia is used to determine the optimal N_{frag}

.

and localization than the normal procedure. The reason is that in this case, the clustering procedure automatically picks a better AO subspace for the localization, while the normal procedure, which picks the Fe atom as the fragment, is not ideal.

Table 5.1 The d-d gaps of $[Fe(H_2O)_6]_2{}^+$ computed using different embedding schemes and [MR] methods.

Method	d-d Spin Gap (eV)
Full DSRG-MRPT3	1.77
ASET(mf)-[DSRG-MRPT3](Normal partition, t=0.5)	2.80
ASET(2)-[DSRG-MRPT3](Normal partition, t=0.5)	2.33
ASET(mf)-[DSRG-MRPT3](clustering procedure, size $N_A = 100$)	2.26
ASET(2)-[DSRG-MRPT3](clustering procedure, size $N_A = 100$)	1.94

For now, it is still difficult to conclude how good this new algorithm is

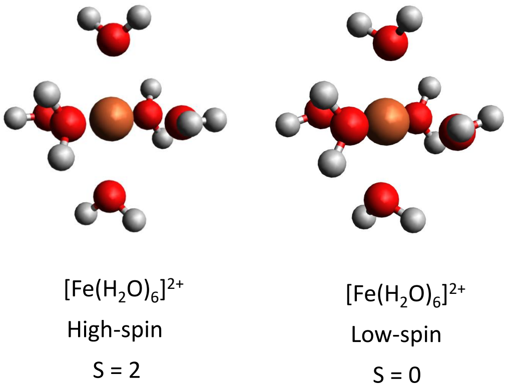


Figure 5.5 The geometries of high-spin and low-spin $[\mathrm{Fe(H_2O)_6}]_2{}^{+}$.

based on those primitive results; more research is needed to structure this idea into formal research.

5.3 Basis-set extrapolation using ASET

The other idea to apply ASET is to use it as a tool for basis set extrapolation. In quantum chemistry, the size of the basis set determines the numerical accuracy of the solution. The general practice is to balance the cost and accuracy by selecting the largest basis set sizes so that the computations can run efficiently. However, using embedding methods, it is possible to compute results similar to a large basis set, with the computational cost equivalent to a smaller basis set. Watson and Chan firstly explore this idea,[174] where they reduce the computational cost to minimum AO basis level while still retaining qualitatively correct descriptions of phenomena involving valence electrons which generally need at least double-zeta basis sets to describe correctly. With the ASET(2) scheme we developed in chapter 3, we can push this idea one step forward: embed any small basis into a large basis. This scheme will provide a new way of extrapolating basis sets closer to the complete basis set (CBS) limit.

We show here a primitive example of this approach. We set the small basis as cc-pVDZ and the large basis as cc-pVQZ. Then the embedding scheme is formulated as follows:

- The whole system (A+B) is computed using cc-pVQZ with a lower-level theory.
- The fragment (A) will be an AO subspace with a size equal to that of cc-pVDZ.
- The environment (B) will be all AO labels in cc-pVQZ basis set but not in the fragment (A).
- Compute the ASET with A embedded in A+B.

One major problem of this scheme is how to pick an AO subspace with a size equal to that of cc-pVDZ from an AO set generated with cc-pVQZ basis? There are two approaches: 1) Pick basis functions in cc-pVQZ with the same ξ in their radical part as cc-pVDZ. 2) After SCF, pick N_A lowest MOs. N_A is set to the size of cc-pVDZ basis set in the whole system.

In the following example, we test the first approach on a C_2 dissociation problem. Fig. 5.6 shows the ASET(2)-[DSRG-MRPT3] dissociation curve of

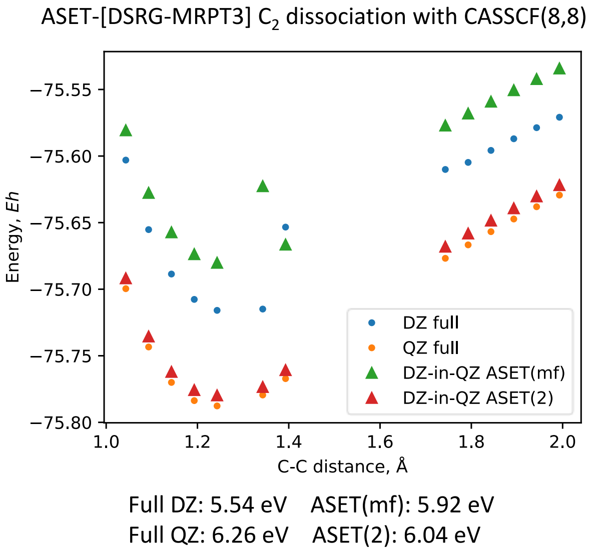


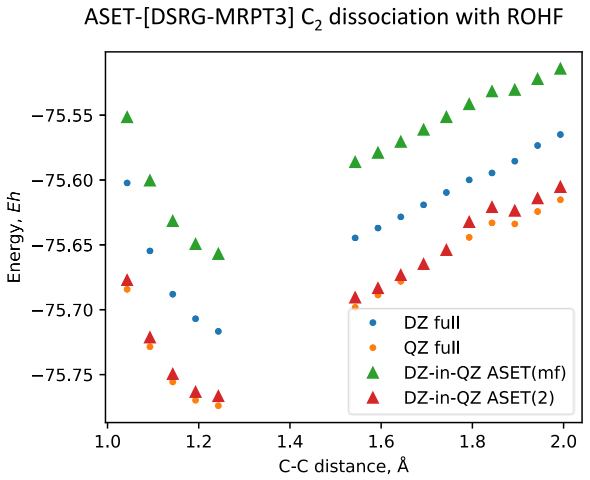
Figure 5.6 C_2 dissociation using DZ-in-QZ ASET. The reference wavefunction is CASSCF(8,8).

C_2 computed using a CASSCF(8,8) reference wavefunction; while Fig. 5.7 uses a ROHF reference wavefunction. As shown in both cases, the DZ-in-QZ ASET(2) results are very close to the full QZ computations. These examples prove the feasibility of this idea.

5.4 Summary and perspectives

This chapter briefly mentioned two directions I have explored to expand the ASET theory. One is the automatic fragmentation using a clustering algorithm; the other involves using ASET for basis set expansion or extrapolation. Both ideas seem to work well, and should be persued in the future.

There are more other ideas that we can explore, including



Full DZ: 6.39 eV ASET(mf): 5.97 eV
Full QZ: 6.64 eV ASET(2): 6.61 eV

Figure 5.7 C_2 dissociation using DZ-in-QZ ASET. The reference wavefunction is ROHF, then a (8,8) active space is used for DSRG computations.

- Combining ASET with quantum computing: using ASET(2) downfolded Hamiltonian as the starting point of various quantum algorithms to introduce environment contributions to the quantum computations.
- Using random phase approximation (RPA) as the mean-field level: This treatment can extend ASET to solid-state problems.
- MR-WFT-in-DFT embedding: incorporate DFT potential into the ASET downfolding process.

To conclude, ASET is a promising method worth more development in the future. However, we expect that the most important impact that ASET will have is enabling interesting applications to systems that could not be studied before. Our embedding method can used for a variety of applications where it is necessary to reduce the cost and focus a high-level multireference computation to a smaller partition of the system. Given the simplicity and versatility of ASET, we expect that it has the potential to become a very impactful method in the landscape of electronic structure theory.

Bibliography

[1] Carlo Adamo and Vincenzo Barone. Toward reliable density functional methods without adjustable parameters: The pbe0 model. *J. Chem. Phys.*, 110(13):6158–6170, 1999.

[2] Alexey V. Akimov and Oleg V. Prezhdo. Large-scale computations in chemistry: A bird's eye view of a vibrant field. *Chem. Rev.*, 115(12): 5797–5890, 2015. doi: 10.1021/cr500524c.

[3] K Andersson, P-Å Malmqvist, Björn O Roos, Andrzej J Sadlej, and K Wolinski. Second-order perturbation theory with a casscf reference function. J. Phys. Chem., 94(14):5483–5488, 1990.

[4] K Andersson, P-Å Malmqvist, and Björn O Roos. 2nd-order perturbation-theory with a complete active space self-consistent field reference function. J. Phys. Chem., 96(2):1218–1226, 1992.

[5] C. Angeli, R. Cimiraglia, S. Evangelisti, T. Leininger, and J.-P. Malrieu. Introduction of n-electron valence states for multireference

perturbation theory. J. Chem. Phys., 114(23):10252–10264, 2001. doi: 10.1063/1.1361246.

[6] Celestino Angeli, Renzo Cimiraglia, and Jean-Paul Malrieu. n-electron valence state perturbation theory: A spinless formulation and an efficient implementation of the strongly contracted and of the partially contracted variants. J. Chem. Phys., 117(20):9138–9153, 2002. doi: 10.1063/1.1515317.

[7] Jie J. Bao, Sijia S. Dong, Laura Gagliardi, and Donald G. Truhlar. Automatic selection of an active space for calculating electronic excitation spectra by ms-caspt2 or mc-pdft. J. Chem. Theory Comput., 14(4):2017–2025, 2018. doi: 10.1021/acs.jctc.8b00032.

[8] Rodney J Bartlett and Monika Musiał. Coupled-cluster theory in quantum chemistry. Rev. Mod. Phys., 79(1):291, 2007.

[9] Nicholas Bauman, Jaroslav Chládek, Libor Veis, Jiri Pittner, and Karol Kowalski. Variational quantum eigensolver for approximate diagonalization of downfolded hamiltonians using generalized unitary coupled cluster ansatz. Quantum Sci. Technol, 2021.

[10] Nicholas P Bauman, Eric J Bylaska, Sriram Krishnamoorthy, Guang Hao Low, Nathan Wiebe, Christopher E Granade, Martin Roetteler, Matthias Troyer, and Karol Kowalski. Downfolding of many-body Hamiltonians using active-space models: Extension of

the sub-system embedding sub-algebras approach to unitary coupled cluster formalisms. *J. Chem. Phys.*, 151(1):014107, 2019. ISSN 0021-9606. doi: 10.1063/1.5094643.

[11] Nicholas P Bauman, Guang Hao Low, and Karol Kowalski. Quantum simulations of excited states with active-space downfolded hamiltonians. *J. Chem. Phys.*, 151(23):234114, 2019.

[12] A. D. Becke. A new mixing of Hartree-Fock and local density-functional theories. *J. Chem. Phys.*, 98:1372–1377, January 1993. doi: 10.1063/1.464304.

[13] Axel D Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, 38(6):3098, 1988.

[14] Simon J. Bennie, Basile F. E. Curchod, Frederick R. Manby, and David R. Glowacki. Pushing the limits of eom-ccsd with projector-based embedding for excitation energies. *J. Phys. Chem. Lett.*, 8 (22):5559–5565, 2017. doi: 10.1021/acs.jpclett.7b02500.

[15] Moritz Bensberg and Johannes Neugebauer. Direct orbital selection for projection-based embedding. *J. Chem. Phys.*, 150(21):214106, 2019. doi: 10.1063/1.5099007.

[16] Gregory J. O. Beran. Modeling polymorphic molecular crystals with

electronic structure theory. *Chem. Rev.*, 116(9):5567–5613, 2016. doi: 10.1021/acs.chemrev.5b00648.

[17] A Daniel Boese and Peter Saalfrank. Co molecules on a nacl(100) surface: Structures, energetics, and vibrational davydov splittings at various coverages. *J. Phys. Chem. C*, 120(23):12637–12653, 2016. ISSN 1932-7447. doi: 10.1021/acs.jpcc.6b03726.

[18] George H. Booth and Garnet Kin-Lic Chan. Spectral functions of strongly correlated extended systems via an exact quantum embedding. *Phys. Rev. B*, 91:155107, 2015. doi: 10.1103/PhysRevB.91.155107.

[19] Baird H Brandow. Linked-cluster expansions for the nuclear many-body problem. *Rev. Mod. Phys.*, 39(4):771, 1967.

[20] Ireneusz W Bulik, Gustavo E Scuseria, and Jorge Dukelsky. Density matrix embedding from broken symmetry lattice mean fields. *Phys. Rev. B*, 89(3):035140, January 2014.

[21] P. J. Bygrave, N. L. Allan, and F. R. Manby. The embedded many-body expansion for energetics of molecular crystals. *J. Chem. Phys.*, 137(16):164102, 2012. doi: 10.1063/1.4759079.

[22] Laimutis Bytautas, Thomas M. Henderson, Carlos A. Jiménez-Hoyos, Jason K. Ellis, and Gustavo E. Scuseria. Seniority and orbital symme-

try as tools for establishing a full configuration interaction hierarchy. *J. Chem. Phys.*, 135(4):044119, 2011. doi: 10.1063/1.3613706.

[23] James M Callahan, Malte F Lange, and Timothy C Berkelbach. Dynamical correlation energy of metals in large basis sets from downfolding and composite approaches. *J. Chem. Phys.*, 154(21):211105, 2021.

[24] Paolo Celani and Hans-Joachim Werner. Analytical energy gradients for internally contracted second-order multireference perturbation theory. *J. Phys. Chem.*, 119(10):5044–5057, 2003.

[25] Jun Chen, Seenivasan Hariharan, Jórg Meyer, and Hua Guo. Potential energy landscape of co adsorbates on nacl (100) and implications in isomerization of vibrationally excited co. *J. Phys. Chem. C*, 124(35):19146–19156, 2020.

[26] Dhabih V. Chulhai and Jason D. Goodpaster. Improved accuracy and efficiency in quantum embedding through absolute localization. *J. Chem. Theory Comput.*, 13(4):1503–1508, 2017. doi: 10.1021/acs.jctc.7b00034.

[27] Daniel Claudino and Nicholas J. Mayhall. Automatic partition of orbital spaces based on singular value decomposition in the context of embedding theories. *J. Chem. Theory Comput.*, 15(2):1053–1064, 2019. doi: 10.1021/acs.jctc.8b01112.

[28] Daniel Claudino and Nicholas J. Mayhall. Simple and efficient truncation of virtual spaces in embedded wave functions via concentric localization. *J. Chem. Theory Comput.*, 15(11):6085–6096, 2019. doi: 10.1021/acs.jctc.9b00682.

[29] Daniel T. Colbert and William H. Miller. A novel discrete variable representation for quantum mechanical reactive scattering via the S-matrix Kohn method. *J. Chem. Phys.*, 96(3):1982–1991, feb 1992. ISSN 0021-9606. doi: 10.1063/1.462100.

[30] David J. Coughtrie, Robin Giereth, Daniel Kats, Hans-Joachim Werner, and Andreas Köhn. Embedded multireference coupled cluster theory. *J. Chem. Theory Comput.*, 14(2):693–709, 2018. doi: 10.1021/acs.jctc.7b01144.

[31] T. Daniel Crawford, Elfi Kraka, John F. Stanton, and Dieter Cremer. Problematic p-benzyne: Orbital instabilities, biradical character, and broken symmetry. *J. Chem. Phys.*, 114(24):10638–10650, 2001. doi: 10.1063/1.1373433.

[32] Tanner Culpitt, Kurt R. Brorsen, and Sharon Hammes-Schiffer. Communication: Density functional theory embedding with the orthogonality constrained basis set expansion procedure. *J. Chem. Phys.*, 146(21):211101, 2017. doi: 10.1063/1.4984777.

[33] Stefan Dapprich, István Komáromi, K.Suzie Byun, Keiji Morokuma,

and Michael J Frisch. A new oniom implementation in gaussian98. part i. the calculation of energies, gradients, vibrational frequencies and electric field derivatives1dedicated to professor keiji morokuma in celebration of his 65th birthday.1. *Comput. Theor. Chem.*, 461-462:1–21, 1999. ISSN 0166-1280. doi: https://doi.org/10.1016/S0166-1280(98)00475-8.

[34] Dipayan Datta, Liguo Kong, and Marcel Nooijen. A state-specific partially internally contracted multireference coupled cluster approach. *J. Chem. Phys.*, 134(21):214116, 2011.

[35] Wallace D. Derricotte and Francesco A. Evangelista. Localized intrinsic valence virtual orbitals as a tool for the automatic classification of core excited states. *J. Chem. Theory Comput.*, 13(12):5984–5999, 2017. doi: 10.1021/acs.jctc.7b00493.

[36] J Emiliano Deustua, Jun Shen, and Piotr Piecuch. Converging High-Level Coupled-Cluster Energetics by Monte Carlo Sampling and Moment Expansions. *Phys. Rev. Lett.*, 119(22):223003, November 2017.

[37] Jessalyn A DeVine, Marissa L Weichman, Benjamin Laws, Jing Chang, Mark C Babin, Garikoitz Balerdi, Changjian Xie, Christopher L Malbon, W Carl Lineberger, David R Yarkony, et al. Encod-

ing of vinylidene isomerization in its anion photoelectron spectrum. Science, 358(6361):336–339, 2017.

[38] Diksha Dhawan, Mekena Metcalf, and Dominika Zgid. Dynamical self-energy mapping (dsem) for quantum computing. arXiv preprint arXiv:2010.05441, 2020.

[39] Feizhi Ding, Frederick R. Manby, and Thomas F. Miller. Embedded mean-field theory with block-orthogonalized partitioning. *J. Chem. Theory Comput.*, 13(4):1605–1615, 2017. doi: 10.1021/acs.jctc.6b01065.

[40] Andreas Dreuw and Martin Head-Gordon. Single-reference ab initio methods for the calculation of excited states of large molecules. *Chem. Rev.*, 105(11):4009–4037, 2005.

[41] Achintya Kumar Dutta, Marcel Nooijen, Frank Neese, and Róbert Izsák. Automatic active space selection for the similarity transformed equations of motion coupled cluster method. *J. Chem. Phys.*, 146(7):074103, 2017. doi: 10.1063/1.4976130.

[42] Francesco A. Evangelista. Adaptive multiconfigurational wave functions. *J. Chem. Phys.*, 140(12):124114, 2014. doi: 10.1063/1.4869192.

[43] Francesco A. Evangelista. A driven similarity renormalization group

approach to quantum many-body problems. *J. Chem. Phys.*, 141(5): 054109, 2014. doi: 10.1063/1.4890660.

[44] Francesco A. Evangelista. Perspective: Multireference coupled cluster theories of dynamical electron correlation. *J. Chem. Phys.*, 149 (3):030901, 2018. doi: 10.1063/1.5039496.

[45] Francesco A. Evangelista and Jürgen Gauss. An orbital-invariant internally contracted multireference coupled cluster approach. *J. Chem. Phys.*, 134(11):114102, 2011. doi: 10.1063/1.3559149.

[46] Francesco A Evangelista and Jürgen Gauss. On the approximation of the similarity-transformed hamiltonian in single-reference and multireference coupled cluster theory. *Chem. Phys.*, 401:27–35, 2012.

[47] Francesco A Evangelista, Andrew C Simmonett, Wesley D Allen, Henry F Schaefer III, and Jürgen Gauss. Triple excitations in state-specific multireference coupled cluster theory: Application of mk-mrccsdt and mk-mrccsdt-n methods to model systems. *J. Chem. Phys.*, 128(12):124104, 2008.

[48] Francesco A Evangelista, Chenyang Li, Kevin P Hannon, Jeffrey B Schriber, Tianyuan Zhang, Chenxi Cai, Nan He, Nicholas Stair, Shuhe Wang, and Renke Huang. Forte: an open-source suite of quantum chemistry methods for strongly correlated electrons, 2022.

[49] Edoardo Fertitta and George H Booth. Rigorous wave function embedding with dynamical fluctuations. *Phys. Rev. B*, 98(23):235132, 2018.

[50] Edoardo Fertitta and George H Booth. Energy-weighted density matrix embedding of open correlated chemical fragments. *J. Chem. Phys.*, 151(1):014115, July 2019.

[51] Edoardo Fertitta and George H Booth. Energy-weighted density matrix embedding of open correlated chemical fragments. *J. Chem. Phys.*, 151(1):014115, 2019.

[52] James Finley, Per-Åke Malmqvist, Björn O. Roos, and Luis Serrano-Andrés. The multi-state caspt2 method. *Chem. Phys. Lett.*, 288(2): 299 – 306, 1998. ISSN 0009-2614. doi: https://doi.org/10.1016/S0009-2614(98)00252-8.

[53] Vladimir Fock. Näherungsmethode zur lösung des quantenmechanischen mehrkörperproblems. *Zeitschrift für Physik*, 61(1):126–148, 1930.

[54] Mark E. Fornace, Joonho Lee, Kaito Miyamoto, Frederick R. Manby, and Thomas F. Miller. Embedded mean-field theory. *J. Chem. Theory Comput.*, 11(2):568–580, 2015. doi: 10.1021/ct5011032.

[55] Karl F Freed. Theoretical foundations of purely semiempirical quantum chemistry. *J. Chem. Phys.*, 60(5):1765–1788, 1974.

[56] Emmanuel Fromager, Stefan Knecht, and Hans Jørgen Aa Jensen. Multi-configuration time-dependent density-functional theory based on range separation. *J. Chem. Phys.*, 138(8):084101, February 2013.

[57] Carlo Alberto Gaggioli, Samuel J. Stoneburner, Christopher J. Cramer, and Laura Gagliardi. Beyond density functional theory: The multiconfigurational approach to model heterogeneous catalysis. *ACS Catal.*, 9(9):8481–8502, 2019. doi: 10.1021/acscatal.9b01775.

[58] Antoine Georges, Gabriel Kotliar, Werner Krauth, and Marcelo J. Rozenberg. Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions. *Rev. Mod. Phys.*, 68:13–125, Jan 1996. doi: 10.1103/RevModPhys.68.13.

[59] Giovanni Ghigo, Björn O. Roos, and Per-Åke Malmqvist. A modified definition of the zeroth-order hamiltonian in multiconfigurational perturbation theory (caspt2). *Chem. Phys. Lett.*, 396(1):142 – 149, 2004. ISSN 0009-2614. doi: https://doi.org/10.1016/j.cplett.2004.08.032.

[60] Soumen Ghosh, Pragya Verma, Christopher J. Cramer, Laura Gagliardi, and Donald G. Truhlar. Combining wave function methods with density functional theory for excited states. *Chem. Rev.*, 118(15):7249–7292, 2018. doi: 10.1021/acs.chemrev.8b00193.

[61] M. J. Gillan, D. Alfè, P. J. Bygrave, C. R. Taylor, and F. R. Manby. Energy benchmarks for water clusters and ice structures from an embedded many-body expansion. *J. Chem. Phys.*, 139(11):114101, 2013. doi: 10.1063/1.4820906.

[62] Jérôme F Gonthier and C David Sherrill. Density-fitted open-shell symmetry-adapted perturbation theory and application to π-stacking in benzene dimer cation and ionized dna base pair steps. *J. Chem. Phys.*, 145(13):134106, 2016.

[63] Jason D. Goodpaster, Taylor A. Barnes, Frederick R. Manby, and Thomas F. Miller. Accurate and systematically improvable density functional theory embedding for correlated wavefunctions. *J. Chem. Phys.*, 140(18):18A507, 2014. doi: 10.1063/1.4864040.

[64] Niranjan Govind, Yan Alexander Wang, and Emily A. Carter. Electronic-structure calculations by first-principles density-based embedding of explicitly correlated systems. *J. Chem. Phys.*, 110(16): 7677–7688, 1999. doi: 10.1063/1.478679.

[65] Giovanni Granucci and Maurizio Persico. Benzene—o2 interaction

potential from ab initio calculations. *Chem. Phys. Lett.*, 205(2-3): 331–336, 1993.

[66] Stefan Grimme. Semiempirical gga-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.*, 27(15):1787–1799, 2006.

[67] Matthias Hanauer and Andreas Köhn. Pilot applications of internally contracted multireference coupled cluster theory, and how to choose the cluster operator properly. *J. Chem. Phys.*, 134(20): 204111, 2011. doi: 10.1063/1.3592786.

[68] Kevin P. Hannon, Chenyang Li, and Francesco A. Evangelista. An integral-factorized implementation of the driven similarity renormalization group second-order multireference perturbation theory. *J. Chem. Phys.*, 144(20):204111, 2016. doi: 10.1063/1.4951684.

[69] Douglas R Hartree. The wave mechanics of an atom with a non-coulomb central field. part i. theory and methods. In *Math. Proc. Camb. Philos. Soc.*, volume 24, pages 89–110. Cambridge university press, 1928.

[70] Christof Hättig. Optimization of auxiliary basis sets for ri-mp2 and ri-cc2 calculations: Core–valence and quintuple-ζ basis sets for h to ar and qzvpp basis sets for li to kr. *Phys. Chem. Chem. Phys.*, 7:59–

66, 2005. doi: 10.1039/B415208E. URL `http://dx.doi.org/10.1039/B415208E`.

[71] Nan He and Francesco A. Evangelista. A zeroth-order active-space frozen-orbital embedding scheme for multireference calculations. *J. Chem. Phys.*, 152(9):094107, 2020. doi: 10.1063/1.5142481.

[72] Nan He, Chenyang Li, and Francesco A Evangelista. Second-order active-space embedding theory. *J. Chem. Theory Comput.*, 2022.

[73] Bence Hégely, Péter R. Nagy, György G. Ferenczy, and Mihály Kállay. Exact density functional and wave function embedding schemes based on orbital localization. *J. Chem. Phys.*, 145(6):064107, 2016. doi: 10.1063/1.4960177.

[74] Bence Hégely, Péter R. Nagy, and Mihály Kállay. Dual basis set approach for density functional and wave function embedding schemes. *J. Chem. Theory Comput.*, 14(9):4600–4615, 2018. doi: 10.1021/acs.jctc.8b00350.

[75] J Heidberg, E Kampshoff, and M Suhren. Correlation field, structure, and phase transition in the monolayer co adsorbed on nacl (100) as revealed from polarization fourier-transform infrared spectroscopy. *J. Chem. Phys.*, 95(12):9408–9411, 1991.

[76] Matthew R. Hermes and Laura Gagliardi. Multiconfigurational self-

consistent field theory with density matrix embedding: The localized active space self-consistent field method. J. Chem. Theory Comput., 15(2):972–986, 2019. doi: 10.1021/acs.jctc.8b01009.

[77] Edward G Hohenstein and C David Sherrill. Density fitting and cholesky decomposition approximations in symmetry-adapted perturbation theory: Implementation and application to probe the nature of π-π interactions in linear acenes. J. Chem. Phys., 132(18): 184111, 2010.

[78] Chen Huang and Emily A. Carter. Potential-functional embedding theory for molecules and materials. J. Chem. Phys., 135(19): 194104, 2011. doi: 10.1063/1.3659293.

[79] Chen Huang, Michele Pavone, and Emily A. Carter. Quantum mechanical embedding theory based on a unique embedding potential. J. Chem. Phys., 134(15):154110, 2011. doi: 10.1063/1.3577516.

[80] Marcella Iannuzzi, Barbara Kirchner, and Jürg Hutter. Density functional embedding for molecular systems. Chem. phys. lett., 421(1-3):16–20, 2006.

[81] Miho Isegawa, Bo Wang, and Donald G. Truhlar. Electrostatically embedded molecular tailoring approach and validation for peptides. J. Chem. Theory Comput., 9(3):1381–1393, 2013. doi: 10.1021/ct300845q.

[82] Hongyan Jiang, Marvin Kammler, Feizhi Ding, Yvonne Dorenkamp, Frederick R. Manby, Alec. M. Wodtke, Thomas F. Miller, Alexander Kandratsenka, and Oliver Bünermann. Imaging covalent bond formation by h atom scattering from graphene. *Science*, 364(6438): 379–382, 2019. doi: 10.1126/science.aaw6378.

[83] Russell D Johnson III. Nist computational chemistry comparison and benchmark database, nist standard reference database number 101. Release 16a http://cccbdb. nist. gov/(accessed July 11, 2021), 2013.

[84] Richard R. Jones and Robert G. Bergman. p-benzyne. generation as an intermediate in a thermal isomerization reaction and trapping evidence for the 1,4-benzenediyl structure. *J. Am. Chem. Soc.*, 94 (2):660–661, 1972. doi: 10.1021/ja00757a071.

[85] Mihály Kállay, Péter G. Szalay, and Péter R. Surján. A general state-selective multireference coupled-cluster algorithm. *J. Chem. Phys.*, 117(3):980–990, 2002. doi: 10.1063/1.1483856.

[86] Stefan K. Kehrein and Andreas Mielke. Diagonalization of System plus Environment Hamiltonians. *J. Stat. Phys.*, 90(3-4):889–898, 1998. ISSN 0022-4715. doi: 10.1023/a:1023289323069.

[87] T Kinoshita, O Hino, and Rodney J Bartlett. Coupled-cluster

method tailored by configuration interaction. J. Chem. Phys., 123 (7):074106, August 2005.

[88] Thorsten Klüner, Niranjan Govind, Yan Alexander Wang, and Emily A. Carter. Prediction of electronic excited states of adsorbates on metal surfaces from first principles. Phys. Rev. Lett., 86: 5954–5957, Jun 2001. doi: 10.1103/PhysRevLett.86.5954.

[89] Thorsten Klüner, Niranjan Govind, Yan Alexander Wang, and Emily A. Carter. Periodic density functional embedding theory for complete active space self-consistent field and configuration interaction calculations: Ground and excited states. J. Chem. Phys., 116 (1):42–54, 2002. doi: 10.1063/1.1420748.

[90] Gerald Knizia. Intrinsic atomic orbitals: An unbiased bridge between quantum theory and chemical concepts. J. Chem. Theory Comput., 9 (11):4834–4843, 2013. doi: 10.1021/ct400687b.

[91] Gerald Knizia and Garnet Kin-Lic Chan. Density matrix embedding: A simple alternative to dynamical mean-field theory. Phys. Rev. Lett., 109:186404, Nov 2012. doi: 10.1103/PhysRevLett.109.186404.

[92] Gerald Knizia and Garnet Kin-Lic Chan. Density matrix embedding: A strong-coupling quantum embedding theory. J. Chem. Theory Comput., 9(3):1428–1432, 2013. doi: 10.1021/ct301044e.

[93] Andreas Köhn, Matthias Hanauer, Leonie Anna Mück, Thomas-Christian Jagau, and Jürgen Gauss. State-specific multireference coupled-cluster theory. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 3 (2):176–197, March 2013.

[94] G. Kotliar, S. Y. Savrasov, K. Haule, V. S. Oudovenko, O. Parcollet, and C. A. Marianetti. Electronic structure calculations with dynamical mean-field theory. *Rev. Mod. Phys.*, 78:865–951, Aug 2006. doi: 10.1103/RevModPhys.78.865.

[95] Alisa Krishtal, Debalina Sinha, Alessandro Genova, and Michele Pavanello. Subsystem density-functional theory as an effective tool for modeling ground and excited states, their dynamics and many-body interactions. J. Phys. Condens. Matter, 27(18):183202, 2015. ISSN 0953-8984. doi: 10.1088/0953-8984/27/18/183202.

[96] Joseph B Kruskal. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proc. Am. Math. Soc.*, 7(1):48–50, 1956.

[97] Anna I Krylov. Equation-of-Motion Coupled-Cluster Methods for Open-Shell and Electronically Excited Species: The Hitchhiker's Guide to Fock Space. *Annu. Rev. Phys. Chem.*, 59:433–462, 2008.

[98] Ambarish Kulkarni, Samira Siahrostami, Anjli Patel, and Jens K. Nørskov. Understanding catalytic activity trends in the oxygen

reduction reaction. *Chem. Rev.*, 118(5):2302–2312, 2018. doi: 10.1021/acs.chemrev.7b00488.

[99] Werner Kutzelnigg and Debashis Mukherjee. Normal order and extended wick theorem for a multiconfiguration reference wave function. *J. Chem. Phys.*, 107(2):432–449, 1997.

[100] M. Ladd and M.F.C. Ladd. Crystal Structures: Lattices and Solids in Stereoview. Horwood Series in Science. Elsevier Science, 1999. ISBN 9781898563631. URL `https://books.google.com/books?id=pCtguVnfgj0C`.

[101] Tran Nguyen Lan, Alexei A. Kananenka, and Dominika Zgid. Communication: Towards ab initio self-energy embedding theory in quantum chemistry. *J. Chem. Phys.*, 143(24):241102, 2015. doi: 10.1063/1.4938562.

[102] Bryan T. G. Lau, Gerald Knizia, and Timothy C. Berkelbach. Regional embedding enables high-level quantum chemistry for surface science. J. Phys. Chem. Lett., 12(3):1104–1109, 2021. doi: 10.1021/acs.jpclett.0c03274.

[103] Jascha A Lau, Arnab Choudhury, Li Chen, Dirk Schwarzer, Varun B Verma, and Alec M Wodtke. Observation of an isomerizing double-well quantum system in the condensed phase. *Science*, 367(6474): 175–178, 2020.

[104] C. Lee, W. Yang, and R. G. Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, 37:785–789, January 1988. doi: 10.1103/PhysRevB.37.785.

[105] Sebastian J. R. Lee, Matthew Welborn, Frederick R. Manby, and Thomas F. Miller. Projection-based wavefunction-in-dft embedding. *Acc. Chem. Res.*, 52(5):1359–1368, 2019. doi: 10.1021/acs.accounts.8b00672.

[106] Sebastian JR Lee, Feizhi Ding, Frederick R Manby, and Thomas F Miller III. Analytical gradients for projection-based wavefunction-in-dft embedding. *J. Chem. Phys.*, 151(6):064112, 2019.

[107] Chenyang Li and Francesco A. Evangelista. Multireference driven similarity renormalization group: A second-order perturbative analysis. *J. Chem. Theory Comput.*, 11(5):2097–2108, 2015. doi: 10.1021/acs.jctc.5b00134.

[108] Chenyang Li and Francesco A. Evangelista. Towards numerically robust multireference theories: The driven similarity renormalization group truncated to one- and two-body operators. *J. Chem. Phys.*, 144(16):164114, 2016. doi: 10.1063/1.4947218.

[109] Chenyang Li and Francesco A. Evangelista. Driven similarity renor-

malization group: Third-order multireference perturbation theory. J. Chem. Phys., 146(12):124132, 2017. doi: 10.1063/1.4979016.

[110] Chenyang Li and Francesco A. Evangelista. Erratum: “driven similarity renormalization group: Third-order multireference perturbation theory” [j. chem. phys. 146, 124132 (2017)]. J. Chem. Phys., 148 (7):079902, 2018. doi: 10.1063/1.5023904.

[111] Chenyang Li and Francesco A. Evangelista. Driven similarity renormalization group for excited states: A state-averaged perturbation theory. J. Chem. Phys., 148(12):124106, 2018. doi: 10.1063/1.5019793.

[112] Chenyang Li and Francesco A. Evangelista. Multireference theories of electron correlation based on the driven similarity renormalization group. Annu. Rev. Phys. Chem, 70(1):245–273, 2019.

[113] Giovanni Li Manni, Rebecca K Carlson, Sijie Luo, Dongxia Ma, Jeppe Olsen, Donald G Truhlar, and Laura Gagliardi. Multiconfiguration Pair-Density Functional Theory. J. Chem. Theory Comput., 10(9): 3669–3680, September 2014.

[114] Florian Libisch, Chen Huang, and Emily A. Carter. Embedded correlated wavefunction schemes: Theory and applications. Acc. Chem. Res., 47(9):2768–2775, 2014. doi: 10.1021/ar500086h.

[115] Cameron F. Logan, Jennifer C. Ma, and Peter Chen. The photoelectron spectrum of the .alpha.,3-dehydrotoluene biradical. *J. Am. Chem. Soc.*, 116(5):2137–2138, 1994. doi: 10.1021/ja00084a061.

[116] W. C. Lu, C. Z. Wang, M. W. Schmidt, L. Bytautas, K. M. Ho, and K. Ruedenberg. 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.*, 120(6):2629–2637, 2004. doi: 10.1063/1.1638731.

[117] Dmitry I. Lyakh, Monika Musiał, Victor F. Lotrich, and Rodney J. Bartlett. Multireference nature of chemistry: The coupled-cluster view. *Chem. Rev.*, 112(1):182–243, 2012. doi: 10.1021/cr2001417.

[118] Dmitry I Lyakh, Monika Musiał, Victor F Lotrich, and Rodney J Bartlett. Multireference nature of chemistry: The coupled-cluster view. *Chem. Rev.*, 112(1):182–243, 2012.

[119] He Ma, Nan Sheng, Marco Govoni, and Giulia Galli. Quantum embedding theory for strongly correlated states in materials. *J. Chem. Theory Comput.*, 17(4):2116–2125, 2021.

[120] Giovanni Macetti and Alessandro Genoni. Quantum mechanics/extremely localized molecular orbital embedding strategy for excited states: Coupling to time-dependent density functional theory

and equation-of-motion coupled cluster. J. Chem. Theory Comput., 16(12):7490–7506, 2020.

[121] Uttam Sinha Mahapatra, Barnali Datta, and Debashis Mukherjee. A size-consistent state-specific multireference coupled cluster theory: Formal developments and molecular applications. J. Chem. Phys., 110(13):6171–6188, 1999. doi: 10.1063/1.478523.

[122] Frederick R. Manby, Martina Stella, Jason D. Goodpaster, and Thomas F. Miller. A simple, exact density-functional-theory embedding scheme. J. Chem. Theory Comput., 8(8):2564–2568, 2012. doi: 10.1021/ct300544e.

[123] David A Mazziotti. Anti-hermitian contracted schrödinger equation: Direct determination of the two-electron reduced density matrices of many-electron molecules. Phys. Rev. Lett., 97(14):143002, 2006.

[124] David A. Mazziotti. Two-electron reduced density matrix as the basic variable in many-electron quantum chemistry and physics. Chem. Rev., 112(1):244–262, 2012. doi: 10.1021/cr2000493.

[125] Andrew W Meredith and Anthony J Stone. A perturbation theory study of adlayer co on nacl(100). J. Chem. Phys., 104(8): 3058–3070, 1996. doi: 10.1063/1.471072.

[126] Mekena Metcalf, Nicholas P Bauman, Karol Kowalski, and Wibe A

De Jong. Resource-efficient chemistry on quantum computers with the variational quantum eigensolver and the double unitary coupled-cluster approach. *J. Chem. Theory Comput.*, 16(10):6165–6175, 2020.

[127] Rana K Mohamed, Paul W. Peterson, and Igor V. Alabugin. Concerted reactions that produce diradicals and zwitterions: Electronic, steric, conformational, and kinetic control of cycloaromatization processes. *Chem. Rev.*, 113(9):7089–7129, 2013. doi: 10.1021/cr4000682.

[128] Chr Møller and Milton S Plesset. Note on an approximation treatment for many-electron systems. *Phys. Rev.*, 46(7):618, 1934.

[129] Apurba Nandi, Peng Zhang, Jun Chen, Hua Guo, and Joel M Bowman. Quasiclassical simulations based on cluster models reveal vibration-facilitated roaming in the isomerization of co adsorbed on nacl. *Nat. Chem.*, 13(3):249–254, 2021.

[130] Eric Neuscamman, Takeshi Yanai, and Garnet Kin-Lic Chan. A review of canonical transformation theory. *Int. Rev. Phys. Chem.*, 29 (2):231–271, April 2010.

[131] Eric Neuscamman, Takeshi Yanai, and Garnet Kin-Lic Chan. A review of canonical transformation theory. *Int. Rev. Phys. Chem.*, 29 (2):231–271, 2010.

[132] Tran Nguyen Lan, Alexei A Kananenka, and Dominika Zgid. Rigorous ab initio quantum embedding for quantum chemistry using green's function theory: Screened interaction, nonlocal self-energy relaxation, orbital basis, and chemical accuracy. *J. Chem. Theory Comput.*, 12(10):4856–4870, 2016.

[133] Samuel O. Odoh, Christopher J. Cramer, Donald G. Truhlar, and Laura Gagliardi. Quantum-chemical characterization of the properties and reactivities of metal–organic frameworks. *Chem. Rev.*, 115 (12):6051–6111, 2015. doi: 10.1021/cr500551h.

[134] Nevin Oliphant and Ludwik Adamowicz. Multireference coupled-cluster method using a single-reference formalism. *J. Chem. Phys.*, 94(2):1229–1235, 1991. doi: 10.1063/1.460031.

[135] Jógvan Magnus Olsen, Kęstutis Aidas, and Jacob Kongsted. Excited states in solution through polarizable embedding. *J. Chem. Theory Comput.*, 6(12):3721–3734, 2010. doi: 10.1021/ct1003803.

[136] Morten N. Pedersen, Erik D. Hedegård, Jógvan Magnus H. Olsen, Joanna Kauczor, Patrick Norman, and Jacob Kongsted. Damped response theory in combination with polarizable environments: The polarizable embedding complex polarization propagator method.

J. Chem. Theory Comput., 10(3):1164–1171, 2014. doi: 10.1021/ct400946k.

[137] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. J. Mach. Learn Res., 12:2825–2830, 2011.

[138] John P Perdew. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. Phys. Rev. B, 33(12): 8822, 1986.

[139] Hung Q. Pham, Varinia Bernales, and Laura Gagliardi. Can density matrix embedding theory with the complete activate space self-consistent field solver describe single and double bond breaking in molecular systems? J. Chem. Theory Comput., 14(4):1960–1968, 2018. doi: 10.1021/acs.jctc.7b01248.

[140] Piotr Piecuch and Ludwik Adamowicz. State-selective multireference coupled-cluster theory employing the single-reference formalism: implementation and application to the H_8 model system. J. Chem. Phys., 100(8):5792–5809, 1994.

[141] Peter Pulay. Localizability of dynamic electron correlation.

Chem. Phys. Lett., 100(2):151–154, 1983. doi: 10.1016/0009-2614(83)80703-9.

[142] Haisheng Ren, Makenzie R Provorse, Peng Bao, Zexing Qu, and Jiali Gao. Multistate Density Functional Theory for Effective Diabatic Electronic Coupling. *J. Phys. Chem. Lett.*, 7(12):2286–2293, June 2016.

[143] Ryan M. Richard and John M. Herbert. A generalized many-body expansion and a unified view of fragment-based methods in electronic structure theory. *J. Chem. Phys.*, 137(6):064113, 2012. doi: 10.1063/1.4742816.

[144] Björn O. Roos, Peter R. Taylor, and Per E.M. Siegbahn. A complete active space scf method (casscf) using a density matrix formulated super-ci approach. *Chem. Phys.*, 48(2):157 – 173, 1980. ISSN 0301-0104. doi: https://doi.org/10.1016/0301-0104(80)80045-0.

[145] Clemens Carel Johannes Roothaan. New developments in molecular orbital theory. *Rev. Mod. Phys.*, 23(2):69, 1951.

[146] Klaus Ruedenberg, Michael W. Schmidt, Mary M. Gilbert, and S.T. Elbert. Are atoms intrinsic to molecular electronic wavefunctions? i. the fors model. *Chem. Phys.*, 71(1):41 – 49, 1982. ISSN 0301-0104. doi: https://doi.org/10.1016/0301-0104(82)87004-3.

[147] Elvira R. Sayfutyarova and Sharon Hammes-Schiffer. Constructing molecular π-orbital active spaces for multireference calculations of conjugated systems. *J. Chem. Theory Comput.*, 15(3):1679–1689, 2019. doi: 10.1021/acs.jctc.8b01196.

[148] Elvira R. Sayfutyarova, Qiming Sun, Garnet Kin-Lic Chan, and Gerald Knizia. Automated construction of molecular active spaces from atomic valence orbitals. *J. Chem. Theory Comput.*, 13(9):4063–4078, 2017. doi: 10.1021/acs.jctc.7b00128.

[149] Elvira R Sayfutyarova, Qiming Sun, Garnet Kin-Lic Chan, and Gerald Knizia. Automated construction of molecular active spaces from atomic valence orbitals. *J. Chem. Theory Comput.*, 13(9):4063–4078, 2017.

[150] Bruno Schuler, Shadi Fatayer, Fabian Mohn, Nikolaj Moll, Niko Pavliček, Gerhard Meyer, Diego Pe na, and Leo Gross. Reversible bergman cyclization by atomic manipulation. *Nat, Chem.*, 8(1):220–224, 2016. doi: 10.1038/nchem.2438.

[151] Tobias Schwabe, Jógvan Magnus Haugaard Olsen, Kristian Sneskov, Jacob Kongsted, and Ove Christiansen. Solvation effects on electronic transitions: Exploring the performance of advanced solvent potentials in polarizable embedding calculations.

J. Chem. Theory Comput., 7(7):2209–2217, 2011. doi: 10.1021/ct200258g.

[152] Charles J. C. Scott and George H. Booth. Extending density matrix embedding: A static two-particle theory, 2021.

[153] Lu Jeu Sham and Walter Kohn. One-particle properties of an inhomogeneous interacting electron gas. *Phys. Rev.*, 145(2):561, 1966.

[154] Isaiah Shavitt and Rodney J Bartlett. Many-body methods in chemistry and physics: MBPT and coupled-cluster theory. Cambridge university press, 2009.

[155] Jun Shen and Piotr Piecuch. Merging Active-Space and Renormalized Coupled-Cluster Methods via the CC(P;Q) Formalism, with Benchmark Calculations for Singlet-Triplet Gaps in Biradical Systems. *J. Chem. Theory Comput.*, 8(12):4968–4988, December 2012.

[156] C David Sherrill and Henry F Schaefer III. The configuration interaction method: Advances in highly correlated approaches. In *Adv. Quantum Chem.*, volume 34, pages 143–269. Elsevier, 1999.

[157] Shreya Sinha and Peter Saalfrank. “inverted” co molecules on nacl (100): a quantum mechanical study. *Phys. Chem. Chem. Phys.*, 23 (13):7860–7874, 2021.

[158] John C Slater. The theory of complex spectra. *Phys. Rev.*, 34(10): 1293, 1929.

[159] Daniel G. A. Smith, Lori A. Burns, Andrew C. Simmonett, Robert M. Parrish, Matthew C. Schieber, Raimondas Galvelis, Peter Kraus, Holger Kruse, Roberto Di Remigio, Asem Alenaizan, Andrew M. James, Susi Lehtola, Jonathon P. Misiewicz, Maximilian Scheurer, Robert A. Shaw, Jeffrey B. Schriber, Yi Xie, Zachary L. Glick, Dominic A. Sirianni, Joseph Senan O'Brien, Jonathan M. Waldrop, Ashutosh Kumar, Edward G. Hohenstein, Benjamin P. Pritchard, Bernard R. Brooks, Henry F. Schaefer, Alexander Yu. Sokolov, Konrad Patkowski, A. Eugene DePrince, Uğur Bozkaya, Rollin A. King, Francesco A. Evangelista, Justin M. Turney, T. Daniel Crawford, and C. David Sherrill. Psi4 1.4: Open-source software for high-throughput quantum chemistry. *J. Chem. Phys.*, 152(18):184108, 2020. doi: 10.1063/5.0006002.

[160] Kristian Sneskov, Tobias Schwabe, Jacob Kongsted, and Ove Christiansen. The polarizable embedding coupled cluster method. *J. Chem. Phys.*, 134(10):104108, 2011. doi: 10.1063/1.3560034.

[161] Christopher J. Stein and Markus Reiher. Automated selection of active orbital spaces. *J. Chem. Theory Comput.*, 12(4):1760–1771, 2016. doi: 10.1021/acs.jctc.6b00156.

[162] Tamar Stein, Thomas M. Henderson, and Gustavo E. Scuseria. Seniority zero pair coupled cluster doubles theory. *J. Chem. Phys.*, 140 (21):214113, 2014. doi: 10.1063/1.4880819.

[163] Philip J Stephens, Frank J Devlin, Cary F Chabalowski, and Michael J Frisch. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.*, 98(45):11623–11627, 1994.

[164] Jianwei Sun, Adrienn Ruzsinszky, and John P Perdew. Strongly constrained and appropriately normed semilocal density functional. *Phys. Rev. Lett.*, 115(3):036402, 2015.

[165] Qiming Sun and Garnet Kin-Lic Chan. Quantum embedding theories. *Acc. Chem. Res.*, 49(12):2705–2712, 2016. doi: 10.1021/acs.accounts.6b00356.

[166] Mats Svensson, Stéphane Humbel, Robert D. J. Froese, Toshiaki Matsubara, Stefan Sieber, and Keiji Morokuma. Oniom: A multilayered integrated mo + mm method for geometry optimizations and single point energy predictions. a test for diels-alder reactions and pt(p(t-bu)3)2 + h2 oxidative addition. *J. Phys. Chem.*, 100(50):19357–19363, 1996. doi: 10.1021/jp962071j.

[167] Péter G. Szalay and Rodney J. Bartlett. Multi-reference averaged quadratic coupled-cluster method: a size-extensive modification of

multi-reference ci. *Chem. Phys. Lett.*, 214(5):481–488, 1993. doi: 10.1016/0009-2614(93)85670-J.

[168] Péter G. Szalay and Rodney J. Bartlett. Approximately extensive modifications of the multireference configuration interaction method: A theoretical and practical analysis. *J. Chem. Phys.*, 103 (9):3600–3612, 1995. doi: 10.1063/1.470243.

[169] Péter G. Szalay, Thomas Müller, Gergely Gidofalvi, Hans Lischka, and Ron Shepard. Multiconfiguration self-consistent field and multireference configuration interaction methods and applications. *Chem. Rev.*, 112(1):108–181, 2012. doi: 10.1021/cr200137a.

[170] Andrew G Taube and Rodney J Bartlett. Improving upon CCSD(T): Lambda CCSD(T). I. Potential energy surfaces. *J. Chem. Phys.*, 128 (4):044110, 2008.

[171] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa,

Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2.

[172] N-T Vu, A Jakalian, and DB Jack. A simulation of the phase transition in monolayer co/nacl (001). *J. Chem. Phys.*, 106(6):2551–2554, 1997.

[173] Shuhe Wang, Chenyang Li, and Francesco A Evangelista. Analytic energy gradients for the driven similarity renormalization group multireference second-order perturbation theory. *J. Chem. Theory Comput.*, 17(12):7666–7681, 2021.

[174] Thomas J Watson and Garnet Kin-Lic Chan. Correct Quantum Chemistry in a Minimal Basis from Effective Hamiltonians. *J. Chem. Theory Comput.*, 12(2):512–22, 2016. ISSN 1549-9618. doi: 10.1021/acs.jctc.5b00138.

[175] Franz Wegner. Flow-equations for hamiltonians. Ann. Phys., 506 (2):77–91, 1994.

[176] Florian Weigend. A fully direct ri-hf algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency. *Phys. Chem. Chem. Phys.*, 4:4285–4291, 2002. doi: 10.1039/B204199P.

[177] Florian Weigend and Reinhart Ahlrichs. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for h to rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.*, 7:3297–3305, 2005. doi: 10.1039/B508541A.

[178] Florian Weigend, Andreas Köhn, and Christof Hättig. Efficient use of the correlation consistent basis sets in resolution of the identity mp2 calculations. *J. Chem. Phys.*, 116(8):3175–3183, 2002.

[179] Matthew Welborn, Takashi Tsuchimochi, and Troy Van Voorhis. Bootstrap embedding: An internally consistent fragment-based method. *J. Chem. Phys.*, 145(7):074102, 2016. doi: 10.1063/1.4960986.

[180] Matthew Welborn, Frederick R. Manby, and Thomas F. Miller. Even-handed subsystem selection in projection-based embedding. *J. Chem. Phys.*, 149(14):144101, 2018. doi: 10.1063/1.5050533.

[181] Hans Henning Wenk, Michael Winkler, and Wolfram Sander. One century of aryne chemistry. *Angew. Chem. Int. Ed.*, 42(5):502–528, 2003. doi: 10.1002/anie.200390151.

[182] Paul G. Wenthold, Robert R. Squires, and W. C. Lineberger. Ultraviolet photoelectron spectroscopy of the o-, m-, and p-benzyne negative ions. electron affinities and singlet–triplet splittings for o-, m-,

and p-benzyne. *J. Am. Chem. Soc.*, 120(21):5279–5290, 1998. doi: 10.1021/ja9803355.

[183] Hans-Joachim Werner and Peter J. Knowles. A second order multiconfiguration scf procedure with optimum convergence. *J. Chem. Phys.*, 82(11):5053–5063, 1985. doi: 10.1063/1.448627.

[184] Tomasz A. Wesolowski, Sapana Shedge, and Xiuwen Zhou. Frozen-density embedding strategy for multilevel simulations of electronic structure. *Chem. Rev.*, 115(12):5891–5928, 2015. doi: 10.1021/cr500502v.

[185] Tomasz Adam Wesolowski and Arieh Warshel. Frozen density functional approach for ab initio calculations of solvated molecules. *J. Phys. Chem.*, 97(30):8050–8053, Jul 1993. doi: 10.1021/j100132a040. URL `https://pubs.acs.org/doi/pdf/10.1021/j100132a040`.

[186] Tomasz Adam Wesolowski, Olivier Parisel, Yves Ellinger, and Jacques Weber. Comparative study of benzene$\cdots$x (x= o2, n2, co) complexes using density functional theory: The importance of an accurate exchange- correlation energy density at high reduced density gradients. *J. Phys. Chem. A*, 101(42):7818–7825, 1997.

[187] Paul Westhaus. Cluster expansions of canonically transformed

hamiltonians and effective interactions. *Int. J. Quantum Chem.*, 7 (S7):463–477, 1973.

[188] Henryk A Witek, Yoong-Kee Choe, James P Finley, and Kimihiko Hirao. Intruder state avoidance multireference møller–plesset perturbation theory. *J. Comput. Chem.*, 23(10):957–965, 2002.

[189] David E. Woon and Thom H. Dunning. Gaussian basis sets for use in correlated molecular calculations. v. core-valence basis sets for boron through neon. *J. Chem. Phys.*, 103(11):4572–4585, 1995. doi: 10.1063/1.470645.

[190] Sebastian Wouters, Carlos A. Jiménez-Hoyos, Qiming Sun, and Garnet K.-L. Chan. A practical guide to density matrix embedding theory in quantum chemistry. *J. Chem. Theory Comput.*, 12(6):2706–2719, 2016. doi: 10.1021/acs.jctc.6b00316.

[191] Takeshi Yanai and Garnet Kin-Lic Chan. Canonical transformation theory for multireference problems. *J. Chem. Phys.*, 124(19): 194106, 2006.

[192] Hong-Zhou Ye, Nathan D Ricke, Henry K Tran, and Troy Van Voorhis. Bootstrap Embedding for Molecules. *J. Chem. Theory Comput.*, 15 (8):4497–4506, June 2019.

[193] Hong-Zhou Ye, Nathan D. Ricke, Henry K. Tran, and

Troy Van Voorhis. Bootstrap embedding for molecules. *J. Chem. Theory Comput.*, 15(8):4497–4506, 2019. doi: 10.1021/acs.jctc.9b00529.

[194] Dominika Zgid and Garnet Kin-Lic Chan. Dynamical mean-field theory from a quantum chemical perspective. *J. Chem. Phys.*, 134(9): 094115, 2011. doi: 10.1063/1.3556707.

[195] Tianyuan Zhang, Chenyang Li, and Francesco A Evangelista. Improving the efficiency of the multireference driven similarity renormalization group via sequential transformation, density fitting, and the noninteracting virtual orbital approximation. *J. Chem. Theory Comput.*, 15(8):4399–4414, 2019.

[196] Qing Zhao, Xing Zhang, John Mark P Martirez, and Emily A Carter. Benchmarking an embedded adaptive sampling configuration interaction method for surface reactions: H2 desorption from and ch4 dissociation on cu (111). *J. Chem. Theory Comput.*, 16(11):7078–7088, 2020.