Automatic partition of orbital spaces based on singular value decomposition in the context of embedding theories

Daniel Claudino and Nicholas J. Mayhall[*]

Department of Chemistry, Virginia Tech, Blacksburg, VA 24060, USA

We present a simple approach for orbital space partitioning to be employed in the projection-based embedding theory developed by Goodpaster and coworkers [*J. Chem. Theory Comput.* 2012, 8, 2564]. Once the atoms are assigned to the desired subspaces, the molecular orbitals are projected onto the atomic orbitals centered on active atoms and then singular value decomposed. The right singular vectors are used to rotate the initial molecular orbitals, taking the largest gap in the singular values spectrum to define the most suitable partition of the occupied orbital space. This scheme is free from numerical parameters, contrary to the Mulliken charge threshold or the completeness criterion previously used. The performance of this new prescription is assessed in a test set of several distinct reactions, the deprotonation of decanoic acid, the torsional potential of a retinal derivative, and the critical points along the reaction coordinate of an example of the Menshutkin S_N2 reaction inside a carbon nanotube.

I. INTRODUCTION

Despite the fact that, for a given one-particle basis, the exact electronic, non-relativistic wave function is formally known, i.e., the full configuration interaction (FCI), its factorially growing computational cost renders most of its practical application to problems of chemical and physical interest prohibitively expensive. Several truncations of the FCI expansion or alternative *ansätze*, which are collectively referred to as wave function theories (WFT) hereafter, have been successful in providing tunable accuracy and systematic convergence toward the FCI limit, but even these approximations are often times too demanding computationally. One viable avenue is to resort to density functional theory (DFT), whose inclusion of electron correlation within a mean-field framework greatly offsets the computational bottleneck imposed by the more accurate WFT approaches. On the other hand, DFT is plagued by the lack of a formal route toward the FCI limit and its considerable inability to remove the self-interaction error, among other problems. Fortunately, DFT can be comparable to WFT methods in many instances, if not always quantitatively, at least in allowing for more qualitative predictions.

Most of the current understanding of chemistry relies on the knowledge based on the qualitative picture of chemical sub-units. This idea readily translates into the notion of local environments, relying on the assumption that the chemistry of an entire class of compounds can be reduced to one or few of the so-called functional groups. This fact naturally leads to the idea of multilevel approaches, which works under the assumption that the bulk of the sought chemistry/physics can be achieved if the region(s) primarily responsible for the phenomenon under consideration is (are) treated at a higher, more accurate level of theory, while the remainder of the system in question, deemed lower in importance, can be incorporated through methods that are more amenable from the computational viewpoint.

Several examples of multilevel schemes can be found in the literature. Some of the earlier endeavors on this front attempted to describe the background by embedding the high-level region in the electrostatic potential of the environment, treating it in an additive manner in (QM/MM) methods,[1] or in a subtractive manner in ONIOM-type methods.[2–6]

While electrostic interactions dominate in the long-range, short term interactions are more nuanced, requiring more robust embedding procedures.[7–9] In density matrix embedding theory (DMET),[8,10–13] the active region is treated as an open-quantum system coupled to a similarly sized quantum bath, which together are embedded in a mean field background. The benefits of DMET lie in the fact that the active region is allowed to remain entangled with the nearby atoms, enabling the approach to be used with strongly correlated systems.

If strong correlation between the active and the environment regions is not present, then more efficient embedding strategies based on DFT might be used.[14–16] Given the particular advantages of WFT and DFT, it seems desirable to combine their strengths within a multilevel framework. Recently Manby, Miller, and co-workers started exploring this route.[17–21] In their approach, the entire system is initially computed at the DFT level, yielding a set of Kohn-Sham (KS) orbitals. These orbitals are in turn made local according to the Pipek-Mezey (PM) localization procedure,[22] and identified with the atoms in the "active" and "environment" subsystems, named A and B, respectively. It is worth noting that the entire system can be partitioned in more than two subsystems without loss of generality of the arguments presented here. This step is carried out by evaluating the Mulliken charges and including all orbitals whose Mulliken charges have a contribution from one of the active atoms that is 0.4 or greater. Inspired by the work of Huzinaga and Cantu,[23] the two subsystems can be made orthogonal, that is, they can be treated independently, by means of introducing a projector with the goal of isolating the contribution from the orbitals in B from the total density of the system. This projector is accompanied by a level-shift parameter μ that ensures exact

orthogonality between A and B in the limit of $\mu \to \infty$, which is, for practical purposes, taken to be 10^6, with negligible error. This approach, with some slight modifications, has been applied successfully to a wide range of chemical problems.[24–28]

Although the projection-based embedding has already seen significant successes, its use as a "black box" method presents some challenges. The Mulliken charge threshold employed in the partition of the A and B orbital subspaces is rather arbitrary, while the very concept of Mulliken charge can have deleterious effects, as it does not have a basis set limit and is not always in accordance with well-established chemical intuition backed by empirical evidence.[29,30] In fact, such charges tend to degrade as the basis set is enlarged. It can also pose problems in the underlying Pipek-Mezey localization step, as it in turn depends on the Mulliken charge analysis.

One attempt to tackle these deficiencies was proposed by Hégely et al.[31] The Mulliken charge analysis above was replaced by the algorithm proposed by Boughton and Pulay (BP)[32] commonly used to categorize local orbitals in terms of spatial domains, and instead of the projector and level-shift parameter, the Huzinaga equation was explicitly solved. Even though the BP scheme is believed to be more reliable than the Mulliken analysis, it is still based on a seemingly arbitrary 0.015 least-squares residual and the presence of Mulliken charges is not entirely prevented, as the authors used PM localized orbitals. Resorting to the solution of the Huzinaga equation, according to the authors, "eliminates the bias caused by the arbitrary level shift parameter of the original theory, though it has a very small effect on the numerical results."[31] This scheme has also been explored in conjunction with the dual basis scheme in order to further alleviate the underlying computational burden.[33]

Alternatively, Hammes-Schiffer and co-workers proposed a mechanism through which they simultaneously optimize the orbitals in A and B such that the resulting subspaces are mutually orthogonal by construction.[34] By doing so, the orbital energies in B no longer need to be pushed to very high energies, meaning that it is also free from the level-shift parameter. This approach is reported to display smaller errors than the Huzinaga-based embedding scheme discussed above. However, A and B are again comprised of PM orbitals and are assigned to which subspace based on a predetermined Mulliken charge threshold (not reported), thus some of the drawbacks pointed out above may also afflict this prescription.

In order to remedy these weaknesses, we propose an alternative way to construct the orbitals and their partitioning into the appropriate subsystems. The motivating idea is that *conventional orbital localization procedures (like PM) seek too stringent of a localization condition, finding orbitals of near atomic character. However, a good embedding calculation only needs the occupied orbitals regionally localized*, and this is the strategy we take in this paper. It is independent of Mulliken charges and is made completely black-box, meaning that it does not rely on PM or any other explicit localization algorithm.

II. THEORY

A. Projection-based embedding theory

The main working equations follow from the first paper from Manby et al.[17] and are summarized below in order to facilitate the subsequent discussion. An initial calculation is carried out at the low level of theory, DFT in the present case, involving the entire system and which is used to describe the environment. At the end of this step, one is left with a set of molecular orbitals (MO) $\{\phi_i\}$, with i indexing the occupied MOs, and an atomic orbital (AO) density matrix γ. Upon proper partition of the occupied orbital space into the two subsystems A and B, it follows that $\{\phi_i\}=\{\phi_i^A\}\bigcup\{\phi_i^B\}$ and $\gamma = \gamma^A + \gamma^B$.

The total energy of the unpartitioned system can be written as:

$$E[\gamma^A + \gamma^B] = \underbrace{\text{tr}(\gamma^A \mathbf{h}) + \mathbf{g}[\gamma^A]}_{\text{Energy of isolated A}}$$
$$+ \underbrace{\text{tr}(\gamma^B \mathbf{h}) + \mathbf{g}[\gamma^B]}_{\text{Energy of isolated B}} \quad (1)$$
$$+ \underbrace{\mathbf{g}[\gamma^A, \gamma^B]}_{\text{Non-additive two-electron term}}$$

where the $\mathbf{h}$ is the one-electron Hamiltonian matrix and $\mathbf{g}$ groups the two-electron terms that contain the Coulomb, exchange and (potentially) correlation contributions. The term $\mathbf{g}[\gamma^A, \gamma^B]$ gathers the non-additive two-electron terms that account for the interaction between subsystems. The intersystem interaction can be suppressed and later incorporated with the aid of the projector $\mathbf{P} = \mathbf{S}\gamma^B\mathbf{S}$, where $\mathbf{S}$ is the AO overlap matrix. Thus, the Fock matrix for subsystem A, $\mathbf{F}^A$, becomes:

$$\mathbf{F}^A = \mathbf{h}^{A \text{ in } B} + \mathbf{g}[\gamma^A_{\text{emb}}] \quad (2)$$

where the effect of the interaction of A with B is accounted for via an effective one-particle potential that is taken as an additional term to the usual $\mathbf{h}$ and subsequent calculation yields γ^A_{emb}, the density matrix of A embedded in B, that is, self-consistently determined by $\mathbf{F}^A$. Explicitly:

$$\mathbf{h}^{A \text{ in } B} = \mathbf{h} + \mathbf{g}[\gamma^A + \gamma^B] - \mathbf{g}[\gamma^A] + \mu\mathbf{P} \quad (3)$$

with μ being the level-shift parameter, which is set to 10^6. The final energy expression of A embedded in B becomes:

$$E[\boldsymbol{\gamma}_{\mathrm{emb}}^A; \boldsymbol{\gamma}^A, \boldsymbol{\gamma}^B]$$
$$= E[\boldsymbol{\gamma}_{\mathrm{emb}}^A] + E[\boldsymbol{\gamma}^B] + \mathbf{g}[\boldsymbol{\gamma}^A, \boldsymbol{\gamma}^B] \qquad (4)$$
$$+ \mathrm{tr}[(\boldsymbol{\gamma}_{\mathrm{emb}}^A - \boldsymbol{\gamma}^A)(\mathbf{h}^{\mathrm{A\ in\ B}} - \mathbf{h})]$$

It is straightforward to confirm that Equation 4 reduces to Equation 1 in the case of exact self-embedding, i.e., that A and B are treated at the same level of theory and μ is large enough to ensure that $\{\phi_i^A\} \perp \{\phi_i^B\}$. For the purposes of this paper, where we would like A to be treated at some WFT level, the last term in Equation 4 corrects for the fact that $\boldsymbol{\gamma}_{\mathrm{emb}}^A \neq \boldsymbol{\gamma}^A$. Analogously, embedding a wave function in A within a DFT potential provided by B follows from Equation 4:

$$E[\Psi^A; \boldsymbol{\gamma}^A, \boldsymbol{\gamma}^B] = \langle \Psi^A | \hat{H}^{\mathrm{A\ in\ B}} | \Psi^A \rangle$$
$$+ E[\boldsymbol{\gamma}^B] + \mathbf{g}[\boldsymbol{\gamma}^A, \boldsymbol{\gamma}^B] \qquad (5)$$
$$- \mathrm{tr}[\boldsymbol{\gamma}^A(\mathbf{h}^{\mathrm{A\ in\ B}} - \mathbf{h})]$$

B. Singular value decomposed orbitals

Given that there are well-established limits in the formalism shown above, i.e., the self-embedding case and the exact orthogonalization of orbitals in A and B, the concern regarding the performance of this approach shifts to the construction of the orbital subspaces. First, the orbitals should be, to a great extent, spatially restricted to the atoms that constitute each subsystem, and for that, using a localization scheme tends to be the natural choice, such as the Ruedenberg-Edminston,[35] Boys,[36] Pipek-Mezey,[22] and intrinsic bonding orbitals.[30] We propose a different route. Starting from the expansion of the occupied MOs in terms of the AO basis, one has $\phi = \boldsymbol{\chi}\mathbf{C}_{\mathrm{occ}}$, where $\boldsymbol{\chi}$ represents the AO basis of choice. Alternatively, one can work with orthogonalized AOs, which can be accomplished by turning to $\bar{\mathbf{C}}_{\mathrm{occ}} = \mathbf{S}^{1/2}\mathbf{C}_{\mathrm{occ}}$. In this basis, we define the following projector $\hat{Q}_A$:

$$\hat{Q}_A = \sum_{\bar{\mu} \in A} |\bar{\mu}\rangle\langle\bar{\mu}| \qquad (6)$$

where $\{\bar{\mu}\}$ is the set of orthogonal AOs associated with atoms in subsystem A. It is simple to verify that the orthogonal complement of $\hat{Q}_A$ is $\hat{Q}_B$ and that these projectors are idempotent. In practical terms, we split $\bar{\mathbf{C}}_{\mathrm{occ}}$ into matrices that are comprised of only the AOs belonging to the centers in each subsystem. By letting $\bar{\mathbf{C}}_{\mathrm{occ}}^A$ be the MO coefficients from AOs in subsystem A, we have that:

$$\hat{Q}_A = \sum_{\bar{\mu} \in A} |\bar{\mu}\rangle\langle\bar{\mu}| \Rightarrow \mathbf{Q}_A\bar{\mathbf{C}}_{\mathrm{occ}} = \bar{\mathbf{C}}_{\mathrm{occ}}^A \qquad (7)$$

New orbitals can obtained by rotating the original MOs $\mathbf{C}_{\mathrm{occ}}$ according to the right-singular vectors of $\bar{\mathbf{C}}_{\mathrm{occ}}^A$, with the corresponding singular values signaling the relative importance of their associated vectors. The singular value decomposition (SVD) of $\bar{\mathbf{C}}_{\mathrm{occ}}^A$ is given by:

$$\bar{\mathbf{C}}_{\mathrm{occ}}^A = \mathbf{U}^A \boldsymbol{\Sigma}^A \mathbf{V}^{*A}. \qquad (8)$$

The column vectors in $\mathbf{V}^A$ are identical to the eigenvectors of $\bar{\mathbf{C}}_{\mathrm{occ}}^{\dagger A} \bar{\mathbf{C}}_{\mathrm{occ}}^A$. A new set of orbitals are then obtained through a rotation of the original MOs, $\mathbf{C}_{\mathrm{occ}}$. We call this procedure **S**ubsystem **P**rojected **A**O **DE**composition, and will refer to it by the acronym SPADE hereafter, with the orbitals it generates named SPADE orbitals. Thus:

$$\mathbf{C}_{\mathrm{occ}}^{\mathrm{SPADE}} = \mathbf{C}_{\mathrm{occ}}\mathbf{V}^A \qquad (9)$$

Using only an appropriate partition of the atoms into active (A) and environment (B) subsystems, the corresponding orbital subspaces ensue from the distribution of singular values. The largest difference between successive singular value signals the most adequate partition of the set of MOs. To illustrate this point, we take ethanol as an example. Upon selection of the OH- group as the active subsystem, the active orbital subspace is expected to be spanned by five orbitals, in accordance with Ref. 17, where PM orbitals and 0.4 threshold in the Mulliken charge contribution from the oxygen and hydrogen atoms in A were used. Application of the procedure above yields a set of singular values $\{\sigma_i\}$, whose squares can be seen as effective occupation numbers, with i being an orbital label. Plotting $\{\sigma_i^2\}$ and the singular value difference $\Delta\sigma_i^2 = \sigma_i^2 - \sigma_{i+1}^2$ results in Figure 1.

It is apparent that singular value decomposing the MO coefficients in accordance with Equation 8 produces a singular value distribution in which the largest difference $\sigma_i^2 - \sigma_{i+1}^2$ agrees with the atomic partition designated initially, so the active orbital subspace should contain five orbitals. These orbitals are plotted side-by-side in Figure 2 with the five Pipek-Mezey orbitals whose Mulliken charge contribution from the OH- group exceeds 0.4.

The attractiveness of turning to the SPADE orbitals as shown above is two-fold. First, due to the square of the singular values being a quantity analogous to occupation numbers, they can be seen as similar to the natural population analysis, which has been proven capable of avoiding (some of) the flaws of the Mulliken analysis. Second, it follows from the fact that $\bar{\mathbf{C}}_{\mathrm{occ}}^A$ only contains the coefficients from atoms in A that the rotated orbitals will be mostly constrained to their corresponding subsystems. The SPADE procedure automatically "localizes" the orbitals, but contrary to most localization schemes that tend to localize the orbitals in atoms or bonds, the SPADE orbitals are local only in the sense that they remain in their native subsystems, which is one of the requirements for successful embedding.
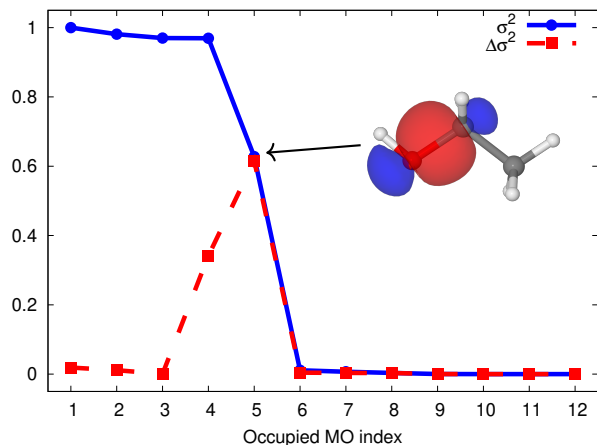
FIG. 1: Distribution of the square of the singular values (σ^2) from $\boldsymbol{\Sigma}^A$ in Equation 8 and $\Delta\sigma_i^2 = \sigma_i^2 - \sigma_{i+1}^2$ from a calculation on ethanol with the OH- group as the active subsystem at the PBE/6-31G* level of theory. The orbital depicted is the SPADE orbital preceding the largest gap, which connects the active and the environment subsystems.

C. Physical rationale for singular value partitioning

In the previous section, the decision to partition the orbital space at the point of largest gap in the singular value spectrum was largely empirical, in that it agreed with chemical intuition. However, one can view this in another, perhaps more physical way. Performing an SVD on the projected occupied MO vectors is identical to computing the eigenvectors of the one particle density matrix projected onto the same AO space. As a single determinant state, before projection this density matrix is idempotent, with all eigenvalues either equal to 1 or 0. However, after projection onto a subsystem, the resulting density matrix is non-idempotent, with occupation numbers (eigenvalues of projected density matrix) distributed in between 1 and 0. This non-idempotency tells us that the active subsystem is (statically) correlated with the rest of the system. In fact, if the active system and environment were separated to infinite distance, this projector would not mix the occupied and virtual spaces, leaving the projected density matrix idempotent.

While electron correlation is one mechanism by which a density matrix can become non-idempotent, finite temperature effects can also have this response, and the use of finite temperature in non-interacting systems to approximate static correlation has been explored successfully in a number of works such as the TAO-DFT method.[37] If we assume that the non-idempotent projected density can be associated with the ground state of a smaller non-interacting system at finite temperature, then the distribution of electron occupations (n_i) should follow the Fermi-Dirac distribution:

$$n_i = \frac{1}{1 + e^{(e_i - \mu)/k_B T}}. \qquad (10)$$

The point on this sigmoid distribution at which the second derivative $\partial^2 n_i / \partial e_i^2$ goes to zero, is the Fermi level. By choosing the largest gap in the eigenvalue spectrum of the projected density matrix, we are essentially choosing to partition our system at the Fermi level of the finite temperature non-interacting system. Under this model, an active system selection which produces a relatively "low temperature" spectrum should be expected to yield good embedding results, as it is an indication of only weak correlation between the active subsystem and environment. However, this rationale is contingent upon the spectrum being sigmoidal, and poor active subsystem selection could possibly yield a spectrum which does not follow this behavior, resulting in a poor partitioning.

III. COMPUTATIONAL DETAILS

The embedding potential that is added to the core Hamiltonian in order to obtain $\mathbf{h}^{A \text{ in } B}$ is constructed with the aid of Numpy, taking advantage of the Python API in the Psi4 quantum chemical package.[38] A local version of Psi4 was modified in order to allow for the required implementation of the effective Fock matrices. The procedure for construction of the SPADE orbitals is based on the linear algebra module of NumPy.

The effect of the choice of density functional is tested by carrying out calculations with a pure and a hybrid functional, and this is accomplished at the low level of theory with the PBE[39,40] and the B3LYP functionals,[41,42] respectively. The WFT methods used as high level of theory in the embedded portion of the calculations were chosen to be the Møller-Plesset perturbation theory to second-order (MP2), coupled cluster single and double excitations (CCSD),[43] and CCSD with perturbative triple excitation correction through fourth-order (CCSD(T)).[44–46] The basis sets used in this work were the augmented correlation consistent polarized valence double zeta (aug-cc-pVDZ)[47,48] and Pople's 6-31G*[49] and the latter with addition of diffuse functions, 6-31+G*. Calculations where the electron repulsion integrals are approximated by density-fitting[50–53] are expanded in the cc-pVDZ-JKFIT[54] (SCF) and cc-pVDZ-RI[55,56] (MP2) auxiliary basis sets.

IV. RESULTS AND DISCUSSION

A. Reaction energies

Instead of examining the capabilities of projection-based embedding with orbitals originated from the SPADE prescription put forth in Section II in one or few selected cases, we start investigating the performance of the SPADE approach by presenting results from embedding calculations carried out for the molecules in the test set of reactions suggested by Mayhall and Raghavachari.[57] We regard this as a balanced set that
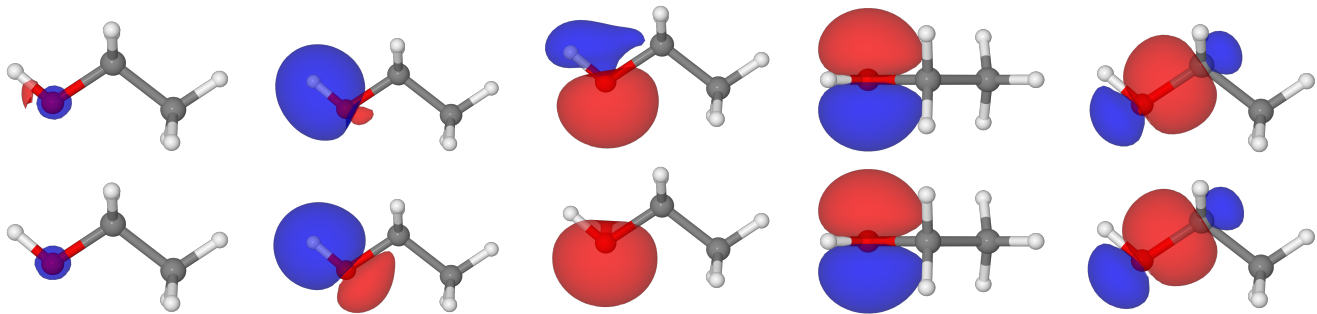
FIG. 2: The five SPADE (top) and PM (bottom) embedded orbitals of ethanol when the active subsystem is the OH- group at the PBE/6-31G* level of theory.

samples different situations, with the active subsystem undergoing chemically distinct reactions, involving electron donor and acceptor environment subsystems as well as closed and open shell configurations. The reactions considered in this paper are summarized in Table I.

TABLE I: Test set of reactions (Ref. 57). X = F, CH$_3$, and ‖ denotes the partition into active (right) and environment (left).

Deprotonation	
1)	X_3C-‖-$CH_2OH_2^+ \rightarrow X_3C$-‖-$CH_2OH$ + H^+
2)	X_3C-‖-$CH_2OH \rightarrow X_3C$-‖-CH_2O^- + H^+
3)	X_3C-‖-$CH_2NH_3^+ \rightarrow X_3C$-‖-$CH_2NH_2$ + H^+
4)	X_3C-‖-$CH_2NH_2 \rightarrow X_3C$-‖-$CH_2NH^-$ + H^+
5)	X_3C-‖-$COOH \rightarrow X_3C$-‖-COO^- + H^+

H-abstraction	
6)	X_3C-‖-$CH_2OH \rightarrow X_3C$-‖-$CH_2O^\bullet$ + $H^\bullet$
7)	X_3C-‖-$CH_2NH_2 \rightarrow X_3C$-‖-$CH_2NH^\bullet$ + $H^\bullet$

Ionization energy	
8)	X_3C-‖-$CH_2O^- \rightarrow X_3C$-‖-$CH_2O^\bullet$ + e^-
9)	X_3C-‖-$CH_2NH^- \rightarrow X_3C$-‖-$CH_2NH^\bullet$ + e^-

S_N2	
10)	X_3C-‖-$CH_2F + Cl^- \rightarrow X_3C$-‖-$CH_2Cl + F^-$

The comparison of the reaction energies for the reactions in Table I obtained from embedded calculations carried out with a certain WFT method are compared against the calculations at the same levels of theory without embedding using our SPADE scheme and the PM orbitals selected according to the Mulliken charge cut-off, referred to as PM/Mulliken hereafter. The results are shown in Figure 3.

The general trend when comparing different WFT approaches for a given substituent X (X = F, CH$_3$) is that the overall error profile changes little regardless of the higher level of theory employed. In other words, it means that even though the error is progressively mitigated as the embedded calculations are performed at MP2, CCSD, and CCSD(T), the general qualitative picture of the errors emerging from embedding remains fairly similar. This implies that, since the only difference between the SPADE and the PM/Mulliken embedded calculations is the choice of the orbital partition, the discrepancies between these two schemes for a given reaction can be attributed to the construction of the orbital subspaces and the mean-field characterization these orbitals provided in the embedded calculations.

Albeit the two orbital selection schemes comparing well in most instances, some discrepancies require further explanation. In the case of deprotonation reactions, the departure of the proton should not change the size of the embedded orbital subspace, that is, the embedded Brønsted-Lowry acid and its conjugate base are expected to have partitioned orbital spaces of the same size. It turns out that, while the PM localization produces the anticipated nine orbitals with Mulliken charge over 0.4 from the active atoms for X_3C-CH_2OH, the deprotonated species, X_3C-CH_2O^-, has only eight PM orbitals that meet this criterion, and this is the reason for the much larger error in reactions 2 and 8 when the PM/Mulliken orbitals are employed.

A known deficiency in the concept of Mulliken charges is revealed when they are computed in a basis containing diffuse functions, leading to unphysical charges. The PM orbitals do, in turn, rely on Mulliken charges, thus they may also be plagued by this drawback. In this context, it is important to note that the calculations whose results are reported in this section were carried out using the aug-cc-pVDZ basis set, where "aug" denotes augmentation of the cc-pVDZ set with diffuse Gaussian primitives. It happens that, in the case of X_3C-CH_2NH^-, there is at least one orbital generated by the PM localization procedure that does not comply with the idea of a local orbital in the aug-cc-pVDZ basis set, which is shown in Figure 4.
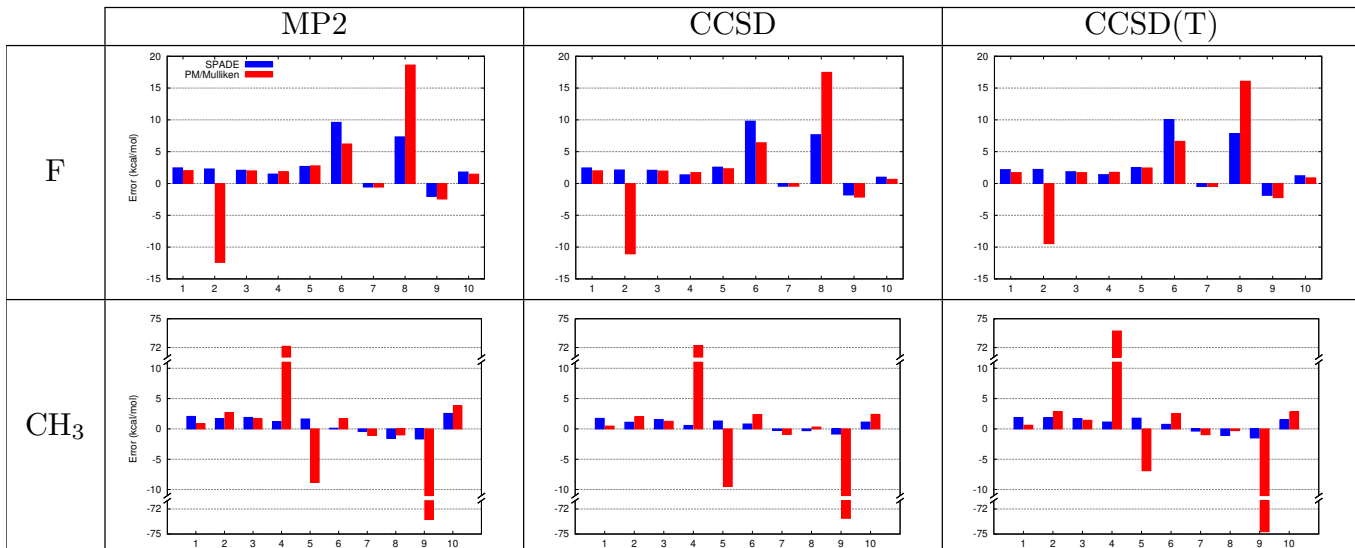
FIG. 3: Errors from the full energies at the high level of theory for the reactions in Table I using PBE/aug-cc-pVDZ as the low level of theory.
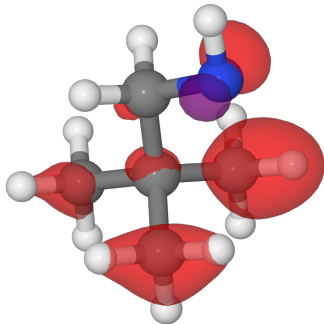

FIG. 4: One of the PM orbitals in $(CH_3)_3C$-CH_2NH^- that does not localize properly at the PBE/aug-cc-pVDZ level of theory.

It is clear that the orbital portrayed in Figure 4 does not localize on any particular atoms or bonds, but it significantly delocalizes over the entire anion. Due to the fact that the test set contains anions, it is strongly recommended that diffuse basis sets be used for a proper treatment of such species. It is well-known that the Mulliken charges do not have a defined basis set limit and can be very sensitive to the choice of basis set, which seems to be the case here. Since this anion also participates in reaction 9, that explains why the errors in these reactions with the PM/Mulliken scheme are so different from those with SPADE orbitals. Even though our method shares with the PM/Mulliken scheme the inability to provide a defined basis set limit, it seems much less prone to suffer from its potential deleterious effects.

B. Convergence with respect to the size of the active subsystem

If the embedding of a subsystem into another is successfully performed, it should be expected that the total energy of the embedded calculation would approach the energy of the entire system computed at the higher level of theory as the size of the active subsystem is increased. In order to investigate the convergence of the energy for different molecular partitions, we chose the deprotonation of the decanoic acid, as reported in Refs. 31 and 34, and shown in Fig. 5. The smallest embedded calculation is performed with an active subsystem consisting of the carboxyl/carboxylate functional groups. The active subsystem is subsequently extended, one neighboring carbon at time, where the hydrogen atoms directly bonded to the a certain carbon atom are also included in the subsystem to which the carbon atom in question is assigned.

Since one of the main conclusions of Subsection IV A is that the main source of error in the current type of embedding is the mean-field treatment in the presence of the embedding potential, rendering the errors in correlation energy from MP2, CCSD, CCSD(T) in close resemblance, analysis of the convergence behavior of the MP2-in-PBE energy should suffice to elucidate the efectiveness of a certain embedding approach. Figure 5 displays the convergence of the deprotonation energy at the MP2 level with respect to the size of the active subsystem, comparing the performance of the PM scheme with the proposed SPADE orbitals. We also assess the effect of the basis set on the quality of the embedded calculations, which is realized by reporting the convergence toward the MP2 limit with the 6-31+G* and aug-cc-pVDZ basis sets.

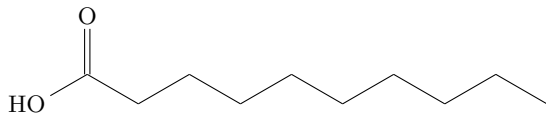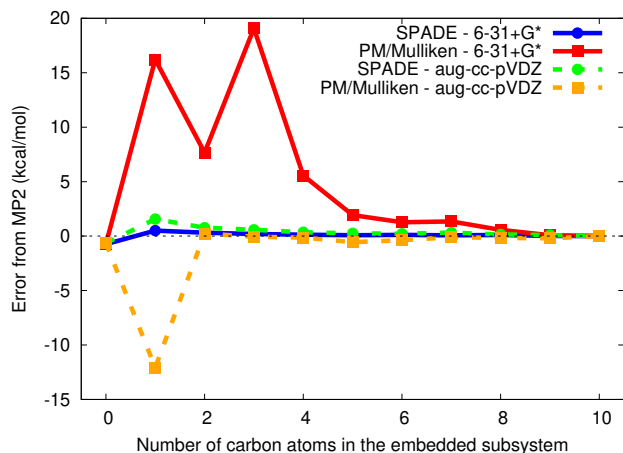Firstly, the embedded calculations with 0 and 10 carbon atoms refer to the limits in the absence of embedding,

FIG. 5: Errors of MP2-in-PBE from the full MP2 energies in the deprotonation of decanoic acid using the PM/Mulliken and the SPADE approaches.

FIG. 6: MP2-in-PBE Oxygen 1s orbital energies in the decanoate anion as a function of the size of the embedded subsystem.

with the former being the entire system computed with the PBE functional (low level), while the latter is the analogous calculation at the MP2 level of theory (high level). Even though the PM/Mulliken results display the expected behavior for larger embedded subsystems, we can see that this scheme seems to be much more sensitive to the employed basis set, which is clear for active subsystems with fewer carbon atoms. The decanoate anion has one fewer orbital than the decanoic acid in the cases where the 6-31+G* basis set is used and one or two carbon atoms are embedded, and also for the embedded calculation with one embedded carbon atom with the aug-cc-pVDZ basis set. The erratic convergence shown by the PM/Mulliken approach in conjunction with the 6-31+G* basis is dramatically reduced by changing to aug-cc-pVDZ after the second embedded carbon.

The source of the distinct convergence pattern between the two basis sets under the PM/Mulliken approach seems to be related to the ability of the PM localization scheme to properly define local orbitals that can be identified with the 1s orbital of the oxygen atoms in the decanoate anion. We tried changing the level shift parameter to 10^5 and 10^7, without noticable change. A plot of the canonical embedded orbital energies referring to this orbital from the four combinations of basis sets and orbital localization/partition is provided in Figure 6.

Apart from the discrepancy seen for the embedded subsystem comprised of only the carboxylate (one carbon), the PM O1s orbital energy from the aug-cc-pVDZ basis set readily resembles those in the SPADE embedded calculations, which can also be attributed to the improper density formed by a smaller orbital subset, as the Mul-
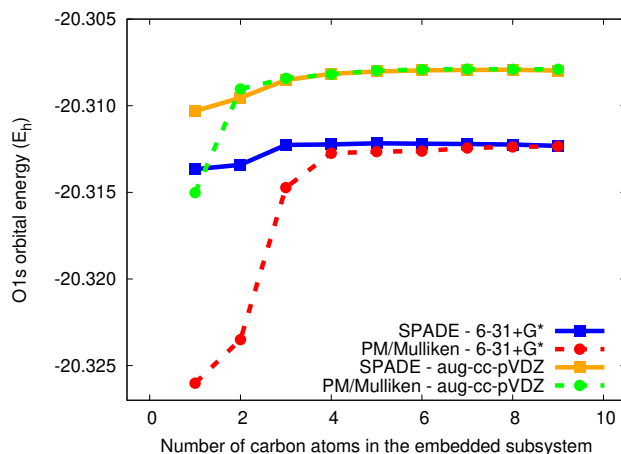
liken charge cut-off leads to the decanoate anion being one orbital short in the embedding step in comparison to its conjugate acid. This inconsistency is exacerbated if the 6-31G+ is used instead, whose O1s orbitals converge to the same values as the SPADE approach rather later in the process of increasing the size of the active subsystem.

The results displayed in Figure 6 only partially explain the oscillatory convergence observed in the embedded calculation of smaller active subsystems with the PM/Mulliken combination presented in Figure 5. The final form of these plots depends on error cancelation from multiple sources. The energies computed with the PBE functional consistently overshoot those at the MP2 level of theory, that is, the more orbitals are embedded, the smaller (less negative) the final embedded energy, and vice-versa. So, in the cases where the decanoate anion has one fewer embedded orbital than the decanoic acid, which, in the case of 6-31+G* basis set, is found when we embed one and two carbons, the error cancellation actually plays in favor of a smaller overall error. It turns out that manually adding the missing orbital in both cases leads to much larger errors. This is also the reason why the one carbon embedded system with the aug-cc-pVDZ basis set underestimates the deprotonation energy of the decanoic acid.

The complications discussed above arising from the utilization of the PM/Mulliken approach are greatly mitigated by turning to SPADE orbital partition. Not only the effect of the basis set is much less pronounced, which is a desirable feature, but the relative size of the partitioned orbital subspaces remain the same for both the decanoic acid and its conjugate base as the size of the embedded subsystem grows. This scheme is responsible for errors that are very small, even for only a single carbon in the active subsystem, which are also noticeably comparable between the two basis sets being tested.
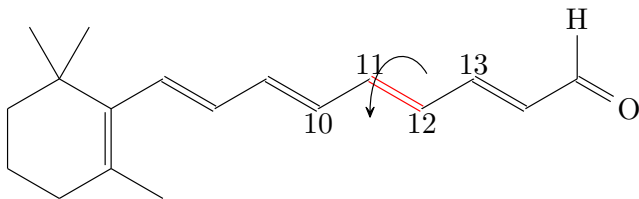
FIG. 7: Structure of the trans isomer of the retinal derivative. The C11-C12 bond is highlighted in red and the relevant carbon atoms are numbered.

C. Torsional potential of a retinal derivative

One of the biochemical compounds that has enjoyed most popularity and has been devoted much attention is the rhodopsin protein due to its role as the pigment responsible for vision. More precisely, its biological function is primarily associated with the retinal cofactor, which has served as model system for the investigation of the rhodopsin activity.[58–60] A central step in the vision mechanism is the absorption of a photon in the visible range by rhodopsin, after which the retinal chromophore undergoes a cis-trans isomerization by a twist of the C11-C12 bond. This process is emulated with the chosen retinal derivative as depicted in Figure 7.

The examination of torsional potentials may pose difficulties to embedding theories, as the orbitals may exhibit dramatic changes followed by the torsion along a bond, leading to discontinuities in what would otherwise be a smooth potential energy curve. We first show in Figure 8 the torsional potential of the retinal derivative at four levels of theory: Hartree-Fock (HF), MP2, and the functionals PBE and B3LYP.
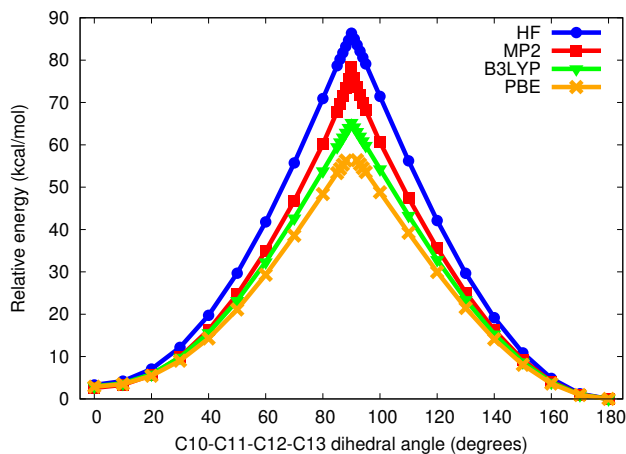

FIG. 8: Torsional potential of the retinal derivative as function of the C10-C11-C12-C13 dihedral angle computed with (non-embedded) HF, MP2, PBE, and B3LYP with the 6-31G* basis set. Energies shifted with respect to the trans isomer.

Figure 8 shows the expected smooth evolution of the torsional potential as the dihedral angle in question is varied. The discontinuity in the PBE potential refers to

the lack of convergence of the underlying SCF procedure for dihedral angles of 89°-91°. That means that this discontinuity will also be present in the torsional potential from embedded calculations where PBE orbitals are used to construct the embedding potential for the posterior WFT treatment. This drawback is satisfactorily sorted out by turning to the hybrid functional B3LYP.

In the embedded calculations for the torsional potential of the retinal derivative, we choose the active subsystem as the four carbons directly associated with the twist of the bond, that is, the C10, C11, C12, and C13 atoms, along with the four hydrogens bonded to them. This case also presents a new situation in comparison with presented so far, since the embedded subsystem is found between two regions that serve as the environment. Figure 9 shows the four combinations of the construction of the orbital subspaces, SPADE and PM/Mulliken, with the functional used to generate the embedding potential, PBE and B3LYP.
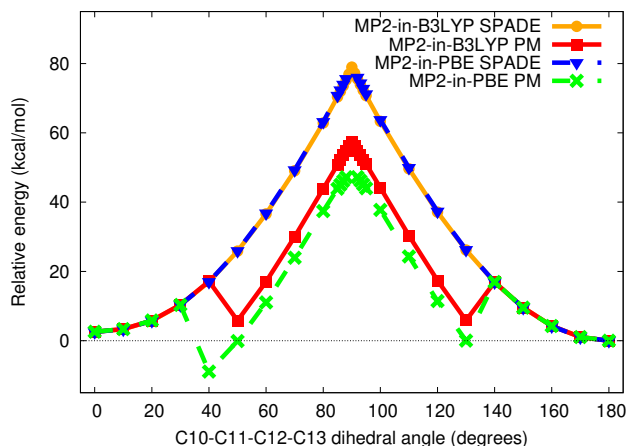

FIG. 9: Torsional potential from the four combinations of the two strategies for the construction of the orbital subspace (SPADE and PM/Mulliken, referred to as PM) and the PBE and B3LYP functionals.

The most important feature of the embedded torsional potential of the retinal derivative as provided in Figure 9 is that the threshold of 0.4 in Mulliken charges in not suitable for the successful application of this approach to the current system. Both SPADE and PM/Mulliken schemes active orbital subspaces consist of 16 embedded orbitals in the cis and trans conformations. As the dihedral angle departs from either end, the active orbital subspace partitioned according to the PM/Mulliken prescription eventually reaches a point where it "loses" one orbitals to the environment, that is, the associated Mulliken charge contribution from one of the atoms falls below 0.4. Specifically, the Pipek-Mezey orbital whose density is found mainly on C13, C14, and C15 draws Mulliken charge from C13, which is part of the active subsystem, and whose contribution is found below 0.4 between 50° and 130° (B3LYP) or 140° (PBE). This drop in Mulliken charge is responsible for the discontinuities

observed in the corresponding potential curves, trait not present in the potential profiles deriving from embedded SPADE orbitals. To better illustrate the variation of the Mulliken charge contribution from C13 to the PM orbital in question, we plot its MP2-in-B3LYP torsional potential, along with the abovementioned Mulliken charge in Figure 10.
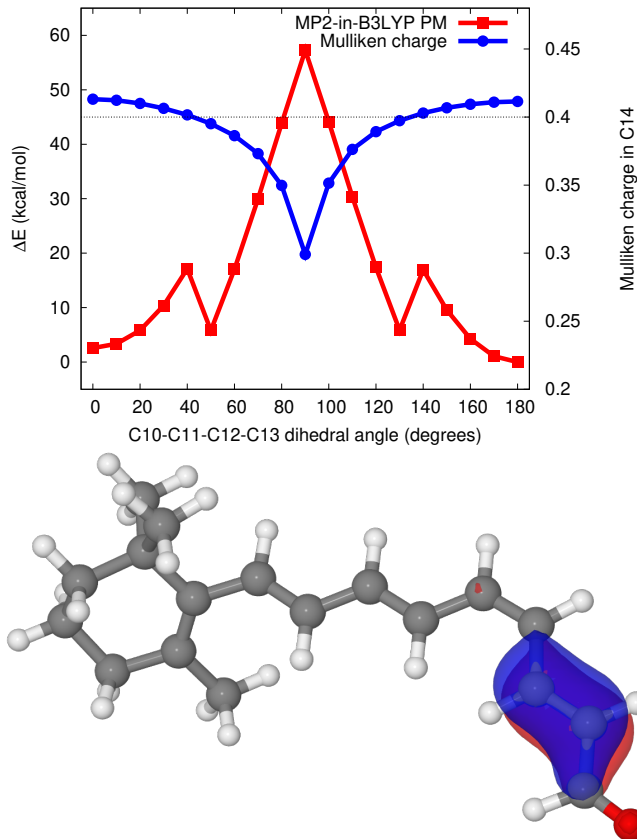

FIG. 10: MP2-in-B3LYP torsional potential using the PM/Mulliken scheme and the evolution of the Mulliken charge contribution from C13 in the orbital depicted above (cis isomer).

From the analysis of Figure 10, it is clear that the discontinuities seen in Figure 9 coincide with the Mulliken charge from C13 falling below the set threshold of 0.4. In turn, our criterion, by virtue of being free of a fixed numerical parameter, does not encounter such a problem. It does not mean that the largest difference between the squares of successive singular values remains constant, but rather that this difference refers to the same two orbitals throughout the rotation along the C11-C12 bond. It can be argued that admitting a more relaxed Mulliken charge cut-off could solve this discrepancy, but this new value would remain arbitrary in the sense that there would be no guarantee that it would be suitable for other problems.

The two functionals tested here, PBE and B3LYP, have significantly distinct natures. While PBE is a pure functional, B3LYP accounts for exchange in a hybrid fashion, that is, 20% of the exchange potential stems from Hartree-Fock, percentage which is determined in a empirical manner. Thus, another important conclusion to be drawn from Figure 9 is that, for the entirety of the SPADE curves and where the PM/Mulliken orbital spaces agree with the SPADE ones, the effect of the intrinsic character of the functional providing the embedding potential does not play a major role.

One of the driving arguments behind the development of multilevel approaches such as embedding methods is to reduce the overall computational expense of dealing with the entire system at the high level of theory. Substantial savings can be attained if the electron repulsion integrals (ERI) are computed via density-fitting. The error due to the use of density-fitted ERIs and the error in the MP2-in-B3LYP potential from the full MP2 results are plotted in Figure 11.
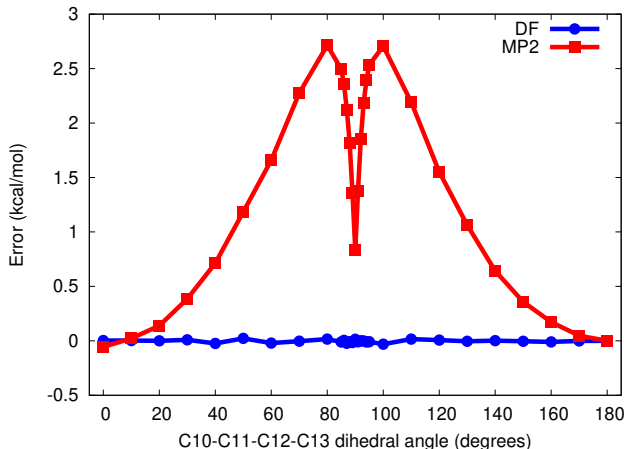

FIG. 11: Errors in the MP2-in-B3LYP energies from the full MP2 results (MP2) and errors of embedding using density-fitted ERIs versus exact ERIs (DF).

It can be concluded from Figure 11 that one can safely turn to density-fitting as a viable and reliable alternative to the commonly costly computation of ERIs by more traditional means. The error from the corresponding full MP2 energies are maximum in about 10° from the maximum in the torsional potential in either direction, and fall as the dihedral angle approaches 90°. Given the size of the retinal derivative, the relatively small size of the embedded subsystem, and the difficulty of the underlying problem, the error from the MP2 energies remaining below 3 kcal/mol throughout the entire range of torsion angles signals the aptitude of the proposed scheme for reliable estimation of torsional potentials.

We would like to acknowledge that, at the closing of this manuscript, it came to our knowledge that Manby and Miller recently introduced a different approach aimed at overcoming the mismatch in the orbital subspaces that gives rise to the discontinuities throughout a potential energy surface, such as those seen in the torsional potential

of the retinal derivative when the PM/Mulliken scheme is employed, as shown in Figure 9.[61]

D. Menshutkin S_N2 reaction inside a carbon nanotube

As a last example, we studied the Menshutkin S_N2 reaction involving ammonia and chloromethane, leading to the methylammonium and chloride ions in gas phase (GP) and inside a carbon nanotube (CNT). It has been proposed that the energy barrier of this reaction can be made lower relative to the gas phase pathway if the reactive system is found inside a CNT, much like the observed effect of a low-dielectric solvent.

The reactive system by itself in gas-phase (GP) is expected to pose no computational challenges, but when considered inside the CNT, it turns into a sizable calculation. It is comprised of 852 electrons in 165 atoms, and with the 6-31G* basis set, there are 2089 basis functions in total. The geometries of the critical points along the reaction coordinate that we consider here, namely the reactants, products, and transition state (TS) are obtained from Song and Martinez.[62]

In Ref. 62, GP and CNT energy profiles for this reaction were presented, with energies shifted with respect to the reactants, at the CASPT2(4,4)/6-31G* level of theory. First, we carried out the same calculations, but at the MP2/6-31G* level of theory, and found no appreciable discrepancy between the energy differences of these methods. This can be interpreted as evidence that the reactive system can, in both chemical environments, be adequately treated by a single reference determinant. Even though our current approach poses no limitation to using such multireference methods, this allows for embedding calculations to be done where the embedded region can be treated at the MP2 level instead of the more costly multireference analog.

In the embedding calculations for this system, we assigned the active subsystem to be the reactive pair, i.e., NH_3+CH_3Cl, while the environment is comprised of the CNT structure. This translates into 18 doubly occupied embedded MOs being carried over to the correlated step. This means that the mean-field calculation, in this case DFT using the B3LYP functional, exceeds the WFT part in computational demand. The results from the embedding calculations, along with pure B3LYP and MP2 results and the numbers reported by Song and Martinez are presented in Figure 12.

First, as stated above, although this transition state involves some static correlation, the results indicate that this does yet not spoil the accuracy of single reference MP2, which is found to be comparable to CASPT2(4,4) level results for both the GP and CNT reactions. In the case of the GP B3LYP energy profile, even the qualitative picture is incorrect, with the energy computed at the given TS geometry being lower than that of the product. The correct energy ordering is recovered in the presence of the CNT, but the B3LYP energy differences are not to be trusted quantitatively when taking the CASPT2(4, 4)/MP2 results as reference.

Even though B3LYP does not seem to be adequate by itself, it is able to provide a reliable embedding potential. From Figure 12, it is evident that the main characteristics of the reaction pathway are preserved in the embedding picture, but can be achieved by much more affordable means. With only 18 out of 426 occupied orbitals being correlated, the embedded TS energies, where the largest deviations from the CASPT2(4,4)/MP2 reference are found, exhibit remarkable accuracy. The SPADE result is 0.1 kcal/mol and 0.9 kcal/mol higher in energy than the CASPT2(4,4) and MP2 numbers, respectively, while the same differences are 0.9 kcal/mol and 1.7 kcal/mol with the PM/Mulliken scheme.

In conclusion, the projection-based embedding theory is able to correctly and quantitatively predict the effect of the CNT to the energy profile of the reaction in question, more precisely the lowering of the energy barrier. The SPADE approach shows marginal advantage over the PM/Mulliken scheme, with energies within chemical accuracy from the reference numbers.

V. CONCLUSION

Projection-based embedding theory has consistently proven itself a promising and reliable multilvel alternative to more costly calculations, without renouncing accuracy. One of its potential disadvantages until recently was the dependence on the concept of Mulliken charges for orbital localization and somewhat arbitrary cut-off values of this quantity for the purpose of partition of the associated orbital subspaces. According to the numerical examples explored so far, it appears as though this SPADE approach provides significant improvements on this front.

The SPADE approach both localizes and partitions the orbitals in a black-box fashion, in a manner that only requires the assignment of the atoms to the desired subsystems. And it does so while avoiding arbitrary numerical parameters and potentially faulty quantities.

It is shown that the proposed scheme exhibits remarkable convergence to the high level of theory with respect to the size of the embedded subsystem by means of investigation of the deprotonation of decanoic acid. Moreover, it is also able to correctly address difficult problems, such as the smooth evolution of the torsional potential energy, without displaying discontinuities, which is illustrated by the case of retinal. Lastly, a sign of the power of the projection-based embedding theory becomes apparent by it successfully capturing the effect of the carbon nanotube over the reaction pathway, even when the CNT is completely treated at the lower level of theory, with moderate advantage of the SPADE procedure.

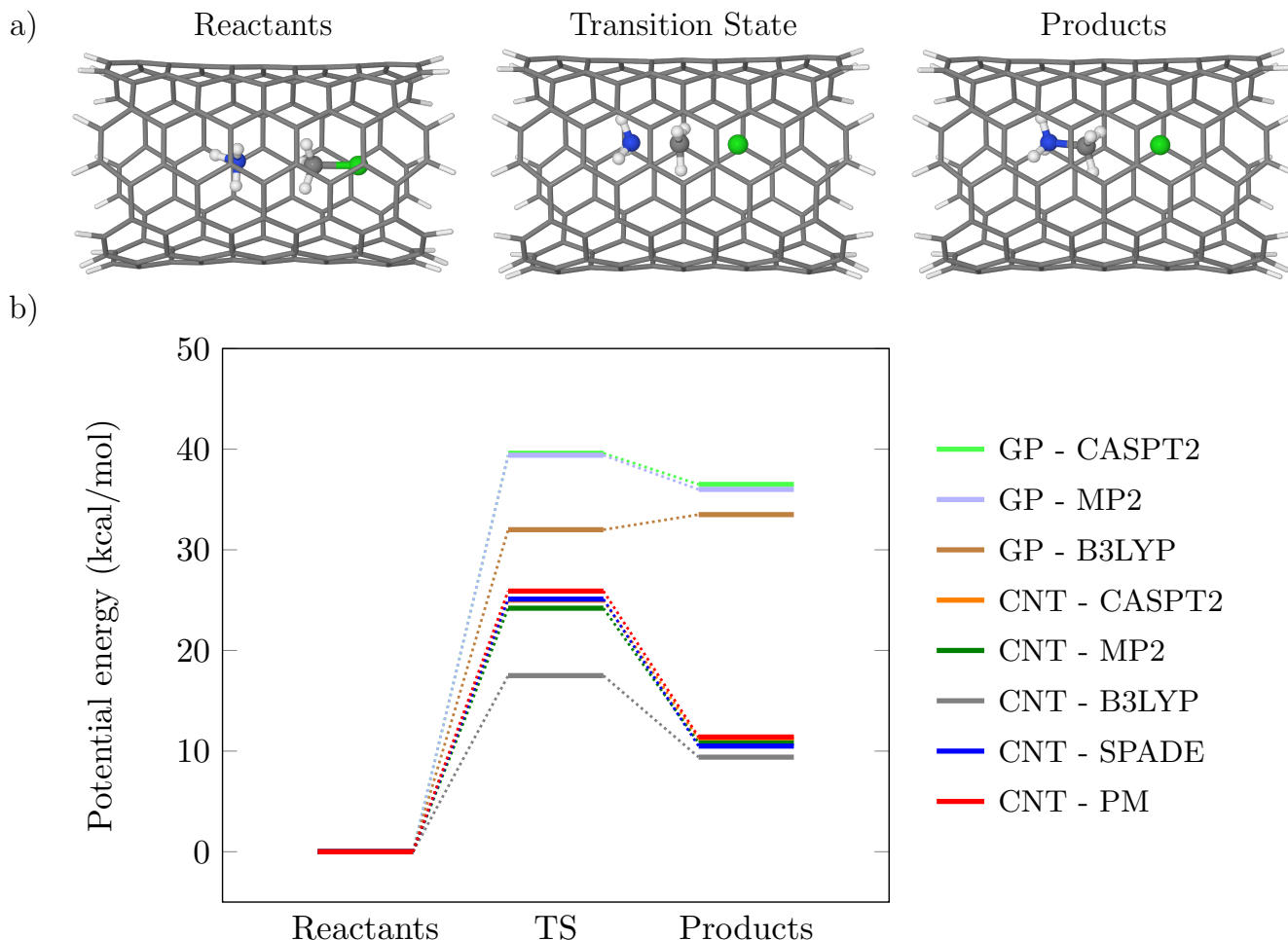Although the results in this study indicate that the SPADE procedure is quite effective, this behavior is not

a)

b)

FIG. 12: The $NH_3 + CH_3Cl \rightarrow H_3NCH_3^+ + Cl^-$ Menshutkin S_N2 reaction. (a) Pictorial representation of the geometries of the reactants, transition state (TS), and products inside the carbon nanotube (CNT). (b) Potential energy diagram in gas phase (GP) in the CNT relative to the energy of the reactants. SPADE and PM are MP2-in-B3LYP calculations using the SPADE and the PM/Mulliken schemes.

always guaranteed. The lack of a rigorous basis set limit may compromise the application of the current strategy to more challenging cases, requiring extremely large basis sets. Furthermore, it is certainly possible that a poor active atom definition might create a singular value spectrum where the number of orbitals in each subsystem change along the course of a reaction. We have been able to observe this effect by choosing only the carbons C11 and C12 in the retinal derivative torsion example, case where the largest gap in the singular value spectrum behaves similarly to the Mulliken charges (Fig. 10). Even though this possibility exists, it is yet to be observed when atoms are adequately assigned, as none of the systems studied with this prescription have fallen into this category, attesting to its robustness.

Future work will focus on extending the idea of SPADE orbitals to the unoccupied MOs, with the goal of effectively truncating the virtual space, thus making WFT-in-DFT embedding even less computationally expensive.

* Electronic address: nmayhall@vt.edu
[1] Warshel, A.; Levitt, M. Theoretical studies of enzymic re-
actions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of*

Molecular Biology **1976**, *103*, 227–249.

[2] Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. 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)$_3$)$_2$ + H$_2$ Oxidative Addition. *The Journal of Physical Chemistry* **1996**, *100*, 19357–19363.

[3] Humbel, S.; Sieber, S.; Morokuma, K. The IMOMO method: Integration of different levels of molecular orbital approximations for geometry optimization of large systems: Test for nbutane conformation and SN2 reaction: RCl+Cl. *The Journal of Chemical Physics* **1998**, *105*, 1959.

[4] Maseras, F.; Morokuma, K. IMOMM: A new integratedab initio + molecular mechanics geometry optimization scheme of equilibrium structures and transition states. *Journal of Computational Chemistry* **1995**, *16*, 1170–1179.

[5] Hratchian, H. P.; Parandekar, P. V.; Raghavachari, K.; Frisch, M. J.; Vreven, T. QM:QM electronic embedding using Mulliken atomic charges: Energies and analytic gradients in an ONIOM framework. *Journal of Chemical Physics* **2008**, *128*, 11.

[6] Mayhall, N. J.; Raghavachari, K.; Hratchian, H. P. ONIOM-based QM:QM electronic embedding method using Löwdin atomic charges: Energies and analytic gradients. *J. Chem. Phys.* **2010**, *132*, 114107.

[7] Lan, T. N.; Kananenka, A. A.; Zgid, D. Communication: Towards ab initio self-energy embedding theory in quantum chemistry. *The Journal of chemical physics* **2015**, *143*, 241102.

[8] Knizia, G.; Chan, G. K.-L. Density Matrix Embedding: A Simple Alternative to Dynamical Mean-Field Theory. *Physical Review Letters* **2012**, *109*, year.

[9] Welborn, M.; Tsuchimochi, T.; Van Voorhis, T. Bootstrap embedding: An internally consistent fragment-based method. *The Journal of Chemical Physics* **2016**, *145*, year.

[10] Wouters, S.; Jiménez-Hoyos, C. A.; Sun, Q.; Chan, G. K.-L. K.-L. A practical guide to density matrix embedding theory in quantum chemistry. *Journal of chemical theory and computation* **2016**, *12*, 2706–2719.

[11] Tsuchimochi, T.; Welborn, M.; Van Voorhis, T. Density matrix embedding in an antisymmetrized geminal power bath. *The Journal of Chemical Physics* **2015**, *143*, 024107.

[12] Wouters, S.; Jiménez-Hoyos, C. A.; Chan, G. K. L. Five years of density matrix embedding theory. *ArXiv preprint* **2016**.

[13] Knizia, G.; Chan, G. K.-L. Density Matrix Embedding: A Strong-Coupling Quantum Embedding Theory. *Journal of Chemical Theory and Computation* **2013**, *9*, 1428–1432.

[14] Goez, A.; Neugebauer, J. Embedding Methods in Quantum Chemistry. In *Frontiers of Quantum Chemistry*; Springer Singapore: Singapore, 2018; pp 139–179.

[15] Wesolowski, T. A.; Shedge, S.; Zhou, X. Frozen-Density Embedding Strategy for Multilevel Simulations of Electronic Structure. *Chemical Reviews* **2015**, *115*, 150429132214004.

[16] Kevorkyants, R.; Wang, X.; Close, D. M.; Pavanello, M. Calculating Hyperfine Couplings in Large Ionic Crystals Containing Hundreds of QM Atoms: Subsystem DFT Is the Key. *The Journal of Physical Chemistry B* **2013**, *117*, 13967–13974.

[17] Manby, F. R.; Stella, M.; Goodpaster, J. D.; Miller, T. F. A Simple, Exact Density-Functional-Theory Embedding Scheme. *Journal of Chemical Theory and Computation* **2012**, *8*, 2564–2568.

[18] Goodpaster, J. D.; Barnes, T. A.; Manby, F. R.; Miller, T. F. Density functional theory embedding for correlated wavefunctions: Improved methods for open-shell systems and transition metal complexes. *The Journal of Chemical Physics* **2012**, *137*, 224113.

[19] Barnes, T. A.; Goodpaster, J. D.; Manby, F. R.; Miller, T. F. Accurate basis set truncation for wavefunction embedding. *The Journal of Chemical Physics* **2013**, *139*, 024103.

[20] Goodpaster, J. D.; Barnes, T. A.; Manby, F. R.; Miller, T. F. Accurate and systematically improvable density functional theory embedding for correlated wavefunctions. *The Journal of Chemical Physics* **2014**, *140*, 18A507.

[21] Bennie, S. J.; Stella, M.; Miller, T. F.; Manby, F. R. Accelerating wavefunction in density-functional-theory embedding by truncating the active basis set. *The Journal of Chemical Physics* **2015**, *143*, 024105.

[22] Pipek, J.; Mezey, P. G. A fast intrinsic localization procedure applicable for *a b i n i t i o* and semiempirical linear combination of atomic orbital wave functions. *The Journal of Chemical Physics* **1989**, *90*, 4916–4926.

[23] Huzinaga, S.; Cantu, A. A. Theory of Separability of ManyElectron Systems. *The Journal of Chemical Physics* **1971**, *55*, 5543–5549.

[24] Barnes, T. A.; Kaminski, J. W.; Borodin, O.; Miller, T. F. Ab Initio Characterization of the Electrochemical Stability and Solvation Properties of Condensed-Phase Ethylene Carbonate and Dimethyl Carbonate Mixtures. *The Journal of Physical Chemistry C* **2015**, *119*, 3865–3880.

[25] Bennie, S. J.; van der Kamp, M. W.; Pennifold, R. C. R.; Stella, M.; Manby, F. R.; Mulholland, A. J. A Projector-Embedding Approach for Multiscale Coupled-Cluster Calculations Applied to Citrate Synthase. *Journal of Chemical Theory and Computation* **2016**, *12*, 2689–2697.

[26] Huo, P.; Uyeda, C.; Goodpaster, J. D.; Peters, J. C.; Miller, T. F. Breaking the Correlation between Energy Costs and Kinetic Barriers in Hydrogen Evolution via a Cobalt Pyridine-Diimine-Dioxime Catalyst. *ACS Catalysis* **2016**, *6*, 6114–6123.

[27] Bennie, S. J.; Curchod, B. F. E.; Manby, F. R.; Glowacki, D. R. Pushing the Limits of EOM-CCSD with Projector-Based Embedding for Excitation Energies. *The Journal of Physical Chemistry Letters* **2017**, *8*, 5559–5565.

[28] Chapovetsky, A.; Welborn, M.; Luna, J. M.; Haiges, R.; Miller, T. F.; Marinescu, S. C. Pendant Hydrogen-Bond Donors in Cobalt Catalysts Independently Enhance CO2 Reduction. *ACS Central Science* **2018**, *4*, 397–404.

[29] Reed, A. E.; Weinstock, R. B.; Weinhold, F. Natural population analysis. *The Journal of Chemical Physics* **1985**, *83*, 735–746.

[30] Knizia, G. Intrinsic Atomic Orbitals: An Unbiased Bridge between Quantum Theory and Chemical Concepts. *Journal of Chemical Theory and Computation* **2013**, *9*, 4834–4843.

[31] Hégely, B.; Nagy, P. R.; Ferenczy, G. G.; Kállay, M. Exact density functional and wave function embedding schemes based on orbital localization. *The Journal of Chemical Physics* **2016**, *145*, 064107.

[32] Boughton, J. W.; Pulay, P. Comparison of the boys and Pipek-Mezey localizations in the local correlation approach and automatic virtual basis selection. *Journal of Compu-*

tational Chemistry **1993**, *14*, 736–740.

[33] Hégely, B.; Nagy, P. R.; Kállay, M. Dual Basis Set Approach for Density Functional and Wave Function Embedding Schemes. *Journal of Chemical Theory and Computation* **2018**, *14*, 4600–4615.

[34] Culpitt, T.; Brorsen, K. R.; Hammes-Schiffer, S. Communication: Density functional theory embedding with the orthogonality constrained basis set expansion procedure. *The Journal of Chemical Physics* **2017**, *146*, 211101.

[35] Kleier, D. A.; Halgren, T. A.; Hall, J. H.; Lipscomb, W. N. Localized molecular orbitals for polyatomic molecules. I. A comparison of the EdmistonRuedenberg and Boys localization methods. *The Journal of Chemical Physics* **1974**, *61*, 3905–3919.

[36] Boys, S. F. Construction of Some Molecular Orbitals to Be Approximately Invariant for Changes from One Molecule to Another. *Reviews of Modern Physics* **1960**, *32*, 296–299.

[37] Chai, J.-D. Density functional theory with fractional orbital occupations. *The Journal of Chemical Physics* **2012**, *136*, 154104.

[38] Parrish, R. M. et al. Psi4 1.1: An Open-Source Electronic Structure Program Emphasizing Automation, Advanced Libraries, and Interoperability. *Journal of Chemical Theory and Computation* **2017**, *13*, 3185–3197.

[39] Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Physical Review Letters* **1996**, *77*, 3865–3868.

[40] Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple [Phys. Rev. Lett. 77, 3865 (1996)]. *Physical Review Letters* **1997**, *78*, 1396–1396.

[41] Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B* **1988**, *37*, 785–789.

[42] Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical Review A* **1988**, *38*, 3098–3100.

[43] Purvis, G. D.; Bartlett, R. J. A full coupledcluster singles and doubles model: The inclusion of disconnected triples. *The Journal of Chemical Physics* **1982**, *76*, 1910–1918.

[44] Urban, M.; Noga, J.; Cole, S. J.; Bartlett, R. J. Towards a full CCSDT model for electron correlation. *The Journal of Chemical Physics* **1985**, *83*, 4041–4046.

[45] Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A fifth-order perturbation comparison of electron correlation theories. *Chemical Physics Letters* **1989**, *157*, 479–483.

[46] Watts, D. Open-shell analytical energy gradients for triple excitation many-body, coupled-clustermethods: MBPT(4), CCSDtT(CCSD), CCSD(T), and QCISD(T). *CHEMICAL PHYSICS LETTERS* **1992**, *200*, 7.

[47] Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *The Journal of Chemical Physics* **1989**, *90*, 1007–1023.

[48] Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron affinities of the firstrow atoms revisited. Systematic basis sets and wave functions. *The Journal of Chemical Physics* **1992**, *96*, 6796–6806.

[49] Hariharan, P. C.; Pople, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theoretica Chimica Acta* **1973**, *28*, 213–222.

[50] Whitten, J. L. Coulombic potential energy integrals and approximations. *The Journal of Chemical Physics* **1973**, *58*, 4496–4501.

[51] Dunlap, B. I.; Connolly, J. W. D.; Sabin, J. R. On some approximations in applications of $X\alpha$ theory. *The Journal of Chemical Physics* **1979**, *71*, 3396–3402.

[52] Feyereisen, M.; Fitzgerald, G.; Komornicki, A. Use of approximate integrals in ab initio theory. An application in MP2 energy calculations. *Chemical Physics Letters* **1993**, *208*, 359–363.

[53] Vahtras, O.; Almlf, J.; Feyereisen, M. Integral approximations for LCAO-SCF calculations. *Chemical Physics Letters* **1993**, *213*, 514–518.

[54] Weigend, F. A fully direct RI-HF algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency. *Physical Chemistry Chemical Physics* **2002**, *4*, 4285–4291.

[55] Weigend, F.; Khn, A.; Httig, C. Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations. *The Journal of Chemical Physics* **2002**, *116*, 3175–3183.

[56] Httig, C. Optimization of auxiliary basis sets for RI-MP2 and RI-CC2 calculations: Corevalence and quintuple-ζ basis sets for H to Ar and QZVPP basis sets for Li to Kr. *Phys. Chem. Chem. Phys.* **2005**, *7*, 59–66.

[57] Mayhall, N. J.; Raghavachari, K. Charge Transfer Across ONIOM QM:QM Boundaries: The Impact of Model System Preparation. *J. Chem. Theory Comput.* **2010**, *6*, 3131–3136.

[58] Honig, B.; Karplus, M. Implications of torsional potential of retinal isomers for visual excitation. *Nature* **1971**, *229*, 558–560.

[59] Warshel, A.; Karplus, M. Calculation of $\pi\pi^*$ excited state conformations and vibronic structure of retinal and related molecules. *Journal of the American Chemical Society* **1974**, *96*, 5677–5689.

[60] Migani, A.; Sinicropi, A.; Ferr, N.; Cembran, A.; Garavelli, M.; Olivucci, M. Structure of the intersection space associated with Z/E photoisomerization of retinal in rhodopsin proteins. *Faraday Discuss.* **2004**, *127*, 179–191.

[61] Welborn, M.; Manby, F. R.; Miller III, T. F. Even-handed subsystem selection in projection-based embedding. *The Journal of Chemical Physics* **2018**, *149*, 144101.

[62] Song, C.; Martínez, T. J. Reduced scaling CASPT2 using supporting subspaces and tensor hyper-contraction. *The Journal of Chemical Physics* **2018**, *149*, 044108.