

# UC Berkeley

## UC Berkeley Previously Published Works

### Title

A static quantum embedding scheme based on coupled cluster theory

### Permalink

<https://escholarship.org/uc/item/1xk7n98w>

### Journal

The Journal of Chemical Physics, 161(16)

### ISSN

0021-9606

### Authors

Shee, Avijit  
Faulstich, Fabian M  
Whaley, K Birgitta  
[et al.](#)

### Publication Date

2024-10-28

### DOI

10.1063/5.0214065

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

# A static quantum embedding scheme based on coupled cluster theory

Avijit Shee,<sup>1</sup> Fabian M. Faulstich,<sup>2</sup> Birgitta Whaley,<sup>1,3</sup> Lin Lin,<sup>4,5</sup> and Martin Head-Gordon<sup>1,6</sup>

<sup>1</sup>*Department of Chemistry, University of California, Berkeley, CA 94720, USA*

<sup>2</sup>*Department of Mathematics, Rensselaer Polytechnic Institute, Troy, NY 12180, USA*

<sup>3</sup>*Berkeley Center for Quantum Information and Computation, Berkeley, CA 94720, USA*

<sup>4</sup>*Department of Mathematics, University of California, Berkeley, CA 94720, USA*

<sup>5</sup>*Applied Mathematics and Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*

<sup>6</sup>*Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*

(\*Electronic mail: ashee@berkeley.edu)

(Dated: 4 October 2024)

We develop a static quantum embedding scheme that utilizes different levels of approximations to coupled cluster (CC) theory for an active fragment region and its environment. To reduce the computational cost, we solve the local fragment problem using a high-level CC method and address the environment problem with a lower-level Møller-Plesset (MP) perturbative method. This embedding approach inherits many conceptual developments from the hybrid MP2 and CC works by Nooijen and Sherrill (J. Chem. Phys. 111, 10815 (1999), J. Chem. Phys. 122, 234110 (2005)). We go beyond those works here by primarily targeting a specific localized fragment of a molecule and also introducing an alternative mechanism to relax the environment within this framework. We will call this approach MP-CC. We demonstrate the effectiveness of MP-CC on several potential energy curves, and a set of thermochemical reaction energies, using CC with singles and doubles as the fragment solver, and MP2-like treatments of the environment. The results are substantially improved by the inclusion of orbital relaxation in the environment. Using localized bonds as the active fragment, we also report results for N=N bond breaking in azomethane and for the central C–C bond torsion in butadiene. We find that when the fragment Hilbert space size remains fixed (e.g., when determined by an intrinsic atomic orbital approach), the method achieves comparable accuracy with both a small and a large basis set. Additionally, our results indicate that increasing the fragment Hilbert space size systematically enhances the accuracy of observables, approaching the precision of the full CC solver.

## I. INTRODUCTION

Coupled cluster (CC) theory<sup>1</sup> provides the best balance between efficiency and accuracy for calculating the electronic structure of weak to moderately correlated molecules. However, CC theory with even the simplest singles and doubles truncation, still has computational cost scaling as  $\mathcal{O}(O^2V^4)$ , where  $O$  and  $V$  are the numbers of occupied and virtual orbitals being correlated. This precludes calculations on large molecular systems using large basis sets. At a lower level of accuracy, such problems are often addressed with composite methods. There is a long history of such methods in quantum chemistry, including the Gaussian- $n$  methods (G1-G4)<sup>2–5</sup>, the Weizmann- $n$  methods (W1-W4)<sup>6–8</sup>, the HEAT protocol<sup>9</sup>, the correlation consistent Composite Approach (ccCA)<sup>10,11</sup>, etc. As a simple example, the basis set correction to an expensive coupled cluster CC calculation such as CCSD(T) might be approximated using a far less demanding approach such as second-order Møller-Plesset theory (MP2) that exhibits similar basis set dependence via:

$$E_{\text{Composite}} \approx E_{\text{CC/small}} + (E_{\text{MP2/large}} - E_{\text{MP2/small}}). \quad (1)$$

In a similar additive fashion, the ONIOM<sup>12,13</sup> model treats different parts of a system with different levels of theory to tackle the system size scaling.

Local correlation methods<sup>14–17</sup> are a more advanced treatment of the system-size scaling that exploits the locality of electron correlations to approach linear scaling of compute ef-

fort with size. In contrast, our goal is to approximate the total energy using a high-level treatment for a specific local region of a big molecule and to control the computational expenses of such calculations by employing a lower-level approximation for the environment. This “QM-in-QM” approach is useful for addressing the chemistry of an active region of a catalyst, or point defects in materials, for example. Or, when the active fragment spans all the atoms in the molecule, a similar approach can be used to address the basis set convergence of the total problem. This will serve as an alternative method to overcome the bottleneck of large basis set calculations with CC, for which many methods have been already proposed, such as explicitly correlated R12/F12 methods<sup>18</sup>, diagrammatic separation-based extrapolation<sup>19</sup>, and cardinal number based extrapolation<sup>20</sup>, among many others.

Our effort to address the challenges of large system size and large basis set calculations can also be viewed from a quantum embedding perspective. In quantum embedding, one identifies an important (preferably local) region of the system, which is denoted as the fragment (F). The remainder of the system is referred to as the environment (E) and is treated by a low-scaling method, such as Hartree-Fock, density functional theory, or a perturbative method. For the fragment problem, one first evaluates a coupling potential ( $\Delta$ ) between the environment and the fragment. This coupling potential is then incorporated into the fragment Hamiltonian treating the fragment problem as an open-quantum system. The fragment problem is solved with a high-accuracy method (typically called a solver). CC methods have proven to be successful as a solver

for many embedding methods, especially when the fragment size needs to be large<sup>21–24</sup>. This highly accurate solution on the fragment is subsequently used to self-consistently update  $\Delta$ . The fragment problem will subsequently be described by a renormalized interaction, and solved by a CC method truncated at a certain excitation rank. The environment problem is then treated by a lower-scaling method such as MP2 theory<sup>25–27</sup>. The cost of the fragment problem remains constant with the increase of system size/basis set size and thus reduces the overall cost.

Here we combine the fragment and the environment problems using two sets of coupled projective equations: one for the environment and one for the fragment. Furthermore, we invoke a self-consistency condition between the fragment and the environment based on the “global” CC amplitudes. This approach inherits many essential ideas from the hybrid MP2 and CCSD methods developed by Nooijen<sup>28</sup> and Sherrill *et al.*<sup>29,30</sup> Both Nooijen and Sherrill *et al.* presented two versions of the theory. Their simplest version did not relax orbitals in the environment; however, in an improved version such contributions were included. In that version, certain classes of environment amplitudes (external and semi-internal) were updated using the same amplitude equations as CCSD. These amplitudes were chosen such that the computational cost remains low. This idea has its origin in earlier work by Adamowicz, Piecuch, and coworkers published between 1993 and 1998, which introduced partitioning of the cluster operator into internal and external components.<sup>31,32</sup>

Our approach differs from Nooijen and Sherrill *et al.*’s approach in two crucial aspects. First, instead of choosing canonical molecular orbital (CMO)s as a basis for the active space, we choose localized MO (LMO)s, which lets us define the active fragment as a bond or region. Second, the relaxation of the environment orbitals is performed differently (vide infra) with the objective in mind that it does not increase the cost of the low-level method even when we increase the excitation rank of the high-level CC method. Another related approach is the multi-level CC (MLCC) theory of Koch and co-workers<sup>33,34</sup>. This approach also utilizes a localized active space and also defines a mechanism for relaxing the environment. Differences between our approach and the MLCC models in terms of the general framework and computational implications will be elaborated in Sec. IV and Appendix A.

The remainder of the manuscript is organized as follows. In Section II, we describe the general framework, then outline various choices for the environmental projection equation, and finally discuss some theoretical aspects of the fragment amplitude equations. We will use CCSD as the fragment solver, although this choice can be extended to higher-level CC models. As in all embedding methods, it is important to choose a suitable local basis that describes the fragment problem well, as the results are not invariant with respect to this choice. Our choice of localized representation is described in Section III. To follow up on the review of existing embedding methods given above, in Section IV we compare some formal properties of our current method with related methods. Implementation details are briefly summarized in Section V. Finally, in Section VI we will illustrate various aspects of the theory via

a set of prototypical numerical examples. We will compare different versions of our embedding theory relative to the full high-level method, and a composite approach, which can be regarded as an uncoupled version of the embedding method.

## II. EMBEDDED COUPLED CLUSTER FORMALISM

Coupled cluster theory is a wave function approach that expresses the ground-state wave function using an exponential parametrization<sup>35,36</sup>. Given a single Slater determinant reference state like the Hartree-Fock (HF) determinant  $|\Phi_0\rangle$ , any intermediately normalized state  $|\Psi\rangle$  obeying  $\langle\Phi_0|\Psi\rangle = 1$  can be uniquely written as  $e^T|\Phi_0\rangle$ , where  $T$  is the corresponding cluster operator. To define the cluster operators, we introduce the notation that  $i, j, k, l, \dots$  label the occupied orbitals and  $a, b, c, d, \dots$  the virtual orbitals. The cluster operator is then given by

$$T = \sum_{\mu \in \mathcal{J}} t_\mu X_\mu = \sum_{k=1}^N \sum_{\substack{\mu \in \mathcal{J} \\ |\mu|=k}} t_\mu X_\mu, \quad (2)$$

where

$$\mathcal{J} = \left\{ \begin{pmatrix} a_1, \dots, a_k \\ i_1, \dots, i_k \end{pmatrix} : 1 \leq k \leq N \right\} \quad (3)$$

and  $X_\mu$  are particle-hole excitation operators<sup>1</sup>. The CC amplitudes, denoted by  $\mathbf{t} = (t_\mu)_{\mu \in \mathcal{J}}$ , are determined through projective equations:

$$\langle\Phi_\mu|\bar{H}|\Phi_0\rangle = 0 \quad \forall \mu \in \mathcal{J}, \quad (4)$$

where

$$\bar{H} = e^{-T} H e^T \quad (5)$$

and  $\langle\Phi_\mu| = \langle\Phi_0|X_\mu^\dagger$  denotes an excited Slater determinant with an excitation rank  $|\mu| > 0$ . The corresponding CC energy is then given by

$$E_{\text{CC}} = \langle\Phi_0|\bar{H}|\Phi_0\rangle. \quad (6)$$

Since the untruncated CC equations (4) rapidly become numerically intractable, truncations are commonly employed. The subject of this work is the CCSD variant, where  $T$  comprises excitation operators up to a maximum excitation rank of two, i.e.,

$$T = \sum_{i,a} t_i^a X_i^a + \frac{1}{4} \sum_{i,j,a,b} t_{ij}^{ab} X_{ij}^{ab}. \quad (7)$$

However, even CCSD has computational effort scaling as  $\mathcal{O}(O^2V^4)$ , where  $O$  and  $V$  denote the number of occupied and virtual orbitals included in the correlation treatment. Even CCSD becomes computationally demanding when naively applied to larger systems.

Here we explore an approximation to CCSD that is inspired by quantum embedding theory. The underlying core concept is well-known: it is the recognition that certain CC amplitudes possess greater physical significance than others if expressed in a suitable basis, such as localized or natural orbitals. The proposed approach aims to identify and accurately evaluate these critical amplitudes, while the less significant amplitudes are approximated via many-body perturbation theory with reduced precision (*vide infra*). We therefore name this approach the many-body perturbation coupled-cluster (MP-CC) method. It aims to maintain high accuracy while simultaneously reducing overall computational complexity.

Underlying the MP-CC approach is a partition of the total orbital space into two sets: the active fragment (F) orbitals and the environment (E) orbitals. The fragment orbitals describe the chemically relevant region that requires higher accuracy. The fragment orbitals are formed from intrinsic atomic orbitals (IAOs)<sup>37</sup> which span only a small chemically relevant subspace of the full basis set of the calculation. The environment orbitals, on the other hand, will describe additional virtual orbitals associated with the full atomic orbital basis, and all occupied orbitals not contained in the fragment orbital space. This allows the use of all valence occupied and some virtual orbitals as the fragment. Alternatively, the fragment can be just a single local site of particular interest (e.g. a catalytic active site), or a set of local sites.

A valid partition splits the occupied and virtual orbitals into environment and fragment subsets, such that

$$O = O_E + O_F \quad \text{and} \quad V = V_E + V_F. \quad (8)$$

Once a partition is made, there are also two classes of cluster operators:  $T_F$ , and  $T_E$ . Note that the  $T_E$  operators contain amplitudes that involve only environment orbitals, as well as mixed amplitudes that contain both fragment and environment orbitals. The latter describes correlations between electrons in fragment and environment occupied levels, or between two occupied fragment levels using environment virtuals, etc.

In the spirit of quantum embedding, we will compute the cluster operators  $T_F$ , and  $T_E$  using different levels of sophistication. More precisely, we propose to determine the fragment cluster operator  $T^F$  using the standard CC formalism (the high-level theory), involving a full expansion of  $\bar{H}$  (see Eq. (5)) to yield  $\bar{H}_F^{\text{high}} = \bar{H}$ . In contrast, the calculation of  $T_E$  is based on a lower-order perturbative expansion of  $\bar{H}$ , resulting in  $\bar{H}_E^{\text{low}}$  (*vide infra*). This lower-order expansion offers various options for gaining a computational advantage, as will be elaborated in Sec. II A. The MP-CC approach results in the following set of coupled projective equations:

$$\langle \Phi_\mu^F | \bar{H}_F^{\text{high}} | \Phi_0 \rangle = 0, \quad (9)$$

$$\langle \Phi_\mu^E | \bar{H}_E^{\text{low}} | \Phi_0 \rangle = 0. \quad (10)$$

We emphasize that both  $\bar{H}_F^{\text{high}}$  and  $\bar{H}_E^{\text{low}}$  in Eqs. (9) and (10) contain the full set of  $T$ -amplitudes i.e.,  $\mathbf{t}_F$  and  $\mathbf{t}_E$ . In other words, although Eq. (9) solves only for  $\mathbf{t}_F$ , the  $\mathbf{t}_E$  amplitudes still enter the equations. Similarly, Eq. (10) solves for  $\mathbf{t}_E$  in the presence of  $\mathbf{t}_F$ . Therefore, Eqs. (9) and (10) should be solved

self-consistently, unless they are uncoupled by the perturbative approximations. The total CC energy remains as Eq. (6).

Moreover, the energy and amplitude equations of the MP-CC approach can be combined into a Lagrangian:

$$\begin{aligned} \mathcal{L}(\mathbf{t}, \boldsymbol{\lambda}) = & \langle \Phi_0 | \bar{H} | \Phi_0 \rangle + \langle \Phi_0 | \Lambda_E \bar{H}_E^{\text{low}} | \Phi_0 \rangle \\ & + \langle \Phi_0 | \Lambda_F \bar{H}_F^{\text{high}} | \Phi_0 \rangle, \end{aligned} \quad (11)$$

where  $\Lambda = \sum_{\mu \in \mathcal{J}} \lambda_\mu X_\mu^\dagger$  consists of the dual variables  $\lambda_\mu$ , ensuring that Eqs. (9) and (10) are satisfied;  $\Lambda_E$  and  $\Lambda_F$  are defined in accordance with  $T_E$  and  $T_F$ . The working equations then correspond to a first-order optimality condition of the Lagrangian, i.e.,

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial \mathbf{t}_\alpha} = 0, & \alpha \in \{E, F\} \\ \frac{\partial \mathcal{L}}{\partial \boldsymbol{\lambda}_\alpha} = 0, & \alpha \in \{E, F\}. \end{cases} \quad (12)$$

Compared to other embedding approaches, the Lagrangian in Eq. (11) enables us to access a broad range of observables beyond the energy, by employing analytic energy derivative techniques<sup>38,39</sup>.

### A. Environment amplitude equations

We employ two distinct approaches to the lower-order expansion  $\bar{H}_E^{\text{low}}$ , both based on perturbative arguments.

#### 1. Unrelaxed approach

In the first approach, we use the simplest possible MP2-like projection equations for the environment amplitudes, i.e.,

$$\langle \Phi_\mu^E | V + [F, T] | \Phi_0 \rangle = 0. \quad (13)$$

While  $T$  contains all cluster amplitudes, Eqs. (13) are used to determine only  $T_E$ . In defining these equations, we have used the partitioned Hamiltonian:  $H = F + V$ , where  $F$  is the mean-field Hamiltonian or Fock operator from HF theory and  $V$  is the two-particle fluctuation potential, expressed in the HF basis. Note that we use normal-ordered operators for  $F$  and  $V$ , though we do not denote it explicitly. By the Brillouin theorem,  $\langle \Phi_s^E | V | \Phi_0 \rangle = 0$  for all singles,  $|\Phi_s^E\rangle$  so the environment orbitals are unrelaxed by singles amplitudes. We therefore refer to Eqs. 13 as the unrelaxed MP-CCSD method. Denoting fragment singles as  $S_F$ , fragment doubles as  $T_F$  and environment doubles as  $T_E^{(1)}$ , this is a wavefunction ansatz of the form:

$$|\Psi_{\text{MP-CC(unrelaxed)}}\rangle = e^{S_F} e^{T_E^{(1)} + T_F} |\Phi_0\rangle. \quad (14)$$

The unrelaxed method yields fixed environment amplitudes that influence the fragment amplitudes but not vice-versa.

## 2. Relaxed approach

In the second approach, we relax the wave function in the sense of Thouless relaxation<sup>40</sup>, wherein the HF determinant  $\Phi_0$  can be transformed into another determinant  $\Phi'_0$  by the exponential of a single substitution operator, denoted as  $S$ :  $\Phi'_0 \propto e^S \Phi_0$ . Higher-body amplitudes describing correlation, such as doubles will still be considered only to first order. We have made this distinction apparent by denoting singles amplitudes as  $S \equiv S_F + S_E$ , and environment doubles amplitudes as  $T_E^{(1)}$  in the relaxed MP-CC wave function ansatz

$$|\Psi\rangle = e^S e^{T_E^{(1)} + T_F} |\Phi_0\rangle. \quad (15)$$

More precisely, we will assign the formal perturbation parameter  $\eta$  to the interaction term of the similarity transformed Hamiltonian:  $\tilde{H} = e^{-S} H e^S$ , that is,  $\tilde{H} = E_{cl} + \tilde{F}^{(0)} + \eta \tilde{V}^{(1)}$ . Here,  $\tilde{F}$  and  $\tilde{V}$  are the one-particle and two-particle parts of the transformed Hamiltonian, respectively, and  $E_{cl}$  is a closed scalar contribution, which will not contribute to the projective equations. Keeping the terms up to first order in the perturbation,  $\eta$ , yields the following projective equations for the singles (s) and doubles (d) amplitudes:

$$\langle \Phi_s^E | \tilde{F} + [\tilde{F}, T_E^{(1)}] + [\tilde{F}, T_F] | \Phi_0 \rangle = 0, \quad (16)$$

$$\langle \Phi_d^E | \tilde{V} + [\tilde{F}, T_E^{(1)}] + [\tilde{F}, T_F] | \Phi_0 \rangle = 0. \quad (17)$$

In this work,  $T_E^{(1)}$  contains the first-order environment doubles amplitudes, while the fragment doubles amplitudes are in  $T_F$ . In the derivation of Eq. (16) and (17) we used the fact that the similarity transformation with  $e^S$  does not increase the operator rank of the Hamiltonian. By comparing Eqs. (16) and (17) with Eq. (10), we obtain the definition of  $\tilde{H}_E^{low}$ .

In Fig. 1, we schematically show the workflow of the proposed method ( $W_F$  is defined in II B, and not relevant for the following discussion). Note that when employing the relaxed method (i.e., the second approach outlined above) in step 3 (see Fig. 1), the first iteration uses  $S = 0$  implying that Eq. (16) and (17) reduce to Eq. (13). In particular, in the first iteration, the two procedures outlined above yield the same inactive amplitudes. However, in the subsequent iterations, non-zero active singles amplitudes generated from the  $T_F$ -amplitude equations will contribute to Eq. (16) and (17), thus yielding non-zero inactive singles amplitudes.

The computational scaling of Eq. (16) and (17) is MP2-like, using a factorized two-body Hamiltonian, as only low-rank singles amplitudes contract with them. The factorization of the two-body Hamiltonian can be carried out using either the density fitting (DF)<sup>41,42</sup> approach or a cholesky decomposition (CD)<sup>43</sup> approach. Following Mester *et al.*<sup>44</sup>, this yields  $\mathcal{O}(O^2 V^2 X)$  computational scaling for our relaxed approach, where  $X$  is the dimension of the auxiliary basis used in the DF or CD scheme.

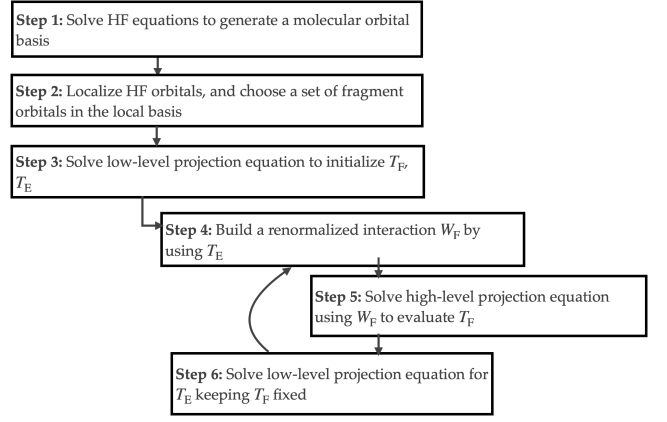


FIG. 1. Schematic view of the current embedding scheme.

## B. Fragment amplitude equations

An important observation is that in the fragment amplitude equations (9), the Hamiltonian  $\tilde{H}_F^{high}$  is screened by the environment amplitudes  $t^E$ . We may rewrite Eq. (9) in the following way:

$$\langle \Phi_\mu^F | e^{-T_F} W_F e^{T_F} | \Phi_0 \rangle = 0, \quad (18)$$

where

$$\begin{aligned} W_F &= [e^{-T_E} H e^{T_E}]_F \\ &= W_{1b,F} + W_{2b,F} + W_{3b,F} + \dots \end{aligned} \quad (19)$$

In Appendix B we show the working equations arising from Eq. (18).

In Eq. (19), we renormalize the bare Hamiltonian by applying a similarity transformation via  $e^{T_E}$  and then restrict the orbital indices to the fragment. Note that the similarity transformation increases the rank of  $W_F$  compared to the bare Hamiltonian. However, when solving Eq. (18), we order the operation of the tensors such that we avoid generating higher rank tensors of  $W_F$ . This particular construction is conceptually important because it is the mechanism by which the low-energy fragment (or active region) Hamiltonian is screened by the “high-energy” environment amplitudes, which is a crucial requirement for accurate quantum embedding theories<sup>45,46</sup>. We note that in a similar manner, Kowalski *et al.* constructed a CC downfolded Hamiltonian by using sub-system embedding sub-algebras (SES)<sup>47</sup>. In Fig. 1, we schematically show that in step 4 and step 5, we first construct  $W_F$ , and then iteratively solve Eq. (18). The maximum complexity of solving Eq. (18) for the fragment amplitudes is  $\mathcal{O}(O_F^2 V_F^4)$ . If there are multiple (disjoint) fragments, it also scales linearly with the number of fragments. The construction of  $W_F$ , on the other hand, is more computationally demanding, and scales as  $\mathcal{O}(O^2 O_F V^2 V_F)$ ; however, this cost is non-iterative.

### III. CHOICE OF BASIS SET FOR ACTIVE SPACE

The choice of a suitable basis to define the active space is crucial for embedding methods. The dimensionality and accuracy of the fragment impurity problem are dependent on this choice, because the basis determines spatial locality on the one hand, and the most suitable virtual orbitals to capture the main electron correlation effects in the fragment region on the other hand. To capture spatial locality, localized orbitals are of course essential. To capture the most relevant virtuals for electron correlation, without performing an active space calculation, we adopt the Atomic Valence Active Spaces (AVAS)<sup>51</sup> approach. AVAS straightforwardly defines valence occupied and virtual orbitals for a fragment, and is trivially able to incorporate higher angular momentum functions into the virtual space as discussed below. AVAS has important advantages over simply using canonical MOs, as illustrated in Appendix C.

The objective of AVAS is to isolate a set of valence/semi-valence atomic orbitals ( $A$ ) denoted by  $p, q$ , which has overlap with the HF occupied ( $|i\rangle$ ) and virtual orbitals ( $|a\rangle$ ) of the full molecule.

To this end, we define the projection operator

$$P_A = \sum_{p,q \in A} [\sigma^{-1}]_{pq} |p\rangle\langle q| \quad \text{where} \quad \sigma_{pq} = \langle p|q\rangle. \quad (20)$$

With this projection operator, we then construct two overlap matrices

$$[S^A]_{ij} = \langle i|P_A|j\rangle \quad \text{and} \quad [S^A]_{ab} = \langle a|P_A|b\rangle \quad (21)$$

in the space of occupied and virtual orbitals, respectively. To isolate the active orbitals,  $[S^A]_{ij}$  and  $[S^A]_{ab}$  are diagonalized and the orbitals  $|i\rangle$  and  $|a\rangle$  are rotated by the diagonalizing matrices to generate two new sets of orbitals:  $|\tilde{i}\rangle$  and  $|\tilde{a}\rangle$ . We then choose the active orbitals to be the vectors in  $|\tilde{i}\rangle$  and  $|\tilde{a}\rangle$  corresponding to non-zero eigenvalues. In the construction of  $[S^A]_{ij}$  matrix, we freeze the core occupied orbitals, which will then be added to the inactive (environment) orbitals.

The atomic orbitals defining the fragment must be specified before the above procedure can be invoked. Commonly, a minimal atomic orbital basis (e.g., such as MINAO<sup>37</sup> or autogenerated via AutoSAD<sup>52</sup>) is employed for constructing  $A$ . An important characteristic of the AVAS scheme is that one can alternatively employ a slightly larger reference basis set so that the active space includes orbitals of higher angular momentum. This aspect is important, for instance, in first-row transition metals, where often “double-shell” or 4d orbitals are required. We will project the AVAS reference basis such that it exactly spans an integer number of fragment occupied and virtual orbitals equal to its dimension in our system. This restriction is especially crucial for the occupied space.

### IV. COMPARISON WITH EXISTING APPROACHES

In the hybrid MP2 and CCSD approaches developed by Nooijen<sup>28</sup> and Sherrill *et al.*<sup>29</sup>, two very similar unre-

laxed schemes were introduced, namely, A-CC/PT and MP2-CCSD(I). Our unrelaxed scheme, as defined in Sec. II A 1, is identical to those approaches for a given choice of fragment orbitals. However, in contrast to the canonical molecular orbitals (CMOs) chosen in their implementations, we choose localized molecular orbitals (LMOs) from the AVAS scheme (Sec. III) to construct the active space, which has the advantage of treating a spatially localized fragment. Furthermore, both Nooijen and Sherrill *et al.* defined two improved versions of their methods, namely R-CC/PT and MP2-CCSD(II), respectively by including selected sets of external and semi-internal singles and semi-internal doubles amplitudes. In our approach, on the other hand, we relax the whole set of environment amplitudes (which corresponds to external plus semi-internal amplitudes) with a low-level method, leading to Eq. (16) and (17).

In the Multi-Level CC (MLCC) work by Koch and co-workers<sup>33</sup>, a CC2<sup>53</sup>-like approach was adopted to define the inactive amplitude equations. The CC2-like equations incorporate  $T_1$  amplitude equations to all orders while treating the  $T_2$  amplitude equations only up to the first order terms in conjunction with a second-order  $[V, T^F]$  term. By contrast, our approach consistently applies a first-order perturbative treatment to both the  $T_1$  and  $T_2$  amplitude equations to determine inactive environment CC amplitudes. Similar to the MLCC approach, the computational complexity of the inactive MP-CC amplitude equations is  $\mathcal{O}(O^2V^2X)$  at the singles and doubles level. However, the future inclusion of  $T_3$  amplitudes in the MP-CC framework – where active non-zero  $T_3$  amplitudes induce further relaxation of the environment orbitals – will not increase the complexity of the low-level projection equation.

In contrast, with the MLCC approach, the complexity of the low-level projection equations would have been higher. However, Koch *et al.* approximate the triple excitations to include only active indices to reduce the complexity. For a more detailed discussion on the differences between MLCC and MP-CC, we refer the interested reader to Appendix A. We have summarized all the comparisons in Table I. In that table we use uppercase letters ( $I, J, K, L, \dots$  for holes;  $A, B, C, D, \dots$  for particles) to denote orbitals in the fragment, and lowercase letters ( $i, j, k, l, \dots$  for holes;  $a, b, c, d, \dots$  for particles) for orbitals in the environment. A side-by-side numerical comparison of MP2-CCSD(II), MLCC and MP-CC is left for future work.

### V. IMPLEMENTATION

The computational results are obtained using a pilot implementation of the MP-CC method at the singles-doubles truncated level, that is MP-CCSD, based on the Python-based Simulations of Chemistry Framework (PySCF)<sup>54–56</sup>. In this implementation, the amplitude equations are separated into different classes for the fragment and environment, allowing us to experiment with the different heuristics mentioned in Sec. II A. We conduct all-electron calculations using CCSD as the computational method in both spin-restricted and spin-unrestricted formulations. For the spin-unrestricted version of MP-CCSD, unrestricted Hartree-Fock (UHF) orbitals will

method	MP-CCSD (Relaxed)	R-CC/PT	MP2-CCSD(II)	MLCCSD
reference	this work	28	30	33
$T_1^F + T_2^F$	All	All	All	All
Classes of $T_1^E$ relaxed	All	All	All	All
Classes of $T_2^E$ relaxed	All	$\{T_{IJ}^{aB}, T_{ij}^{AB}, T_{ij}^{aB}\}$	$\{T_{IJ}^{aB}, T_{ij}^{AB}\}$	All
Active space basis	AVAS	CMO	CMO	CD
Relaxation method	Sec. II A	CCSD	CCSD	CC2
Scaling of relaxation method	$\mathcal{O}(O^2V^2X)$	$\mathcal{O}(O^2V^3V_F)$	$\mathcal{O}(O^2V^3V_F)$	$\mathcal{O}(O^2V^2X)$

TABLE I. Comparison of the MP-CCSD embedding scheme with three other related approaches. Note the relaxation method scaling cost is per iteration, and CD stands for Cholesky decomposition based localized orbitals.

be used. Since all calculations employ localized orbitals, the MP2 equations, as in Eq. (13), are of the non-canonical type.

We want to comment on a few aspects of the implementation of steps (4)-(6) in Fig. 1. We solve the projection equations in both Step 5 and Step 6 iteratively. After solving Step 6 we rebuild  $W^F$ , and solve Step 5 and 6 again, thereby introducing a global outer loop. The entire set of equations are then solved in a double-loop fashion. Similar to the inner loop that involves step 5 and step 6, we employ the DIIS scheme for the outer loop as well. The number of iterations in the outer loop is usually much smaller than that in the inner loop.

## VI. RESULTS

### A. Diatomic potential energy curves

We apply our embedding method to the potential energy curves (PECs) of  $N_2$  and CO to study the following aspects of the theory:

1. To assess the accuracy of different approaches, we analyze the PECs with different approaches, namely MP2, both relaxed and unrelaxed versions of the MP-CCSD method and also a composite method. For the composite method, we modify Eq. (1) to make it an uncoupled approximation to the current embedding scheme:

$$E_{\text{Composite}} = E_{\text{MP2,large}} + E_{\text{CCSD,small(fc)}} - E_{\text{MP2,small(fc)}}, \quad (22)$$

where large and small stand for large, small AO basis sets considered for the current calculation, and fc stands for the frozen core calculation (only if it is used).

We compare the error in energy with respect to the CCSD method, using two metrics: the maximum absolute error (MAE) and the non-parallelity error (NPE).

The spin unrestricted energy at dissociation will be the zero of energy. We furthermore evaluate several spectroscopic parameters, namely, equilibrium bond distance ( $R_e$ ), harmonic frequency ( $\omega_e$ ), and first anharmonicity constant ( $\omega_e x_e$ ) by the least square fit of the PEC with a polynomial with high enough degree such that the parameter values converge.

2. We investigate the method's convergence as we systematically increase the size of the active space aiming to assess systematic improvability.
3. We show consistent behavior of the MP-CCSD method with increased AO basis set size when we keep the active space fixed. This observation is significant, considering the challenges often encountered in achieving similar enhancements within many quantum embedding schemes, for example, with Schmidt decomposition-based DMET<sup>57</sup> and Bootstrap Embedding (BE)-DMET<sup>58</sup> methods.

We will employ three different choices for the small basis (MINAO) used for IAO construction, giving three different active spaces as summarized in Table II. Note that we employ unrestricted orbitals that can break the  $S^2$  symmetry of the electronic Hamiltonian, while allowing us to reach lower variational energies than with restricted orbitals. Notably, UHF shows the correct qualitative behavior in the dissociation limit.

	name	active orbitals	MINAO	$(n_e, n_{\text{orb}})$
A	minimal	1s 2s 2p	STO-3G	(14,10)
B	split valence (SV)	1s 2s 2p 3s 3p	SV	(14,18)
C	SV + polarization	1s 2s 2p 3s 3p 3d	cc-pVDZ	(14,28)

TABLE II. Choices of active space for the  $N_2$  and CO diatomic molecules, and the small basis (denoted as MINAO) used for IAO construction.

#### 1. The nitrogen molecule ( $N_2$ )

The PEC of  $N_2$  often serves as a benchmark problem in electronic structure theory due to the complexity involved in the dissociation of its triple bond. UCCSD for the whole system will serve as the reference method by which we test the embedded MP-CCSD approximations, employing the aug-cc-pCVDZ and aug-cc-pCVTZ basis sets. While not strictly relevant for assessing embedding approaches, we note that UCCSD is not very accurate relative to a higher rank CC method, UCCSDT. While UCCSD shows correct dissociative behavior, its MAE and NPE versus UCCSDT are 29 mH and 20 mH, respectively with the aug-cc-pCVDZ basis set.

Fig. 2 shows PECs obtained with the reference UCCSD method and the various approximations, such as UMP2 and the embedding approaches (using the split valence active space B; see Table II). We observe that the PECs obtained from various methods are not of similar quality. This is especially true close to the Coulson-Fischer point at 1.2Å where

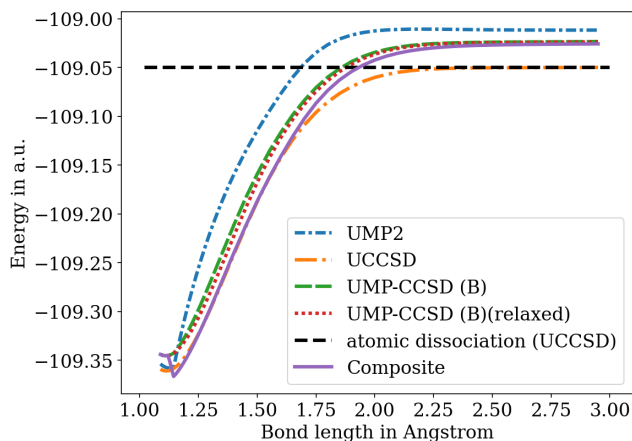


FIG. 2. PEC of  $N_2$  molecule in aug-cc-pCVDZ basis set with various methods. B active space is as described in Table II.

UMP2 and the composite method show a first derivative discontinuity. Its origin lies in the fact that UMP2 is incapable of removing UHF’s spin contamination error.<sup>59</sup> Towards the dissociative limit, all methods are qualitatively satisfactory, due to using unrestricted orbitals.

To focus on the relative errors, Fig. 3 plots the differences between each approximation and the UCCSD reference, as well as showing embedding results in the minimal active space A. Figure 3 shows that UMP2 errors peak near the Coulson-Fischer point and its derivative exhibits discontinuity. The UMP-CCSD method reduces the MAE from 80 mH (UMP2) to 38 mH / 30 mH, and the NPE from 78 mH (UMP2) to 24 mH / 17 mH for A / B choice of active spaces, respectively. Similarly, the composite scheme reduces MAE and NPE to 24 mH and 24 mH, respectively. However, the composite method contains *two* derivative discontinuities which result from the different Coulson-Fischer points in the small and large basis sets, and is clearly undesirable! The unrelaxed variant of UMP-CCSD shows a derivative discontinuity at the same bond distance as UMP2. Very encouragingly, this discontinuity is lifted using the relaxed version of UMP-CCSD, showing the importance of the *environment singles amplitudes* in symmetry restoration.

To further improve the UMP-CCSD energies, we augment the active space with N 3d orbitals (the SVP active space C). We report the relaxed and unrelaxed UMP-CCSD results in aug-cc-pCVDZ (DZ) and aug-cc-pCVTZ (TZ) basis set for active space choices B and C in Fig. 4. Comparing the left and right panels of Fig. 4 shows that MAE and NPE have significantly improved in the SVP active space C compared to the smaller SV active space B. With active space C, the error in energy is well within chemical accuracy ( $< 1$  kcal/mol) for the relaxed method, and it is close ( $\approx 2$  kcal/mol) with the unrelaxed method.

We will now analyze the AO basis set convergence of embedding errors with the aid of Fig. 4. We have enlarged the basis set from aug-cc-pCVDZ (54 basis functions) to aug-cc-pCVTZ (118 basis functions) for the SV (B) and SVP (C) active spaces. For relaxed UMP-CCSD, we observe a nearly

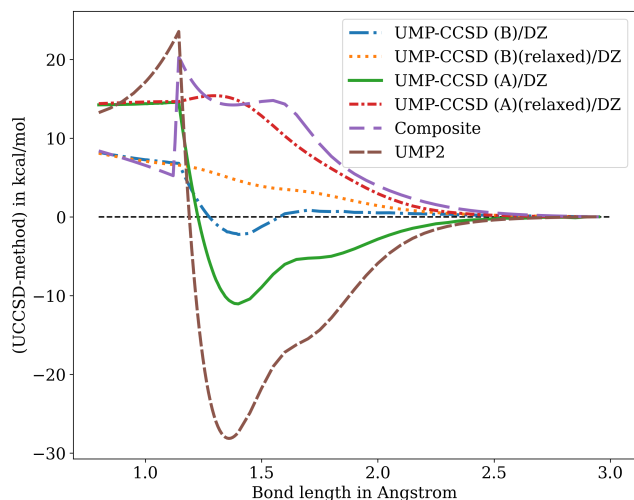


FIG. 3. Electronic energy error with respect to UCCSD along a PEC for  $N_2$  in the aug-cc-pCVDZ basis set. A and B active spaces are as described in Table II. DZ stands for the aug-cc-pCVDZ basis set. We set the energy at dissociation as the zero energy for each method.

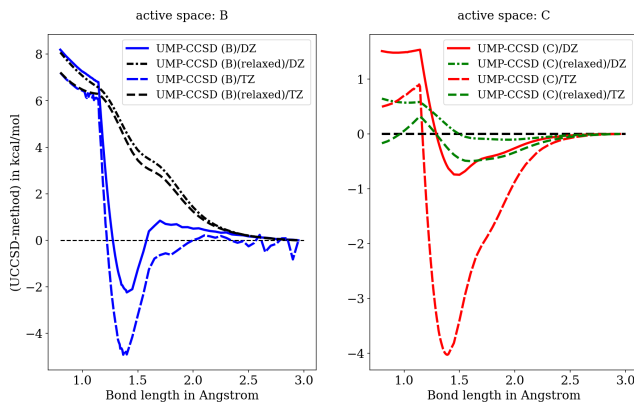


FIG. 4.  $N_2$  electronic energy error with respect to UCCSD in aug-cc-pCVDZ (denoted as DZ) and aug-cc-pCVTZ (denoted as TZ) basis sets. B and C active spaces are as described in Table II. We set the energy at dissociation as the zero energy for each method.

constant error as we increase the AO basis from double- $\zeta$  to triple- $\zeta$  quality. However, for the unrelaxed UMP-CCSD variant, we observe that the error increases substantially from double- $\zeta$  to triple- $\zeta$ . We also observed a few instabilities in the error curve of the unrelaxed UMP-CCSD method, corresponding to at least two solutions. With the relaxed approach, we do not see such instabilities, meaning no evidence of multiple solutions, as well as no loss of accuracy when enlarging the basis set.

The quality of the PECs and the difference curve with various methods, as described so far, influences the spectroscopic parameters as shown in Table III. Compared to UCCSD, we observe that UMP-CCSD reproduces  $R_e$ ,  $\omega_e$  and  $\omega_e x_e$  better than the composite method and UMP2. The failure of the composite method and UMP2 in estimating  $\omega_e$  and  $\omega_e x_e$  can be attributed to the discontinuities in the PEC near the equilib-

Method	$\Delta R_e$ (in Å)	$\Delta \omega_e$ (in $\text{cm}^{-1}$ )	$\Delta \omega_e x_e$ (in $\text{cm}^{-1}$ )
UMP2	0.0059	1230.06	120.3261
Composite	0.0539	327.78	320.5781
UMP-CCSD (C)	0.0003	86.85	-6.8039
UMP-CCSD (C)(relaxed)	0.0006	2.60	-1.7131

TABLE III. Spectroscopic parameters of  $\text{N}_2$  evaluated in aug-cc-pCVTZ basis set. We reported the difference w.r.t. the UCCSD method for all the spectroscopic parameters. In UMP-CCSD (C), C stands for the active space defined in Table II.

rium region. Although the relaxed and unrelaxed variants of UMP-CCSD yield similar MAE and NPE along the PEC, the energy difference plots around the equilibrium region become much smoother when orbital relaxations are considered. This improvement is reflected in the spectroscopic parameters  $\omega_e$  and  $\omega_e x_e$ , which are coefficients of higher order terms in the polynomial. They now become significantly closer to the corresponding UCCSD results when orbital relaxations are taken into account. Hence, the self-consistency cycle improves the behavior around the equilibrium region significantly.

## 2. Carbon monoxide (CO)

We now consider the isoelectronic carbon monoxide molecule. The full PECs for UMP2, UCCSD, relaxed and unrelaxed UMP-CCSD, and the composite method are given in Appendix Fig. 15. Beginning with the cc-pVDZ basis and the large SVP active space C, the energy difference curves for all methods are shown in Fig. 5 (top panel). There are two derivative discontinuities in the UMP2 curve and multiple derivative discontinuities in the composite method curve. The latter arises because the composite method inherits discontinuities from its individual computations; here multiple discontinuities appear at different bond distances for different basis sets, see “UMP2 (large)” and “UMP2 (small)” graphs in Appendix Fig. 14. Using the unrelaxed version of UMP-CCSD, we observe one derivative discontinuity, which can be lifted by including orbital relaxations.

The bottom panel of Fig. 5, plots the basis set dependence of the error in the relaxed embedding method and the composite method. As for  $\text{N}_2$ , we observe less than 1 kcal/mol loss of accuracy when enlarging the basis set from aug-cc-pCVDZ to aug-cc-pCVTZ. For the uncoupled composite scheme, the extent of agreement varies strongly across the PEC, and is qualitatively worse than the relaxed UMP-CCSD method. This issues the value of electronic embedding.

We also calculate the spectroscopic parameters  $R_e$ ,  $\omega_e$  and  $\omega_e x_e$  for the carbon monoxide molecule, which are listed in Table IV. Compared to the preceding  $\text{N}_2$  example, the composite method shows noticeable improvement over UMP2 for all computed quantities. We attribute this to the qualitative agreement of the composite method’s PEC with the PEC of UCCSD near the equilibrium region, see Appendix Fig. 15, and the fact that the discontinuities appear further away from equilibrium. We still observe that  $R_e$  and  $\omega_e$  are best repro-

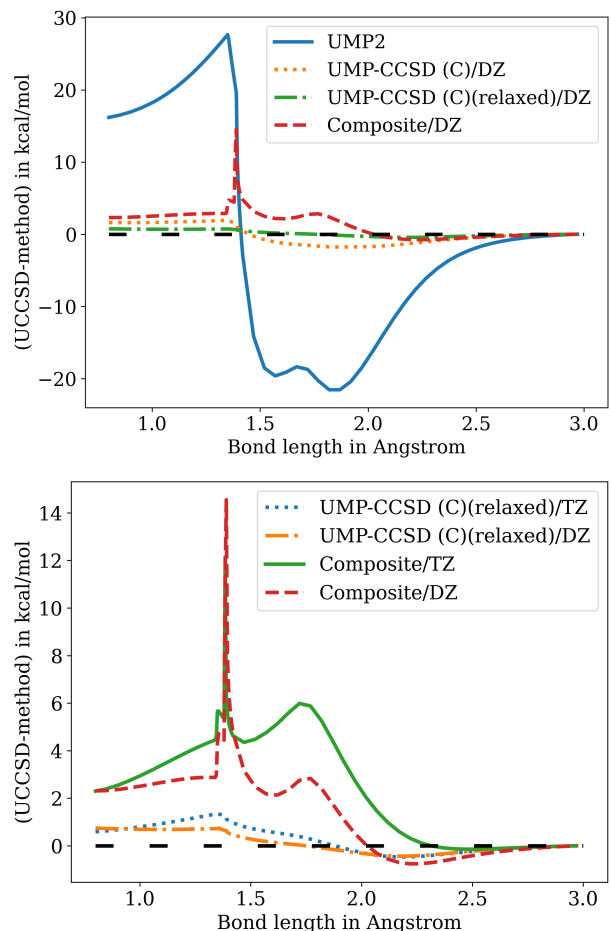


FIG. 5. Electronic energy differences w.r.t. UCCSD along a PEC for CO. DZ and TZ are abbreviations for aug-cc-pCVDZ and aug-cc-pCVTZ basis sets respectively. We set the energy at dissociation as the zero energy for each method.

Method	$\Delta R_e$ (in Å)	$\Delta \omega_e$ (in $\text{cm}^{-1}$ )	$\Delta \omega_e x_e$ (in $\text{cm}^{-1}$ )
UMP2	0.0101	-105.82	1.221
Composite	0.0017	-12.76	-0.0878
UMP-CCSD (C)	0.0007	-6.73	-0.1200
UMP-CCSD (C)(relaxed)	0.0005	-4.65	0.0817

TABLE IV. Spectroscopic parameters of CO evaluated in aug-cc-pCVTZ basis set. We reported the difference to the UCCSD results for all spectroscopic parameters. Note that “(C)” in “UMP-CCSD (C)”, stands for the active choice defined in Table II.

duced with the UMP-CCSD method, showing a consistent improvement over the composite method.

## B. W4 Dataset

We now investigate the accuracy of the embedding approaches on thermochemical energy differences using the well-known W4-11<sup>60</sup> dataset. The cc-pCVTZ basis set is used. We have chosen cc-pVDZ as the small basis reference

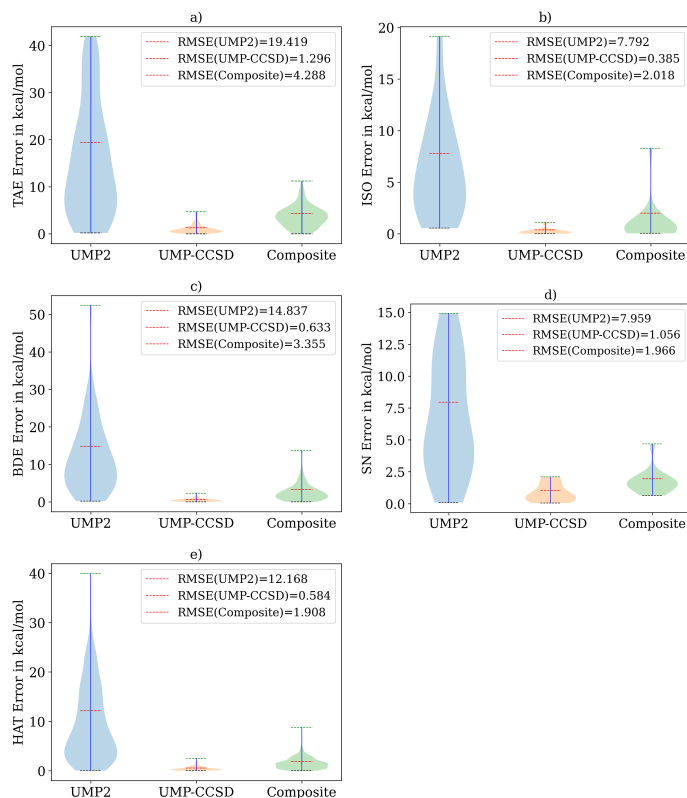


FIG. 6. Violin plot of the errors of UMP2, relaxed UMP-CCSD, and the composite method compared to the corresponding CCSD results for various thermochemical reactions with W4-11 dataset in the cc-pCVTZ basis set. The different panels show a) the total atomization energy, b) the isomerization energy, c) the bond dissociation energy, d) the nucleophilic substitution reaction energy, and e) the heavy atom transfer energy. All atoms are included in the fragment, with active orbitals defined via the use of cc-pVDZ in the AVAS scheme.

to generate the fragment active space (i.e., like active space C in Table II). All atoms are again included in the fragment. We use only the W4-11 single reference (SR) systems, for which UCCSD serves as the reference to test the MP-CCSD embedding models as well as the composite model and MP2. The multi-reference (MR) systems are less accurately described than the SR systems via UCCSD<sup>61</sup> and are excluded. Altogether, we compute 140 total atomization energies (TAEs) and bond dissociation energies (BDEs), 20 isomerization energies (ISO), 505 heavy atom transfer (HAT), and 13 nucleophilic substitution (SN) reactions.

Figure 6 shows violin plots of the absolute errors relative to UCCSD for different thermochemical reactions. We observe that the root mean square error (RMSE) in UMP-CCSD is significantly improved compared to UMP2 for all classes of reactions. Moreover, relaxed UMP-CCSD provides a significant improvement over the composite results. It is encouraging that the RMSE of UMP-CCSD is at or within chemical accuracy for all the reactions considered in this study. Additionally, we observe that the maximal spread of the deviation is also much more favorable for relaxed UMP-CCSD than for the other methods considered.

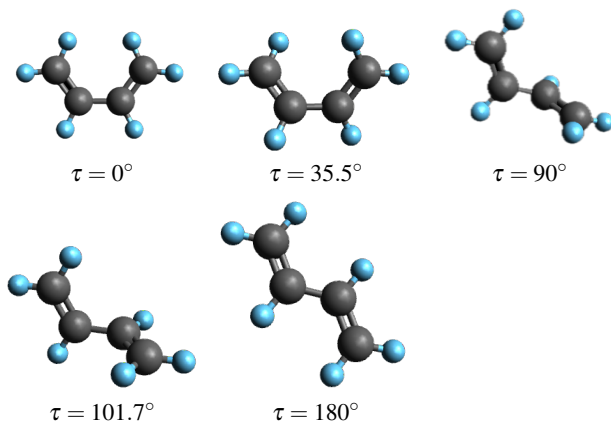


FIG. 7. Rotational isomers of 1,3-butadiene.

### C. Rotational isomers of 1,3-butadiene

In this section, we report on the torsional potential around the central C–C bond of 1,3-butadiene,  $\text{CH}_2=\text{CH}-\text{CH}=\text{CH}_2$ . This conjugated molecule has been extensively studied in the literature,<sup>62–64</sup> as its torsional potential is an archetype of the effect of loss of conjugation between the two double bonds at torsional angles ( $\tau$ ) close to  $90^\circ$ . More specifically, we have studied five different rotamers as shown in Fig. 7. The optimized structures for all those isomers were taken from Feller *et al.*<sup>63</sup>, optimized at the highly accurate CCSD(T)/aug-cc-pVQZ level of theory. Following our previous examples, we have compared the UMP-CCSD and UMP2 methods with the all-electron UCCSD method, using the cc-pCVTZ basis set. Unlike the previous examples, we have chosen a fragment that is localized spatially as well as limited in its active orbitals. Thus we use the minimal STO-3G basis on only the 4 carbon atoms and consider only the C 2s 2p orbitals to comprise the active space. While the molecule has 30 electrons, 22 of them remain in the fragment, which has a total of 25 orbitals. To construct IAOs within the AVAS scheme, we have considered the STO-3G basis set as our minimal basis.

While it is not widely known (and indeed we were surprised to observe it), HF results exhibit spin-polarization all along the torsional potential (see Appendix Table V), so UHF energies lie below RHF energies. This indicates the presence of some biradicaloid character in butadiene, which can be interpreted as arising from the  $\bullet\text{CH}_2-\text{CH}=\text{CH}-\bullet\text{CH}_2$  resonance structure. It is well-known that UMP2 cannot recover symmetries that are broken at the UHF level, due to the absence of orbital rotations<sup>65</sup>, sometimes leading to poor results.<sup>66</sup> Indeed this artificial HF symmetry-breaking turns the torsional potential of butadiene from an easy problem for RMP2<sup>62,64</sup> into a challenging problem for UMP2. This is evident in Fig. 8 where the relative energy of all the isomers was plotted relative to the *trans* isomer. UMP2 inverts the correct form of the torsional potential, predicting  $\tau = 90^\circ$  as the most stable structure, whereas it should be very close to the barrier.

By contrast, UCCSD relative energies agree quite well in comparison to the benchmark theoretical data reported by

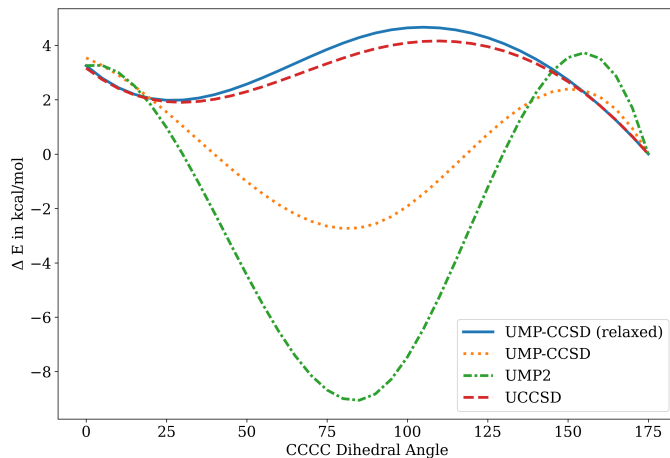


FIG. 8. Relative energy of different rotational isomers along the torsional potential around the central C–C bond of the butadiene molecule,  $\text{H}_2\text{C}=\text{CH}-\text{CH}=\text{CH}_2$ , with various unrestricted methods. The *trans*, that is, the isomer at  $\tau = 180^\circ$  has been considered as the zeroth of energy. See text for computational details.

Feller *et al.*<sup>63</sup>, due to the restoration of spin symmetry breaking to a large extent (see Appendix Table V) in the presence of singles amplitudes. Turning to the UMP-CCSD methods, due to the absence of inactive singles in the unrelaxed version of the UMP-CCSD method, it is still qualitatively incorrect, although improved relative to UMP2. By contrast, upon inclusion of orbital rotations via  $T_1$  the relaxed version of the UMP-CCSD method improves significantly upon both the UMP2 method and the unrelaxed UMP-CCSD method and shows results that are qualitatively and quantitatively very similar to the UCCSD method. The interpretation is straightforward:  $T_1$  has largely removed the symmetry-breaking seen at the UHF level. This is further evidenced in Fig. 9, where the orbital-optimized MP2 (OOMP2)<sup>66</sup> method based on an unrestricted reference also removes the inverted trend of the UMP2 method.

#### D. Azomethane

In order to further test the efficiency of the UMP-CCSD method in capturing system relative energies using a spatially localized active fragment we have studied the PEC of azomethane,  $\text{H}_3\text{C}-\text{N}=\text{N}-\text{CH}_3$ , along the  $\text{N}=\text{N}$  double bond stretching coordinate. All the geometric parameters, except the  $\text{N}=\text{N}$  bond distance of this molecule, were taken from Hermes *et al.*<sup>67</sup>. We use the aug-cc-pCVDZ basis set, with all electrons correlated. For the active fragment, we have chosen only the 2s and 2p orbitals of the 2 nitrogen atoms, as obtained via the AVAS scheme. Also, to construct the IAOs within the AVAS scheme, we have chosen the minimal STO-3G basis set, resulting in a very small fragment consisting of 16 electrons in 16 orbitals.

In Fig. 10 we have plotted the PEC with various methods and also the difference plot with respect to the UCCSD

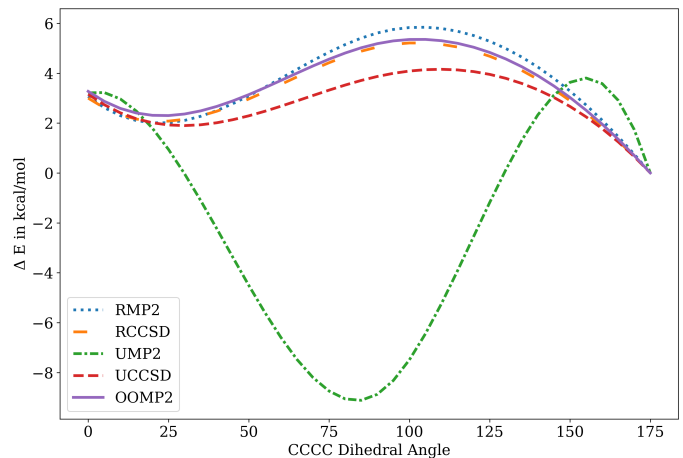


FIG. 9. Relative energy of different rotational isomers of the butadiene molecule (w.r.t. the *trans* isomer) with both restricted and unrestricted methods. See text for computational details.

method. First we identify the Coulson-Fischer (CF) point by plotting the PEC with the RHF and UHF methods, which separate at that point. The UMP2 method shows a pronounced first derivative discontinuity (kink) at the CF geometry. The unrelaxed version of the UMP-CCSD method numerically improves upon the UMP2 method, but it cannot remove the kink. However, the relaxed UMP-CCSD method produces a smooth PEC, illustrating the key role of environment  $T_1$  amplitudes.

These characteristics of the PEC w.r.t. the UCCSD method are even more pronounced in the difference plot (right panel of Fig. 10). Interestingly, we observe apparent derivative discontinuities at bond lengths longer than the CF point with both UMP2 and unrelaxed UMP-CCSD, suggesting that the UHF solution changes character there. The relaxed version of the UMP-CCSD method, on the other hand, removes all discontinuities very satisfactorily. This analysis suggests a qualitatively correct behavior of the relaxed version of the UMP-CCSD method for local chemistry. However, the NPE of the PEC remains quite high with the current (very small) choice of active space, which presumably is coming from the larger error towards the dissociation regime.

To reduce the NPE, we expanded the size of the active space, by using N 2s 2p 3s 3p 3d orbitals, where cc-pVDZ basis set was considered as the reference basis set for IAO construction. Fig. 11 shows that this choice significantly reduces the NPE. We then further increased the size of the basis set for the actual calculation to cc-pCVTZ, and chose the latter (larger) active space. We observed that when the relaxed scheme is employed, the error remains almost constant w.r.t. the cc-pCVDZ basis set, which corroborates our previous observations for diatomic dissociation.

## VII. CONCLUSION

We have proposed and numerically scrutinized the new MP-CC quantum embedding scheme, where the active fragment

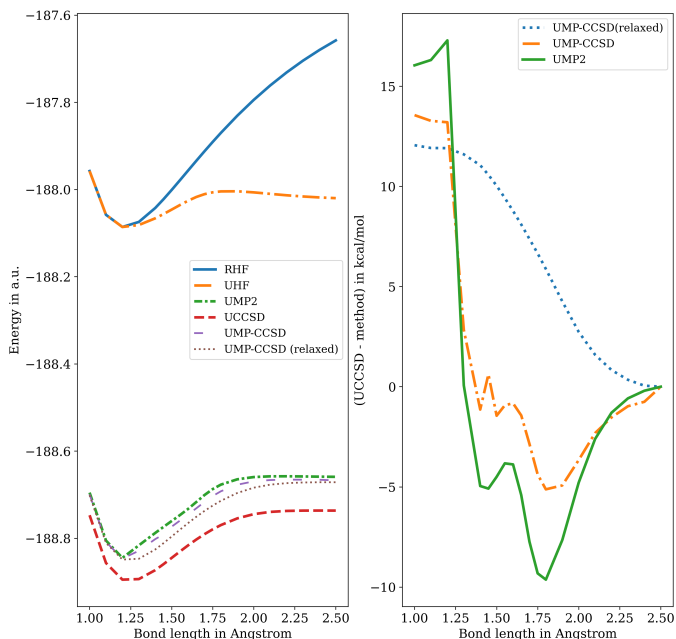


FIG. 10. (left panel) PEC of N=N double bond stretching of azomethane in the cc-pCVDZ basis set. (right panel) Difference plot w.r.t. the UCCSD method. The smallest (valence) active space on the 2 N atoms, as described in the text, is used for this plot. For the difference plot, we set the energy at dissociation as the zero energy for each method

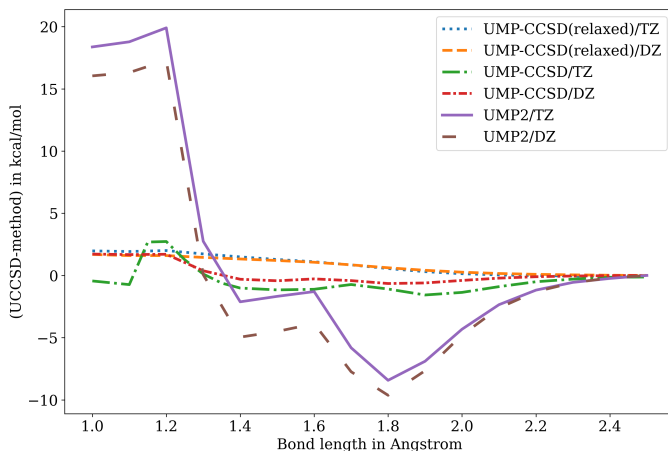


FIG. 11. Energy difference w.r.t. the UCCSD method for azomethane molecule. DZ and TZ stand for the cc-pCVDZ and cc-pCVTZ basis sets, respectively. A larger local active space on the 2 N atoms, as described in the text, is used for this plot. We set the energy at dissociation as the zero energy for each method.

problem is solved at the coupled cluster (CC) level, and the environment problem is solved at a level similar to 2nd order Møller-Plesset (MP) perturbation theory. Instead of constructing a density matrix or Green’s function at the CC level, MP-CC relies only on the cluster amplitudes to separate the environment and the fragment problem. Self-consistency is ensured through amplitude equations coupling the fragment

and the environment. In particular, we do not need to construct a hybridization term between these two layers to take “environment fluctuations” into account. Furthermore, the interaction term for the fragment problem is renormalized by the low-level environment amplitudes.

MP-CC provides a general framework allowing the use of different low-level methods for solving the environment problem, while the high-level method for the fragment problem remains fixed as a CC method. In this work, we fixed the CC method as CCSD, and proposed and compared two different MP2-like environment models: one without orbital relaxation and one with orbital relaxation. The relaxed version significantly improves upon the unrelaxed one by removing a derivative discontinuity in the error curve relative to UCCSD results. By contrast, a composite method (or the uncoupled embedding method) numerically improves upon UMP2, but inherits its first derivative discontinuities of UMP2, making it significantly inferior to relaxed UMP2-CCSD. Unrelaxed UMP2-CCSD is intermediate in performance between the composite theory and the preferable relaxed UMP2-CCSD approach. For relaxed UMP2-CCSD, a significant finding is that we achieve similar accuracy for a fixed active space when the basis set size is increased, which is generally not guaranteed for embedding methods.

The efficacy of relaxed MP-CCSD has been further established by studying rotational isomerism in 1,3-butadiene. The UMP2 method predicts an anomalous trend in the relative energies of the rotamers, which the unrelaxed version of our method cannot rectify. However, the relaxed UMP-CCSD method can recover the correct trend even with a very small active space.

We have investigated the reliability of the proposed orbital relaxed UMP-CCSD method for single reference systems in the W4-11 thermochemistry dataset. Our investigations show that relaxed UMP-CCSD reproduces, on average, the UCCSD results to chemical accuracy.

Furthermore, we have studied the central nitrogen double bond stretching of azomethane as a test case of a localized fragment. The relaxed UMP-CCSD method shows an almost constant energy difference relative to the all-electron UCCSD energy along the PEC, whereas the UMP2 method and the unrelaxed UMP-CCSD methods show non-smooth behavior near the symmetry-breaking region of this molecule.

There is ample scope for future work. The implementation reported and used here is not yet fully optimal, because of an unfavorable  $\mathcal{O}(V^2 O^2 O_F V_F)$  scaling when the renormalized interaction is built before solving the fragment amplitude equations (Fig. 1). Approximation schemes to construct the renormalized interaction could achieve lower polynomial scaling and also facilitate an efficient multi-fragment approach. This could be an alternative to existing embedding approaches such as LASSCF<sup>67,69–71</sup>, DMET<sup>57,72–75</sup>, SEET<sup>76,77</sup>, DMFT<sup>78–80</sup> or g-RISB<sup>81,82</sup>. Moreover, there is also the possibility of utilizing efficient local correlation methods to reduce the complexity further. With respect to the high-level CC method, it would be very desirable to extend the theory to CCSDT; preliminary analysis has shown that MP-CC has a favorable scaling when including triples amplitudes. With respect to the low-level MP

method, it will be interesting to investigate alternatives to MP2 such as Brillouin-Wigner perturbation theory (BWPT)<sup>68</sup>, random phase approximation (RPA) or ring coupled cluster doubles (r-CCD).

## VIII. ACKNOWLEDGEMENTS

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research and Office of Basic Energy Sciences, Scientific Discovery through Advanced Computing (SciDAC) program under Award Number DE-SC0022198 (A.S., K.B.W. and L.L.), and Award Number DE-SC0022364 (M.H-G.) and by the Simons Targeted Grants in Mathematics and Physical Sciences on Moiré Materials Magic (F.F.). L.L. is a Simons Investigator in Mathematics. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 using NERSC award BES-ERCAP0029462. A.S. and L.L. thank Karol Kowalski for insightful discussions. Additionally, the authors thank the anonymous reviewers for their comments, which have contributed to the improvement of this manuscript.

- <sup>1</sup>J. Čížek, J. Chem. Phys. **45**, 4256 (1966), <http://dx.doi.org/10.1063/1.1727484>.
- <sup>2</sup>J. A. Pople, M. Head-Gordon, D. J. Fox, K. Raghavachari, and L. A. Curtiss, J. Chem. Phys. **90**, 5622 (1989).
- <sup>3</sup>L. A. Curtiss, K. Raghavachari, G. W. Trucks, and J. A. Pople, J. Chem. Phys. **94**, 7221 (1991).
- <sup>4</sup>L. A. Curtiss, K. Raghavachari, P. C. Redfern, V. Rassolov, and J. A. Pople, J. Chem. Phys. **109**, 7764 (1998).
- <sup>5</sup>L. A. Curtiss, P. C. Redfern, and K. Raghavachari, J. Chem. Phys. **126**, 084108 (2007).
- <sup>6</sup>J. M. L. Martin and G. de Oliveira, J. Chem. Phys. **111**, 1843 (1999).
- <sup>7</sup>A. D. Boese, M. Oren, O. Atasoylu, J. M. L. Martin, M. Kállay, and J. Gauss, J. Chem. Phys. **120**, 4129 (2004).
- <sup>8</sup>A. Karton, E. Rabinovich, J. M. L. Martin, and B. Ruscic, J. Chem. Phys. **125**, 144108 (2006).
- <sup>9</sup>A. Tajti, P. G. Szalay, A. G. Császár, M. Kállay, J. Gauss, E. F. Valeev, B. A. Flowers, J. Vázquez, and J. F. Stanton, J. Chem. Phys. **121**, 11599 (2004).
- <sup>10</sup>N. J. DeYonker, T. R. Cundari, and A. K. Wilson, The Journal of Chemical Physics **124**, 114104 (2006).
- <sup>11</sup>N. J. DeYonker, B. R. Wilson, A. W. Pierpont, T. R. Cundari, and A. K. Wilson, Molecular Physics **107**, 1107 (2009).
- <sup>12</sup>F. Maseras and K. Morokuma, Journal of Computational Chemistry **16**, 1170 (1995).
- <sup>13</sup>L. W. Chung, W. M. C. Sameera, R. Ramozzi, A. J. Page, M. Hatanaka, G. P. Petrova, T. V. Harris, X. Li, Z. Ke, F. Liu, H.-B. Li, L. Ding, and K. Morokuma, Chemical Reviews **115**, 5678 (2015).
- <sup>14</sup>S. Saebo and P. Pulay, Annu. Rev. Phys. Chem. **44**, 213 (1993).
- <sup>15</sup>J. E. Subotnik, A. Sodt, and M. Head-Gordon, J. Chem. Phys. **125**, 074116 (2006).
- <sup>16</sup>C. Riplinger and F. Neese, J. Chem. Phys. **138**, 034106 (2013).
- <sup>17</sup>Q. Ma and H.-J. Werner, WIREs Comput. Mol. Sci. **8**, e1371 (2018).
- <sup>18</sup>G. Knizia, T. B. Adler, and H.-J. Werner, J. Chem. Phys. **130**, 054104 (2009).
- <sup>19</sup>A. Irmmler and A. Grüneis, J. Chem. Phys. **151**, 104107 (2019).
- <sup>20</sup>T. Helgaker, W. Klopper, H. Koch, and J. Noga, J. Chem. Phys. **106**, 9639 (1997).
- <sup>21</sup>Z.-H. Cui, H. Zhai, X. Zhang, and G. K.-L. Chan, Science **377**, 1192 (2022).
- <sup>22</sup>A. Shee and D. Zgid, J. Chem. Theory Comput. **15**, 6010 (2019).
- <sup>23</sup>T. Zhu and G. K.-L. Chan, Phys. Rev. X **11**, 021006 (2021).
- <sup>24</sup>C.-N. Yeh, A. Shee, S. Iskakov, and D. Zgid, Phys. Rev. B **103**, 155158 (2021).
- <sup>25</sup>H.-J. Werner, G. Knizia, C. Krause, M. Schwilke, and M. Dornbach, J. Chem. Theory Comput. **11**, 484 (2015).
- <sup>26</sup>P. Pinski, C. Riplinger, E. F. Valeev, and F. Neese, J. Chem. Phys. **143**, 034108 (2015).
- <sup>27</sup>Z. Wang, A. Aldossary, and M. Head-Gordon, J. Chem. Phys. **158**, 064105 (2023).
- <sup>28</sup>M. Nooijen, The Journal of Chemical Physics **111**, 10815 (1999).
- <sup>29</sup>A. D. Bochevarov and C. D. Sherrill, The Journal of chemical physics **122** (2005).
- <sup>30</sup>A. D. Bochevarov, B. Temelso, and C. D. Sherrill, The Journal of chemical physics **125** (2006).
- <sup>31</sup>P. Piecuch, N. Oliphant, and L. Adamowicz, The Journal of chemical physics **99**, 1875 (1993).
- <sup>32</sup>P. Piecuch and L. Adamowicz, The Journal of chemical physics **100**, 5792 (1994).
- <sup>33</sup>R. H. Myhre, A. M. J. Sánchez de Merás, and H. Koch, J. Chem. Phys. **141**, 224105 (2014).
- <sup>34</sup>S. D. Folkestad, E. F. Kjønestad, L. Goletto, and H. Koch, J. Chem. Theory Comput. **17**, 714 (2021).
- <sup>35</sup>J. Hubbard, Proceedings of the Royal Society A **240**, 539 (1957).
- <sup>36</sup>N. Hugenholtz, Physica **23**, 533 (1957).
- <sup>37</sup>G. Knizia, J. Chem. Theory Comput. **9**, 4834 (2013).
- <sup>38</sup>N. C. Handy and I. Schaefer, Henry F., J. Chem. Phys. **81**, 5031 (1984).
- <sup>39</sup>H. Koch, H. J. A. Jensen, P. Jørgensen, T. Helgaker, G. E. Scuseria, and I. Schaefer, Henry F., J. Chem. Phys. **92**, 4924 (1990).
- <sup>40</sup>D. J. Thouless, Nuclear Physics **21**, 225 (1960).
- <sup>41</sup>J. L. Whitten, The Journal of Chemical Physics **58**, 4496 (1973).
- <sup>42</sup>B. I. Dunlap, J. W. D. Connolly, and J. R. Sabin, The Journal of Chemical Physics **71**, 3396 (1979).
- <sup>43</sup>H. Koch and A. Sánchez de Merás, The Journal of Chemical Physics **113**, 508 (2000).
- <sup>44</sup>D. Mester, P. R. Nagy, and M. Kállay, The Journal of Chemical Physics **146**, 194102 (2017).
- <sup>45</sup>E. G. C. P. van Loon, M. Rösner, M. I. Katsnelson, and T. O. Wehling, Phys. Rev. B **104**, 045134 (2021).
- <sup>46</sup>X.-J. Han, P. Werner, and C. Honerkamp, Phys. Rev. B **103**, 125130 (2021).
- <sup>47</sup>K. Kowalski, J. Chem. Phys. **158**, 054101 (2023).
- <sup>48</sup>J. Pipek and P. G. Mezey, J. Chem. Phys. **90**, 4916 (1989).
- <sup>49</sup>S. F. Boys, Rev. Mod. Phys. **32**, 296 (1960).
- <sup>50</sup>A. Damle, L. Lin, and L. Ying, J. Chem. Theory Comput. **11**, 1463 (2015).
- <sup>51</sup>E. R. Sayfutyarova, Q. Sun, G. K.-L. Chan, and G. Knizia, J. Chem. Theory Comput. **13**, 4063 (2017).
- <sup>52</sup>M. Gimferrer, A. Aldossary, P. Salvador, and M. Head-Gordon, J. Chem. Theory Comput. **18**, 309 (2022).
- <sup>53</sup>O. Christiansen, H. Koch, and P. Jørgensen, Chemical Physics Letters **243**, 409 (1995).
- <sup>54</sup>Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt, N. A. Bogdanov, G. H. Booth, J. Chen, Z.-H. Cui, J. J. Eriksen, Y. Gao, S. Guo, J. Hermann, M. R. Hermes, K. Koh, P. Koval, S. Lehtola, Z. Li, J. Liu, N. Mardirossian, J. D. McClain, M. Motta, B. Mussard, H. Q. Pham, A. Pulkin, W. Purwanto, P. J. Robinson, E. Ronca, E. R. Sayfutyarova, M. Scheurer, H. F. Schurkus, J. E. T. Smith, C. Sun, S.-N. Sun, S. Upadhyay, L. K. Wagner, X. Wang, A. White, J. D. Whitfield, M. J. Williamson, S. Wouters, J. Yang, J. M. Yu, T. C. Zhu, T. C. Berkelbach, S. Sharma, A. Y. Sokolov, and G. K.-L. Chan, J. Chem. Phys. **153**, 024109 (2020).
- <sup>55</sup>Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, *et al.*, Wiley Interdisciplinary Reviews: Computational Molecular Science **8**, e1340 (2018).
- <sup>56</sup>Q. Sun, Journal of computational chemistry **36**, 1664 (2015).
- <sup>57</sup>G. Knizia and G. K.-L. Chan, J. Chem. Theory Comput. **9**, 1428 (2013).
- <sup>58</sup>H.-Z. Ye and T. Van Voorhis, J. Phys. Chem. Lett. **10**, 6368 (2019).
- <sup>59</sup>W. Kurlancheek and M. Head-Gordon, Mol. Phys. **107**, 1223 (2009).
- <sup>60</sup>A. Karton, S. Daon, and J. M. Martin, Chem. Phys. Lett. **510**, 165 (2011).
- <sup>61</sup>J. Lee, H. Q. Pham, and D. R. Reichman, J. Chem. Theory Comput. **18**, 7024 (2022), pMID: 36255074.
- <sup>62</sup>M. Head-Gordon and J. A. Pople, The Journal of Physical Chemistry **97**, 1147 (1993).

- <sup>63</sup>D. Feller and N. C. Craig, J. Phys. Chem. A **113**, 1601 (2009).
- <sup>64</sup>A. Karpfen and V. Parasuk, Mol. Phys. **102**, 819 (2004).
- <sup>65</sup>R. H. Nobes, J. A. Pople, L. Radom, N. C. Handy, and P. J. Knowles, Chemical physics letters **138**, 481 (1987).
- <sup>66</sup>J. Lee and M. Head-Gordon, J. Chem. Theory Comput. **14**, 5203 (2018), pMID: 30130398.
- <sup>67</sup>M. R. Hermes and L. Gagliardi, Journal of Chemical Theory and Computation **15**, 972 (2019), pMID: 30620876.
- <sup>68</sup>K. Carter-Fenk and M. Head-Gordon, J. Chem. Phys. **158**, 234108 (2023).
- <sup>69</sup>M. R. Hermes, R. Pandharkar, and L. Gagliardi, J. Chem. Theory Comput. **16**, 4923 (2020).
- <sup>70</sup>H. Q. Pham, V. Bernales, and L. Gagliardi, J. Chem. Theory Comput. **14**, 1960 (2018).
- <sup>71</sup>V. Agarawal, D. King, M. Hermes, and L. Gagliardi, ChemRxiv (2024), 10.26434/chemrxiv-2024-sbp4r.
- <sup>72</sup>G. Knizia and G. K.-L. Chan, Phys. Rev. Lett. **109**, 186404 (2012).
- <sup>73</sup>H. Q. Pham, V. Bernales, and L. Gagliardi, J. Chem. Theory Comput. **14**, 1960 (2018).
- <sup>74</sup>M. Nusspickel and G. H. Booth, Phys. Rev. X **12**, 011046 (2022).
- <sup>75</sup>M. Nusspickel, B. Ibrahim, and G. H. Booth, J. Chem. Theory Comput. **19**, 2769 (2023).
- <sup>76</sup>A. A. Kananenka, E. Gull, and D. Zgid, Phys. Rev. B **91**, 121111(R) (2015).
- <sup>77</sup>S. Isakov, C.-N. Yeh, E. Gull, and D. Zgid, Phys. Rev. B **102**, 085105 (2020).
- <sup>78</sup>A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, Rev. Mod. Phys. **68**, 13 (1996).
- <sup>79</sup>G. Kotliar, S. Y. Savrasov, K. Haule, V. S. Oudovenko, O. Parcollet, and C. A. Marianetti, Rev. Mod. Phys. **78**, 865 (2006).
- <sup>80</sup>V. I. Anisimov, A. I. Poteryaev, M. A. Korotin, A. O. Anokhin, and G. Kotliar, J. Phys.: Condensed Matter **9**, 7359 (1997).
- <sup>81</sup>T.-H. Lee, T. Ayral, Y.-X. Yao, N. Lanata, and G. Kotliar, Phys. Rev. B **99**, 115129 (2019).
- <sup>82</sup>T.-H. Lee, N. Lanata, and G. Kotliar, Phys. Rev. B **107**, L121104 (2023).
- <sup>83</sup>J. F. Stanton, J. Gauss, J. D. Watts, and R. J. Bartlett, The Journal of Chemical Physics **94**, 4334 (1991).

## Appendix A: Comparison with the MLCC method

In the multi-level CC (MLCC) method, Koch and co-workers defined a wave function ansatz

$$|\Psi\rangle = e^X |\Phi_0\rangle; \quad X = T_1 + T_2^{act} + T_2^{inact} \quad (A1)$$

where,  $T_1$  is both an active and inactive set of singles amplitudes,  $T_2^{act}$ , and  $T_2^{inact}$  stand for active, and inactive doubles amplitudes. Then to derive the inactive projection equations as in Eq. (10), they assign a perturbative order for  $T_2^{inact}$ , but the rest are treated in all orders. Furthermore, the projection equation for inactive  $T_1$  amplitudes is the same as CC amplitude equations, but the one for  $T_2^{inact}$  is truncated at the first order. The projection equations for  $T_2$  active amplitudes do not differ from CC. This scheme is similar to using CC2<sup>53</sup> equations for the inactive amplitudes. Then, to include triples excitation, they augment X with active triples amplitudes,  $T_3^{act}$ . Therefore, their treatment of the doubles and triples amplitudes is not on the same footing. With this ansatz their low-level projection equations contain singles and doubles equations, but no triples equations. When triples are included in that work, both for the singles and doubles equations, there is a  $[V, T_3]$  term. But because of the active restriction in  $T_3$ , it maximally scales as  $n_{v,act}^3 n_{o,act}^3 n_v$ .

Our projection equations for the active amplitudes are full CC amplitude equations similar to the MLCC method, but the inactive projection equations are derived differently than that method. We do not define a wave function ansatz for the total problem but tried to pick a subset of terms/diagrams from the full set of terms of the CC amplitude equation. This subset of terms is chosen based on a certain heuristic that ensures the low-scaling nature of the inactive amplitude equation. It allows us to choose a projection equation like Eqs. (16) and (17), where we have truncated both the singles and doubles equations in the same order of perturbation, unlike the CC2 projection equations. This choice produces a minimal model that can induce orbital relaxation for the environment when the active space/fragment  $T_1$  amplitudes become nonzero. When we include active  $T_3$  equations for the fragment, it does not give rise to any new sets of inactive  $T_3$  amplitudes because of the lack of  $[V, T_3]$  or  $[V, T_2]$  terms which are second-order terms (note that, the perturbative order argument here is different than the Møller-Plesset perturbation series as we treat  $T_1$  amplitudes to all orders for the low-level method). Therefore, with our approach when triples are included, the complexity of the low-level equations does not increase compared to the singles-doubles equations.

## Appendix B: Working equations for fragment amplitude equation:

We have summarized the working equations for the fragment amplitude equation in this section. We will use uppercase letters (I, J, K, L,... for hole type; A, B, C, D,... for particle type) to denote orbitals in the fragment, and lowercase letters (i, j, k, l,... for hole type; a, b, c, d,... for particle type) for orbitals in the environment. The coupled cluster residue has been denoted as  $R$ ,  $f$ ,  $V$  tensors stand for bare interaction, and  $F$ ,  $W$  stand for similarity transformed one- and two-particle interactions. We have not shown all the explicit expressions for  $F$  and  $W$ , as they can be found in many coupled cluster literature<sup>83</sup>.

### 1. One-body residue:

$$R_I^A = F_I^A + F_E^A t_I^E - F_I^M t_M^A - V_{EI}^{AM} t_M^E + F_E^M t_{IM}^{AE} + \frac{1}{2} V_{EF}^{AM} \tau_{IM}^{EF} - \frac{1}{2} W_{EI}^{MN} t_{MN}^{EA} \quad (B1)$$

where,

$$F_I^A = f_I^A + F_e^A t_I^e - F_I^m t_m^A - V_{eI}^{Am} t_m^e + F_e^m t_{Im}^{Ae} + \frac{1}{2} V_{ef}^{Am} \tau_{Im}^{ef} - \frac{1}{2} W_{eI}^{mn} t_{mn}^{eA} \quad (B2)$$

$$\tau_{ij}^{ab} = t_{ij}^{ab} + P(ij) \frac{1}{2} t_i^a t_j^b \quad (B3)$$

$$P(pq)f(p, q) = f(p, q) - f(q, p) \quad (B4)$$

### 2. Two-body residue:

$$\begin{aligned} R_{IJ}^{AB} = & W_{IJ}^{AB} + P(IJ) V_{EJ}^{AB} T_I^E - P(AB) W_{IJ}^{AM} T_M^B + P(AB) F_E^A T_{IJ}^{EB} - P(IJ) F_I^M T_{MJ}^{AB} + \frac{1}{2} V_{EF}^{AB} \tau_{IJ}^{EF} \\ & + \frac{1}{2} W_{IJ}^{MN} \tau_{MN}^{AB} + P(AB) P(IJ) W_{EI}^{AM} T_{JM}^{EB} \end{aligned} \quad (B5)$$

where,

$$\begin{aligned} W_{IJ}^{AB} = & V_{IJ}^{AB} + P(IJ) V_{eJ}^{AB} T_I^e - P(AB) W_{IJ}^{Am} T_m^B + F_e^A T_{IJ}^{eB} - P(IJ) F_I^m T_{mJ}^{AB} + \frac{1}{2} V_{ef}^{AB} \tau_{IJ}^{ef} \\ & + \frac{1}{2} W_{IJ}^{mn} \tau_{mn}^{AB} + P(AB) P(IJ) W_{eI}^{Am} T_{Jm}^{eB} \end{aligned} \quad (B6)$$

The most expensive term in the above amplitude equations is the following one:

$$W_{eI}^{Am} = 0.5 * V_{ef}^{mn} T_{nI}^{Af} \quad (B7)$$

### Appendix C: The choice of basis for the active space:

We have numerically compared two different choices for the active space basis: canonical molecular orbital (CMO)s, and intrinsic atomic orbital (IAO) basis within the active valence active space (AVAS) scheme. AVAS scheme allows us to choose the active orbitals for a specific fragment of the system. For this comparison, we have chosen the example of  $N_2$  molecule as described in VI. We have used aug-cc-pCVDZ basis set and B-type active space as described in Table II for this study. Our findings are summarized in Fig. 12 where we have employed the relaxed version of our embedding scheme. We found that only the active space chosen with the AVAS scheme can remove the kink near the Coulson-Fischer point.

### Appendix D: Additional plots

We have added Fig. 13 and 14 to aid the analysis of composite method over the PEC of  $N_2$  and CO respectively. Additionally, Fig. 15 was added to show the section of PEC of CO where multiple discontinuities arise for the UMP2 method and the composite method.

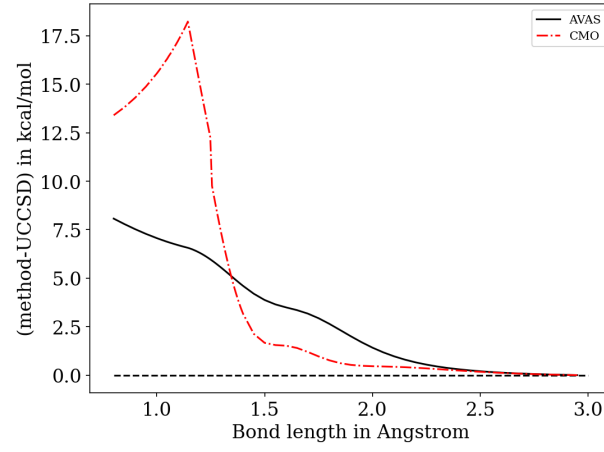


FIG. 12. Energy difference between the UCCSD and relaxed UMP-CCSD methods for  $N_2$  molecule in aug-cc-pCVDZ basis set. Active space B as described in Table II was considered for this study.

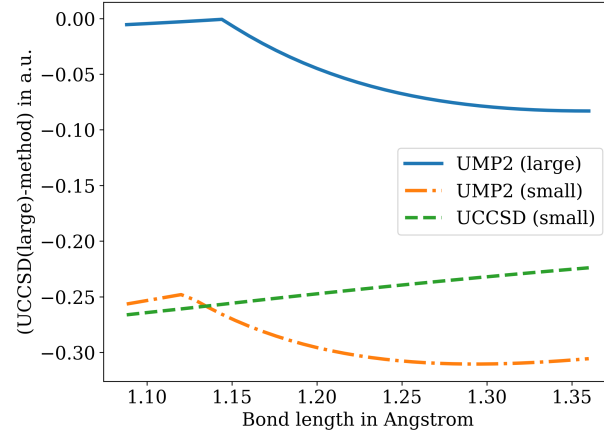


FIG. 13. Correlation energy differences w.r.t. UCCSD in aug-cc-pCVDZ basis set along a PEC for  $N_2$ . large and small basis sets stand for aug-cc-pCVDZ and SVP respectively.

#### Appendix E: Measure of spin symmetry breaking in 1,3-butadiene

In this section, we have tabulated the  $\langle S^2 \rangle$  value as a measure of spin symmetry breaking in various rotational isomers of 1,3-butadiene molecule.

Isomer	$\langle S^2 \rangle_{UHF}$	$\langle S^2 \rangle_{UCCSD}$
$\tau = 0^\circ$	0.450596	0.011599
$\tau = 35.5^\circ$	0.310997	0.004780
$\tau = 90^\circ$	0.003347	0.000019
$\tau = 101.7^\circ$	0.039688	0.000247
$\tau = 180^\circ$	0.384190	0.008753

TABLE V.  $\langle S^2 \rangle$  for different rotamers of 1,3-butadiene from UHF and UCCSD methods.

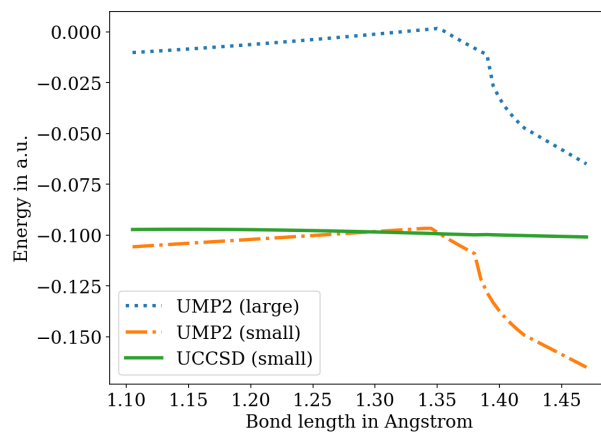


FIG. 14. Correlation energy differences w.r.t. UCCSD in aug-cc-pCVDZ basis set along a PEC for CO. large and small basis sets stand for aug-cc-pCVDZ and cc-pVDZ respectively.

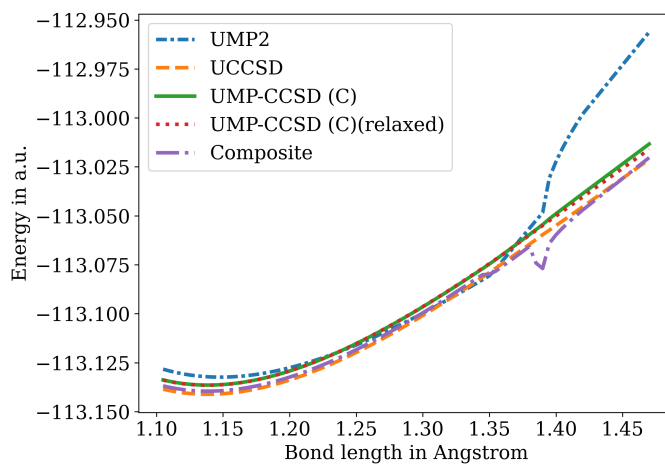


FIG. 15. Total energy differences w.r.t. UCCSD in aug-cc-pCVDZ basis set along a PEC for CO.