Projector-Based Quantum Embedding for Molecular Systems: An Investigation of Three Partitioning Approaches

Jonathan M. Waldrop,[*,†] Theresa L. Windus,[*,†,‡] and Niranjan Govind[*,¶]

†*Ames Laboratory, Ames, IA 50011*

‡*Department of Chemistry, Iowa State University, Ames, IA 50011*

¶*Physical and Computational Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA 99352*

E-mail: jwaldrop@ameslab.gov; twindus@iastate.edu; niri.govind@pnnl.gov

Abstract

Projector-based embedding is a relatively recent addition to the collection of methods that seek to utilize chemical locality to provide improved computational efficiency. This work considers the interactions between the different proposed procedures for this method and their effects on the accuracy of the results. The interplay between the embedded background, projector type, partitioning scheme, and level of atomic orbital (AO) truncation are investigated on a selection of reactions from the literature. The Huzinaga projection approach proves to be more reliable than the level-shift projection when paired with other procedural options. Active subsystem partitioning from the subsystem projected AO decomposition (SPADE) procedure proves slightly better than the combination of Pipek-Mezey localization and Mulliken population screening (PMM). Along with these two options, a new partitioning criteria is proposed based on subsystem von Neumann entropy and the related subsystem orbital occupancy. This

new method overlaps with the previous PMM method, but the screening process is computationally simpler. Finally, AO truncation proves to be a robust option for the tested systems when paired with the Huzinaga projection, with satisfactory results being acquired at even the most severe truncation level.

Introduction

Balancing computational cost with high accuracy is a constant consideration for quantum chemical calculations.[1] Full configuration interaction (FCI) is known to be the best non-relativistic answer for a given chemical system in a given basis set, but its prohibitive computational expense prevents its application to all but the smallest of systems.[2] Various wave function theory (WFT) methods like coupled cluster theory[3] and complete active space approaches[4] provide systematically improvable approximations to FCI with improved cost scaling.[1] That said, these methods are still quite costly for large systems of particular interest. Kohn-Sham density functional theory (KS-DFT)[5,6] provides a cost effective approach and is applicable over a broad range of systems. However, the reliance of KS-DFT on the exchange-correlation approximation, despite being reasonably accurate, is not systematically improvable compared with wave function based methods.[7–10]

One route to balancing cost efficiency and computational accuracy can be found by considering the relatively local nature of chemical interactions. Interesting chemical phenomena tend to be localized to smaller active areas of extended electronic systems, e.g., the active site of a protein or a catalytic reaction site in a zeolite. Based on this observation, it is possible to split a larger system into active and environment subsystems and treat them at varying levels of theory. The active subsystem can be tackled with an accurate level of theory that would be prohibitive for the full system, while the environment is treated at a cost efficient level that provides broadly acceptable accuracy on larger systems. Over the years, a variety of schemes have been developed around the concept of embedding subsystems[11–16] at different theoretical levels, including quantum mechanics/molecular mechanics (QM/MM),[17,18]

ONIOM,[19,20] DFT embedding,[12,21,22] partition DFT,[23,24] potential-functional embedding,[25] embedded mean-field theory (EMFT),[26] Green's function embedding,[27,28] self-energy embedding,[29,30] density matrix embedding theory (DMET),[31] and stochastic embedding DFT.[32] In this manuscript, we focus on the projector-based embedding approach.

Projector-based embedding was introduced by Manby, Miller, and co-workers[33] as a formally exact DFT-in-DFT embedding scheme, which can also be applied to WFT-in-DFT embedding or more generally describable as QM-in-QM. The scheme begins with a calculation on the complete system at the level of theory desired for the environment. The orbitals from this calculation are then localized and partitioned into active and environment sets. The embedded calculation on the active set is then performed by modifying the core Hamiltonian to include the potential due to the frozen environment density and a projector operator that ensures that the two sets of orbitals remain orthogonal to each other. More recently, analytical gradients for this method have also been reported.[34]

Some modifications have been proposed for the initial formulation of the embedding scheme, by way of producing a more "black box" approach. Initially, the partitioning of the orbitals was performed by screening Mulliken populations of the localized orbitals against a predetermined empirical parameter, while a level shift projector with another empirical parameter was used to maintain orthogonality between the system partitions. Claudino and Mayhall[35] sought to remove the empirical aspect from the system partitioning by utilizing singular value decomposition (SVD) to produce both the localized orbitals and the partitioning criteria in an automated fashion. As a replacement for the empirical level shift operator, Hegely et al. proposed the Huzinaga level shift operator[36] which requires no empirical parameters.[37]

Other advances in projector-based embedding have focused on further reducing the cost of embedded correlated WFT methods by truncating the atomic orbital (AO) space of the embedded calculation. The first method of AO truncation involved an empirical screening parameter (much like the above mentioned orbital partitioning scheme) and a specification

of atoms bordering the active system.[38] This method was then replaced by a second algorithm that required only the screening parameter.[39] More recently, the method of absolute localization was proposed which combines the Huzinaga projector with truncation of the AO space only to functions centered on the subsystem atoms.[40,41]

Bensberg and Neugebauer studied the interactions between different combinations of projection methods and AO truncation treatment, while also introducing a scheme for extending a "bottom-up" AO space selection.[42] They found that the Huzinaga projector was the most reliable for both total energies and reaction energies when different levels of AO truncation were used. These authors also introduced a method for selecting active orbital sets across a reaction path by screening out orbitals that are determined to be unchanged throughout the reaction.[43] This selection method was then improved by the application of a orbital alignment process, ensuring that like orbitals are localized similarly across the pathway.[44] This orbital selection method is akin to the "even-handed" selection procedure of the Manby and Miller groups,[45] but removes the dependence on the specification of a set of active atoms.

Here we will study the interactions between the various procedural options for projector-embedding and how they effect the accuracy of the resulting calculations. The following section will provide an overview of the theoretical background for projector-based embedding, including partitioning schemes and AO truncation. Along with the previously reported partitioning schemes, we also introduce a new method for active MO selection. The subsequent sections will describe the systems chosen for testing, the computational details of the calculations, and an analysis of the performance of the different combinations of options.

Theory

For completeness, and to provide context for the methods used and the introduction of our new entropy/occupancy based partitioning, we outline the different types of embedding and partitioning currently in use. For further details, we refer the reader to Ref. 33.

Projector-based Embedding

Starting with a chemical system of interest, the total system is partitioned into an active subsystem (A, atoms where the phenomena of interest is expected to be localized) and an environment subsystem (B, the rest of the atoms in the system). The projector-based embedding calculation begins from an initial calculation over the entire system, which is performed at the desired level of theory for the environment. Using reasonable criteria for associating the molecular orbitals (MOs) with either the active or environment subsystems (see below for descriptions of partitioning schemes), the density of the total system ρ is split into subsystem densities associated with the active (ρ^A) and environment (ρ^B) subsystems,

$$\rho = \rho^A + \rho^B. \tag{1}$$

With this partitioning of the density, the energy of the total system can be represented as

$$E(\rho) = \Big(Tr(\boldsymbol{\gamma}^A \cdot \mathbf{h}) + \mathbf{G}[\boldsymbol{\gamma}^A]\Big) + \Big(Tr(\boldsymbol{\gamma}^B \cdot \mathbf{h}) + \mathbf{G}[\boldsymbol{\gamma}^B]\Big) + \delta\mathbf{G}[\boldsymbol{\gamma}^A, \boldsymbol{\gamma}^B], \tag{2}$$

where $\boldsymbol{\gamma}$ is the total system reduced density matrix, $\boldsymbol{\gamma}^X$ is the reduced density matrix of the active of environment subsystem, $\mathbf{h}$ is the core Hamiltonian of the total system, $\mathbf{G}[\boldsymbol{\gamma}^X]$ collects the two-electron interactions for the active or environment subsystem, and $\delta\mathbf{G}[\boldsymbol{\gamma}^A, \boldsymbol{\gamma}^B]$ collects the two-electron interactions between subsystems, i.e.,

$$\mathbf{G}[\boldsymbol{\gamma}] = \mathbf{G}[\boldsymbol{\gamma}^A] + \mathbf{G}[\boldsymbol{\gamma}^B] + \delta\mathbf{G}[\boldsymbol{\gamma}^A, \boldsymbol{\gamma}^B]. \tag{3}$$

The first set of parentheses in Eq. (2) contains the terms that comprise the energy of the active subsystem, while the second set of parentheses contains the terms for the environment,

$$E(\boldsymbol{\gamma}) = E(\boldsymbol{\gamma}^A) + E(\boldsymbol{\gamma}^B) + \delta\mathbf{G}[\boldsymbol{\gamma}^A, \boldsymbol{\gamma}^B]. \tag{4}$$

Subsequent calculations on the active subsystem can be embedded in the environment subsystem through proper modification of the core Hamiltonian.

$$\mathbf{h}^{A-in-B} = \mathbf{h} + \mathbf{V}^{embed} + \mathbf{V}^{projector}. \tag{5}$$

The first modification is the embedding potential $\mathbf{V}^{embed}$,

$$\mathbf{V}^{embed} = \mathbf{G}[\boldsymbol{\gamma}] - \mathbf{G}[\boldsymbol{\gamma}^A] = \mathbf{G}[\boldsymbol{\gamma}^B] + \delta\mathbf{G}[\boldsymbol{\gamma}^A, \boldsymbol{\gamma}^B], \tag{6}$$

which contains the parts of the electronic potential involving the environment subsystem, as calculated at the initial level of theory. $\mathbf{V}^{embed}$ maintains the two-electron interactions between the subsystems during the embedded calculation. The second modification to the core Hamiltonian is the projector $\mathbf{V}^{projector}$ which projects the space that is occupied by ρ^B out of the embedded calculation. This projection ensures that the orbitals resulting from the embedded calculation are orthogonal to the orbitals that make up ρ^B. The initially proposed form of $\mathbf{V}^{projector}$ was that of an empirical level-shift projector,

$$\mathbf{V}^{projector} = \mu\mathbf{S}\boldsymbol{\gamma}^B\mathbf{S}, \tag{7}$$

where $\mathbf{S}$ is the overlap matrix in AOs and μ is the level-shift parameter (recommended value $\approx 10^6$ a.u.).[33] In addition, the Huzinaga projection operator has also been used,[36,37]

$$\mathbf{V}^{projector} = -\frac{1}{2}(\mathbf{F}\boldsymbol{\gamma}^B\mathbf{S} - \mathbf{S}\boldsymbol{\gamma}^B\mathbf{F}), \tag{8}$$

where $\mathbf{F}$ is the (unmodified) Fock matrix for the total system from the initial calculation.

Using $\mathbf{h}^{A-in-B}$ in place of $\mathbf{h}$, any subsequent calculation on the active subsystem is embedded into the environment subsystem. This method allows for the embedding of either DFT or WFT methods into a mean-field background. The complete energy for such embed-

ded calculations (in the case of WFT-in-DFT) can be formally written as[46]

$$E^{WFT-in-DFT} = E^{DFT} + \left(E_A^{WFT} - E_A^{DFT}\right), \tag{9}$$

with E^{DFT} being the total system energy of the initial calculation, E_A^{DFT} being the energy of subsytem A at the same level of theory and using $\mathbf{h}^{A-in-B}$, and E_A^{WFT} is the WFT energy of subsystem A calculated with $\mathbf{h}^{A-in-B}$. Expressing the total energy as in Eq. 9 casts the embedding problem as a correction to the total DFT energy of the system. If the environment is computed at the Hartree-Fock (HF) level, then both DFT energies in Eq.9 are replaced by the corresponding HF energies. In the case of embedded correlated methods such as Second-Order Møller-Plesset (MP2), an embedded HF is performed and used as the reference.

Algorithm 1 summarizes the embedding workflow.

Density Partitioning Schemes

The partitioning of the density generally consists of localizing the MOs from the initial calculation and assigning them to the active or environment regions based on their association with the active part of the system. The initial formulation[33] accomplished the partitioning through a combination of Pipek-Mezey (PM) localization[47] and Mulliken population screening,[48] labeled here as Pipek-Mezey/Mulliken (PMM) partitioning. Pipek-Mezey localization is performed on the initial MOs to produce a corresponding set of localized MOs (LMOs). The Mulliken population on the active atoms are calculated for each LMO, and any LMO with at least some minimum population (recommended Mulliken population of 0.4) on one of the active atoms is labeled as active. In general, the PM localization can be replaced with any suitable alterative localization scheme and other population analysis methods can replace the Mulliken populations.

More recently, Claudino and Mayhall[35] proposed an alternative method for localizing

Algorithm 1: General Projector-Based Embedding

Perform initial full system DFT calculation;
Acquire full system occupied MO coefficients $\mathbf{C}_{occ}$;
Full system density $\boldsymbol{\gamma} = \mathbf{C}_{occ} \cdot \mathbf{C}_{occ}^{T}$;
Acquire full system Fock matrix $\mathbf{F}$;
Acquire full system DFT energy E^{DFT};
Select active atoms;
Acquire active subsystem MO coefficients $\mathbf{C}_{occ}^{A}$ from partition method;
Active subsystem density $\boldsymbol{\gamma}^{A} = \mathbf{C}_{occ}^{A} \cdot (\mathbf{C}_{occ}^{A})^{T}$;
Environment subsystem density $\boldsymbol{\gamma}^{B} = \boldsymbol{\gamma} - \boldsymbol{\gamma}^{A}$;
Acquire active subsystem electronic potential $\mathbf{G}[\boldsymbol{\gamma}^{A}]$;
Form density projector $\mathbf{V}^{projector}$ as level-shift or Huzinaga;
Embedded core Hamiltonian $\mathbf{h}^{A-in-B} = \mathbf{F} - \mathbf{G}[\boldsymbol{\gamma}^{A}] + \mathbf{V}^{projector}$;
if *AOs truncated* **then**
    Acquire active AO submatrix of $\mathbf{h}^{A-in-B}$;
    Perform active subsystem DFT calculation using $\mathbf{h}^{A-in-B}$;
    Acquire active subsystem DFT energy E_{A}^{WFT};
else
    Active subsystem DFT energy $E_{A}^{DFT} = Tr(\boldsymbol{\gamma}^{A} \cdot \mathbf{h}^{A-in-B}) + \mathbf{G}[\boldsymbol{\gamma}^{A}]$;
end
Perform active subsystem WFT calculation using $\mathbf{h}^{A-in-B}$;
Acquire active subsystem embedded WFT energy E_{A}^{WFT};
WFT-in-DFT Total Energy $E^{WFT-in-DFT} = E^{DFT} - E_{A}^{DFT} + E_{A}^{WFT}$;

and dividing the density, called subsystem projected AO decomposition (SPADE). The MO coefficients are first transformed from the AOs to the set of orthogonalized AOs,

$$\bar{\mathbf{C}}_{occ} = \mathbf{S}^{\frac{1}{2}} \mathbf{C}_{occ}. \tag{10}$$

The transformed coefficients are then split into the parts comprised only of AOs centered on the active or environment atoms, $\bar{\mathbf{C}}_{occ}^{A}$ and $\bar{\mathbf{C}}_{occ}^{B}$ respectively. The part of the coefficients comprised of the active atom AOs is put though singular value decomposition (SVD),

$$\bar{\mathbf{C}}_{occ}^{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathbf{T}}, \tag{11}$$

where $\mathbf{U}$ and $\mathbf{V}$ are, respectively, the left and right singular vectors, and $\boldsymbol{\Sigma}$ is the diagonal

matrix of singular values. The right singular vectors are used to transform the original MOs into the new, localized SPADE orbitals.

$$\bar{C}_{occ}^{SPADE} = \bar{C}_{occ}\mathbf{V} \tag{12}$$

The partitioning of the SPADE orbitals into active and environment sets is determined from the singular values in $\mathbf{\Sigma}$, where the largest difference between successive singular values is used as the dividing line between the MO subsets. Orbitals before this largest difference are considered active and those after are considered environment.

In addition to PMM and SPADE, we introduce another scheme for partitioning the system density based on von Neumann entropy[49,50] and partial orbital occupancies, referred to as entropy/occupation partitioning (EOP). As with PMM, LMOs are first obtained from PM localization. The density matrix of each LMO is calculated and the submatrix corresponding to AOs on active atoms is obtained. The von Neumann entropy of this submatrix, which can be viewed as a measure of how isolated the LMO is to the active subspace, is calculated as

$$S^i = -Tr(\boldsymbol{\gamma}^i \ln \boldsymbol{\gamma}^i), \tag{13}$$

where S^i is the von Neumann entropy for LMO i and $\boldsymbol{\gamma}^i$ is the density matrix of LMO i in the active AO subspace. This value is observed to be directly related to the electronic occupation of LMO i within the active subspace of AOs,

$$N_e^i = Tr(\boldsymbol{\gamma}^i). \tag{14}$$

These partial occupancies function similarly to the Mulliken population in PMM, providing a sense to what degree the electrons in the various LMOs are associated with the active region of the system. The partial occupancies are also compared to a predetermined threshold value, with a recommended default value of 0.2 electrons. While the localization step dominates the

cost of both PMM and EOP, the Mulliken screening process for PMM requires the individual calculation of Mulliken populations on each active atom for each LMO until sufficient charge is found. On the other hand, the partial occupancy in EOP checks the screening of a LMO against all active atoms at once, providing a slightly more efficient algorithm for the screening.

Truncation of AO Space

Up to this point, the cost reduction in the embedded calculation comes from the decrease in the number of correlated occupied orbitals in the active subspace. Hartree-Fock (HF) and DFT benefit from this lowered number of occupied orbitals, but the costs of embedded correlated WFT methods are not necessarily reduced considerably. Correlated methods, such as second-order Møller-Plesset perturbation theory (MP2), scale with the size of the occupied *and* the virtual space, and the embedding simply shifts the balance between the two. Given the localized nature of the embedded calculation, it is reasonable to assume that the complete set of AOs for the full system is not needed for an accurate description of the active region.[38] Truncating the AO set of the embedded calculation leads to fewer virtual molecular orbitals and better scaling for correlated WFTs.

The truncation of the AO space can be accomplished by removing the AOs which are least significant in ρ^A.[39] The significance of an AO is judged by the net Mulliken population of ρ^A on that AO,

$$q_\alpha = \gamma^A_{\alpha\alpha} S_{\alpha\alpha}. \tag{15}$$

AO α is discarded if q_α is less than a predetermined threshold λ, which has a recommend value of 10^{-4}. Additionally, all AOs in a shell are kept if any AO in a shell is to be included, and any AO centered on an active atom is always kept. The submatrices of $\mathbf{h}^{A-in-B}$ and $\boldsymbol{\gamma}^A$ corresponding to the reduced basis set are acquired and used in the calculation of the last two values on the right hand side of Eq. 9. Additionally, the initial $\boldsymbol{\gamma}^A$ submatrix may

be purified to restore idempotency,[51] though this step is not strictly necessary and may fail depending on the state of the submatrix. In this work, the purification is always attempted, but defaults to the initial submatrix in cases of failure. This method of AO truncation defines a continuum with the full system AO space at one end and the space of AOs centered only on active atoms (referred to here as "total truncation") at the other.

In Ref. 40, the authors outline two algorithms for what they call "absolute localization." In this approach, ρ^A is represented within the set of basis functions centered on active atoms and ρ^B is expressed by the basis functions centered on environment atoms. The first algorithm, "orbital localization", is similar to the above descriptions with total truncation, but the Huzinaga projector is redefined in terms of the separate basis sets,

$$\mathbf{V}^{projector} = -\frac{1}{2}(\mathbf{F}^{AB}\boldsymbol{\gamma}^{B}\mathbf{S}^{BA} - \mathbf{S}^{AB}\boldsymbol{\gamma}^{B}\mathbf{F}^{BA}), \tag{16}$$

where the superscripts on the Fock and overlap matrices indicate that they have mixed indices over the separated active and environment AOs and $\boldsymbol{\gamma}^B$ is expressed in only the AOs centered on non-active atoms. The second algorithm, referred to as "iterative freeze-and-thaw", bypasses the need for orbital localization and density partitioning by performing alternating, iterative QM-in-QM calculations on each subsystem in their corresponding restricted basis sets. A full system lower level calculation is still required for the final energy in both algorithms. In likely application cases, the relative size of the active (N_A) and environment (N_B) AO sets are likely to be quite different, i.e., $N_A \ll N_B$. This difference also leads to the comparison that $N_B \approx N_A + N_B$ and that the QM-in-QM iterations on the environment will be similar in computational price to the full system calculations at the environment level of theory. The combined cost of the iterative QM-in-QM calculations and the total system calculation may render the freeze-and-thaw algorithm prohibitively expensive for systems where the full calculations are already costly.

Computational Details

Table 1: Descriptions of reactions studied, along with the MP2/aDZ values (mE_h) calculated for each.

#	Reaction Description	MP2/aDZ Value
1[a]	Deprotonation of 2,2,2-trifluoroethyloxonium	263.91
2[a]	Deprotonation of 2,2,2-trifluoroethanol	586.13
3[a]	Deprotonation of 2,2,2-trifluoroethylammonium	332.55
4[a]	Deprotonation of 2,2,2-trifluoroethylamine	627.06
5[a]	Deprotonation of 2,2,2-trifluoroacetic acid	527.86
6[b]	Deprotonation of phenol	561.78
7[a]	S_N2 reaction of 1,1,1,2-tetrafluoroethane and chloride anion	48.62
8[b]	Barrier height of propyl chloride and chloride anion S_N2 reaction	5.86
9[b]	Acid hydrolysis of dimethylether to methanol and methyl cation	-175.01
10[b]	Ring closing isomerization of 3-methylene-1-heptene to 1-butylcyclobutene	8.69
11[b]	Diels-Alder reaction between 2-methoxy-1,3-butadiene and methyl vinyl ketone	-77.92
12[b]	Barrier height of the above Diels-Alder reaction	10.66

[a]Found in Ref. 35

[b]Found in Ref. 52

To study the interactions of the various options within the projector-based formalisms (e.g., projector type, partition scheme, and level of truncation), a collection of reactions pulled from Refs 35 and 52 (outlined in Table 1) were selected. For each reaction, MP2-in-RHF and MP2-in-B3LYP calculations were performed using varying combinations of the above mentioned options. The comparison of RHF and B3LYP will highlight the effect of including or excluding electronic correlation in the environment. The suitability of the background calculation influences the quality of the embedded result, much like the requirement in perturbation theory that the zeroth-order wave function be close to the exact wave function. Conventional MP2 values were calculated for each system (also provided in Table 1) and the absolute error relative to those results was determined for each embedded calculation. The average errors were calculated with even weighting for each reaction. The

implementation of projector-based embedding used here[53] was constructed in Python using PySCF,[54] and serves as a prototype for the development of a similar implementation in NWChemEx.[55] The conventional MP2 results were also completed in PySCF for consistency. All calculation were performed in the Dunning-type aug-cc-pVDZ (denoted aDZ) basis sets for all atoms.[56,57] Each embedded calculation used either the Huzinaga or level-shift projector ($\mu = 10^6$).

Table 2: Specifications of active and environment atoms for reactants and products in selected reactions, and the number of active atoms associated with the standard or strict partitioning of each. Environment atoms are colored blue, while active atoms are colored red. Molecules that contain no environment atoms are treated at the conventional MP2 level (i.e., not embedded), and no electronic structure calculations were performed for H^+.

		# of Active MOs	
#	Molecules	Standard	Strict
1[a]	$CF_3CH_2OH_2^+$	9	8
	$CF_3CH_2OH + H^+$	9	8
2[a]	CF_3CH_2OH	9	8
	$CF_3CH_2O^- + H^+$	9	8
3[a]	$CF_3CH_2NH_3^+$	9	8
	$CF_3CH_2NH_2 + H^+$	9	8
4[a]	$CF_3CH_2NH_2$	9	8
	$CF_3CH_2NH^- + H^+$	9	8
5[a]	CF_3COOH	12	11
	$CF_3COO^- + H^+$	12	11
6[b]	C_6H_5OH	5	4
	$C_6H_5O^- + H^+$	5	4
7[a]	$CF_3CH_2F + Cl^-$	9	8
	$CF_3CH_2Cl + F^-$	13	12
8[b]	$CH_3CH_2CH_2Cl + Cl^-$	13	12
	$CH_3CH_2CH_2Cl_2^-$	22	21
9[b]	$CH_3OCH_3 + H^+$	9	8
	$CH_3OH + CH_3^+$	5	4
10[b]	$CH_2CHC(CH_2)CH_2CH_2CH_2CH_3$	15	14
	Ring closed	15	14
11[b]	$CH_2CHC(O)CH_3$	15	14
	$CH_2CHC(CH_2)OCH_3$	15	14
	Diels-Alder Product	30	28
12[b]	Transition State	30	28

[a]Found in Ref. 35

[b]Found in Ref. 52

The selection of active atoms and the corresponding number or active MOs for products and reactant is summarized in Table 2. The partitioning schemes are generally expected to select as active all MOs that are intuitively understood to be associated with an active atom, such as core electrons, lone pairs, and bonding orbitals. For PMM and EOP cases, the recommended threshold values were used, except where adjusted to ensure consistent active MO selection within a reaction. Due to similar consistency issues, the SPADE methodology was modified with an override to its active MO selection criteria to allow a predetermined number of the most significant SPADE orbitals to be selected as active, regardless of consecutive differences between the singular values. These issues with consistent active MO identification will be discussed more below.

Because PMM and EOP are using the same localized orbitals and their respective thresholds are set to recover the same number of active MOs, the results for each scheme will be the same. As such, calculations were also performed with a more restrictive EOP selection (referred to as EOP Strict), corresponding to an occupancy threshold of 0.85. This strict partition ensures that the active MOs are ones where the electrons are almost exclusively localized to the AOs on the active atoms and shifts previously active MOs into the environment density. The shifted MOs have non-negligible occupancy within the environment AOs and can be expected to cause issues in total truncation cases, where it would be difficult to reconstruct such an MO without the basis functions from the environment atom(s) with which it is associated.

Regarding AO truncation, the level of truncation resulting from the suggested default threshold (labeled Moderate) was compared to non-truncated results (labeled None). In addition to the default level of truncation, a total truncation down to only the AOs centered on active atoms was tested (labeled Total). This total truncation is equivalent to setting the truncation threshold to infinity, and is similar but distinct from the orbital localization algorithm of the absolute localization method (which has not been studied in this work). The number of AOs used in the various embedded calculations is summarized in Table 3.

Table 3: The number of AOs used at a given level of AO truncation for embedded calculations on reactants and products, depending on partitioning scheme and background level of theory.

Truncation Level		None	Moderate						Total
Partition Scheme		All	SPADE		PMM/EOP		EOP Strict		All
#	Molecules	HF/DFT	HF	DFT	HF	DFT	HF	DFT	HF/DFT
1[a]	Reactant	174	136	137	134	137	113	116	82
	Product	165	129	129	128	128	107	119	73
2[a]	Reactant	165	129	129	128	128	107	119	73
	Product	156	113	120	112	119	111	119	64
3[a]	Reactant	183	138	141	135	140	112	118	91
	Product	174	138	135	137	134	105	123	82
4[a]	Reactant	174	138	135	137	134	105	123	82
	Product	165	129	129	128	128	105	123	73
5[a]	Reactant	170	130	133	126	126	107	112	78
	Product	161	125	125	124	124	111	113	69
6[b]	Reactant	215	144	149	139	146	132	142	32
	Product	206	141	143	140	140	126	131	23
7[a]	Reactant	156	118	120	119	119	96	109	64
	Product	160	118	121	118	120	106	114	68
8[b]	Reactant	159	130	127	128	127	120	124	68
	Product	186	154	158	153	157	141	149	95
9[b]	Reactant	123	113	114	112	105	112	104	73
	Product	82	66	61	65	60	58	55	32
10[b]	Reactant	310	228	226	228	225	218	222	137
	Product	310	227	225	244	236	224	222	137
11[b]	Reactant 1	169	159	155	159	154	153	154	119
	Reactant 2	210	173	174	175	173	175	173	137
	Product	379	335	343	337	341	318	332	256
12[b]	Transition State	379	332	335	330	341	322	335	256

[a]Found in Ref. 35

[b]Found in Ref. 52

Results and Discussion

In this section, the general trends of the various procedural combinations will be analyzed, along with more detailed considerations of the different reaction types.

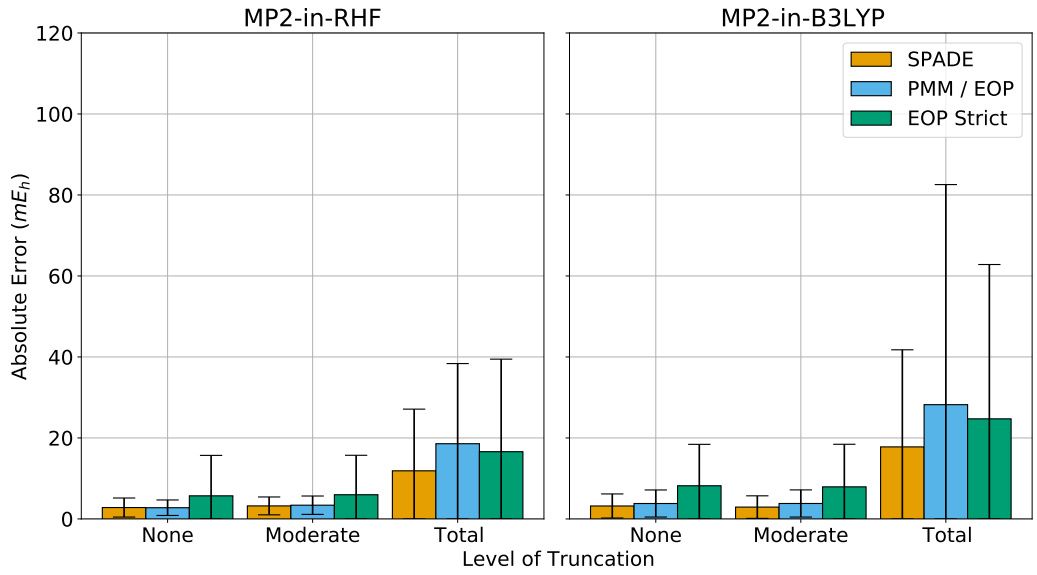


Figure 1: Average unsigned errors (mE_h) and standard deviations for embedded calculations on all studied systems using the level-shift projection. The negative error bars are set to zero.

Overall Average Trends

Figure 1 presents the average absolute errors where the level-shift projector was utilized with the various partition schemes (SPADE in orange, PMM/EOP in blue, and EOP Strict in green) and truncation levels. The RHF and B3LYP results are similar for the cases with no or moderate AO truncation, and the SPADE and PMM/EOP results perform reasonably in these cases. The averages for EOP Strict are worse and have considerably higher deviations. Most obviously, the use of total AO truncation is completely unsuitable with the level-shift, regardless of partition scheme or background. The failure of level-shift projector with considerable AO truncation has been previously noted,[42] and is related to its inability to enforce complete orthogonality between the subsystems. Figure 2 presents the same values for the Huzinaga projector cases. It is observed that the Huzinaga projector performs reasonably in all cases, and has much more restrained deviations (note the different scales for the y axes). The observation that the Huzinaga projector is preferable to the level-shift projector has been previously noted.[37,40,42] Based on the overall preferability of the Huzinaga projector,

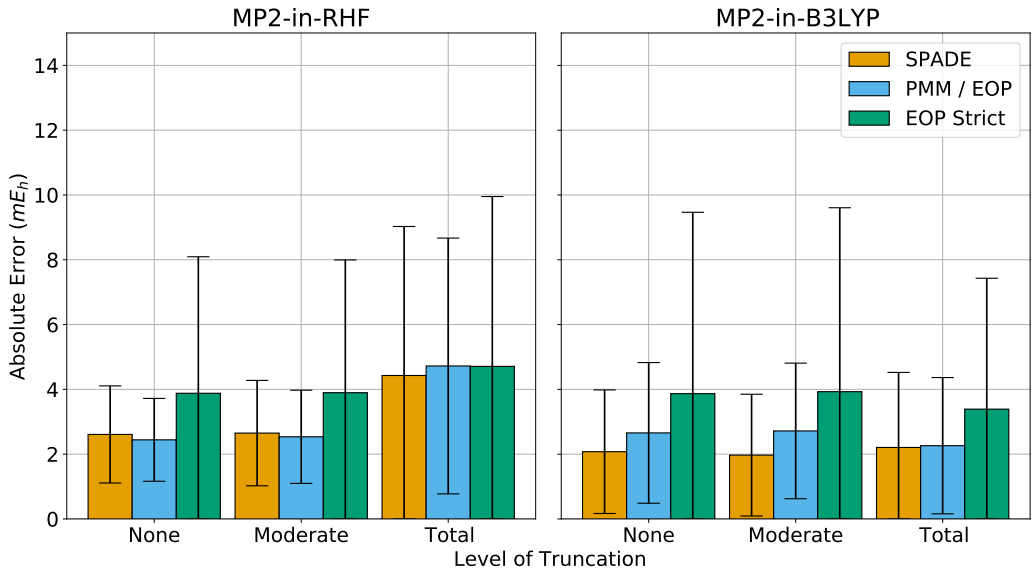


Figure 2: Average unsigned errors (mE_h) and standard deviations for embedded calculations on all studied systems using Huzinaga projection. The negative error bars are set to zero.

the remainder of the paper will focus exclusively on those results. The results from the level-shift calculations are available in the supporting material.

Figure 2 shows that the average error for the Huzinaga based calculations are between 2 and 5 mE_h for all combinations of options. Additionally, the standard deviation for calculations without the strict partitioning selection (SPADE and PMM/EOP) are mostly around 2 mE_h. The differences between the PMM/EOP pair and the SPADE results are minimal, with SPADE performing slightly better within the B3LYP environment. The combination of Huzinaga projection and SPADE partitioning has the least number of empirical parameters (with only the truncation parameter remaining, if used) and provides the most black box procedure. The level of truncation has very little effect on the errors for the tested systems (especially for the MP2-in-B3LYP results). A completely black box procedure can be reached with Huzinaga projection, SPADE partitioning, and total (or no) truncation. The difference in total truncation errors between the RHF and B3LYP backgrounds does point to a potential issue of suitability with the background, so a more moderate truncation level may still be more robust.

The strict EOP results have consistently higher errors and standard deviations, which is partly explained by the fact that less MOs are being handled at the MP2 level. Another contributing factor to the errors is the "close-ness" of some of the active versus environment selections. Based on the locality from which embedding theory is derived, there is a hypothetical distance from the chemical reaction at which particles in the system experience effectively no change. Ideally, the selection of active atoms and related electron density should include all particles effected by the reaction and exclude all particles that are not effected. Some of the tested systems (discussed below) have active/environment selections that can be considered quite close, i.e., some particles of the environment can be expected to experience a non-trivial change due to the reactions.

Deprotonation Reactions

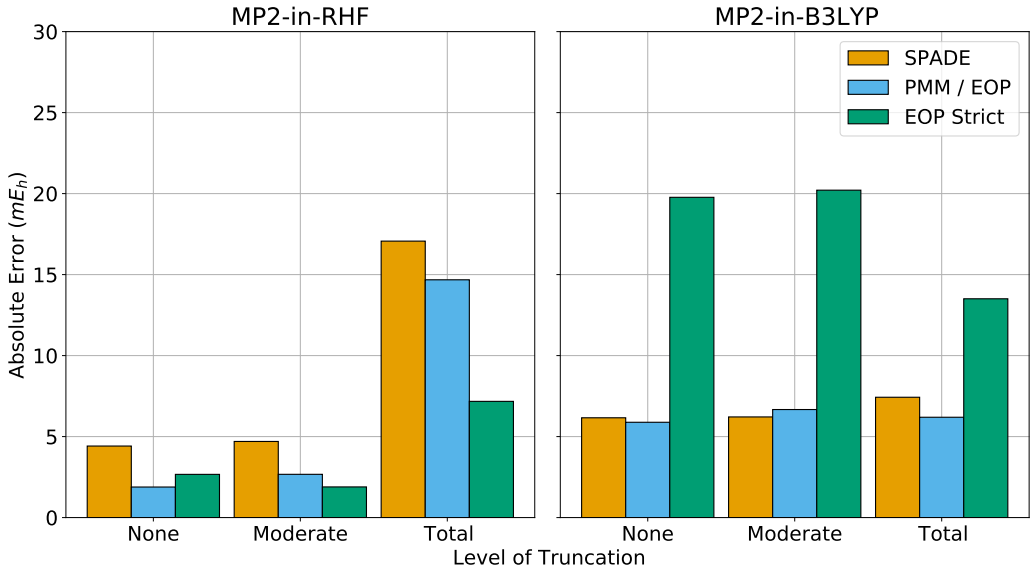


Figure 3: Unsigned errors (mE_h) for embedded calculations of Reaction 6 using Huzinaga projection.

Reactions 1-6 deal with various deprotonation cases, including alcohols, amines, and carboxylic acids. The reactions involving trifluoroethane derivatives (Reactions 1-5) all have

errors that tend to be lower than the corresponding average. The electron withdrawing nature of the fluorine atoms may lead to a particular stable environment density, making these systems particularly suitable for partitioning and embedding. Graphs for these reactions can be found in Figs S1-S5. Generally, the errors are higher for the calculations embedded into the RHF environment, which can be linked to the lack of correlation for non-active electrons. The combination of B3LYP environment, Huzinaga projection, SPADE partitioning, and total truncation often provides the best results for these reactions.

Figure 3 displays the errors for Reaction 6 (deprotonation of the hydroxyl group in phenol). For the calculations in the RHF environment, SPADE has the largest error in each of the truncation levels, while total truncation with any partition proves to be unsuitable for this environment. The calculations in the B3LYP environment are quite uniform with regards to truncation level. The SPADE and PMM/EOP errors are higher than average, while EOP Strict appears to be unsuitable at any level of truncation. The particular failures in both environments can be related to the interaction of active and environment subsystems within the phenol.

The atoms in the hydroxyl group are selected to be active, while the phenyl group is considered the environment. In general, the electronics of the two subsystems are likely still quite sensitive to changes in the other, especially given the likelihood of electron density from the oxygen spreading into the phenyl ring π system in the deprotonated state. The failure of total truncation in the RHF environment is due to the lack of delocalization in the environment. Without the basis functions on the environment atoms, the electrons around the oxygen remain isolated and unable to delocalize into the phenyl ring. This same issue is not seen in the B3LYP case since the charge delocalization is already better represented in the environment since DFT tends to delocalize the electron density more than RHF. In the RHF/EOP Strict case, the electrons corresponding to the C-O bond connecting the subsystems is shifted into the environment. This shift means there is one less occupied orbital and one more virtual orbital in the MP2 calculation in the active space, providing a small

increase in flexibility that allows for better performance in EOP Strict with total truncation. In the B3LYP environment, shifting this bond becomes detrimental for the EOP Strict cases. The representation of electronic correlation within MP2 and B3LYP is understandably not identical, and mixing the handling of the correlation of electrons on the oxygen (those in the C-O bond vs. all others) produces an imbalance that leads to increased error. The better performance of EOP Strict with total truncation is again related to the increased flexibility for correlation in MP2 when there are less electrons.

S_N2 Reactions

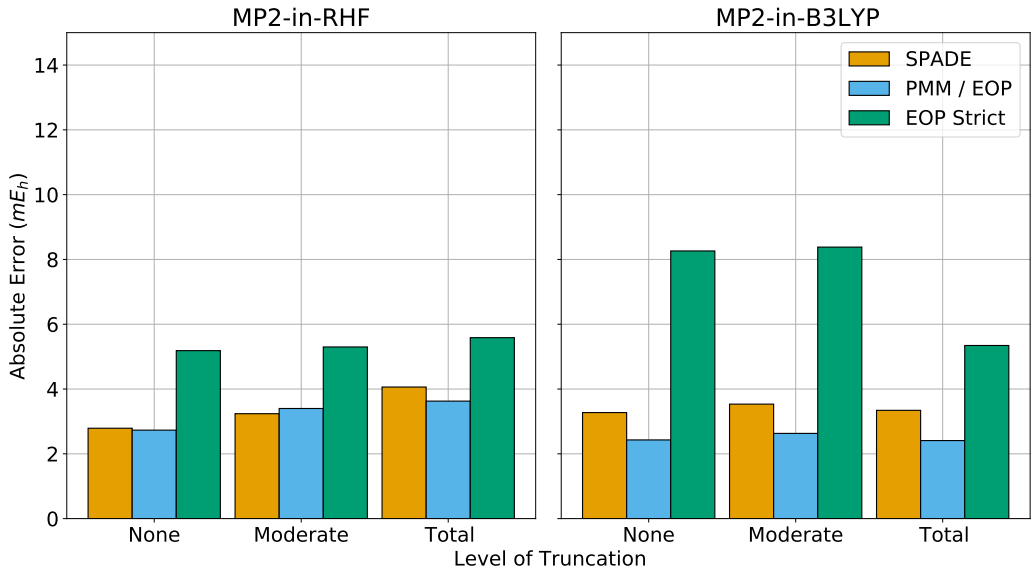


Figure 4: Unsigned errors (mE_h) for embedded calculations of Reaction 7 using Huzinaga projection.

Reactions 7 and 8 investigate the adequacy of the embedding options for S_N2 reaction energies and barrier heights, respectively. The errors for Reaction 7 can be found in Fig. 4, while the results for reaction 8 are in Fig. S6. The errors for Reaction 7 tend to be slightly higher than the overall average, though mostly well within the standard deviation. The exceptions here are the EOP Strict results in the B3LYP environment. These errors are

not as high as those seen in Reaction 6, but have similar origins in the mixing of correlation handling. The best results for this system are seen from the PMM/EOP results in the B3LYP environment, which are insensitve to the truncation level. The calculated errors for reaction 8 are quite good, especially in the B3LYP environment. The SPADE results in B3LYP are also good, though suffer slightly with total truncation. The only standout values are those for total truncation in the RHF environment, which are most likely attributable to the same basis set reduction issue noted for Reaction 6 in the RHF environment.

Acid Hydrolysis and Ring Closing Reactions

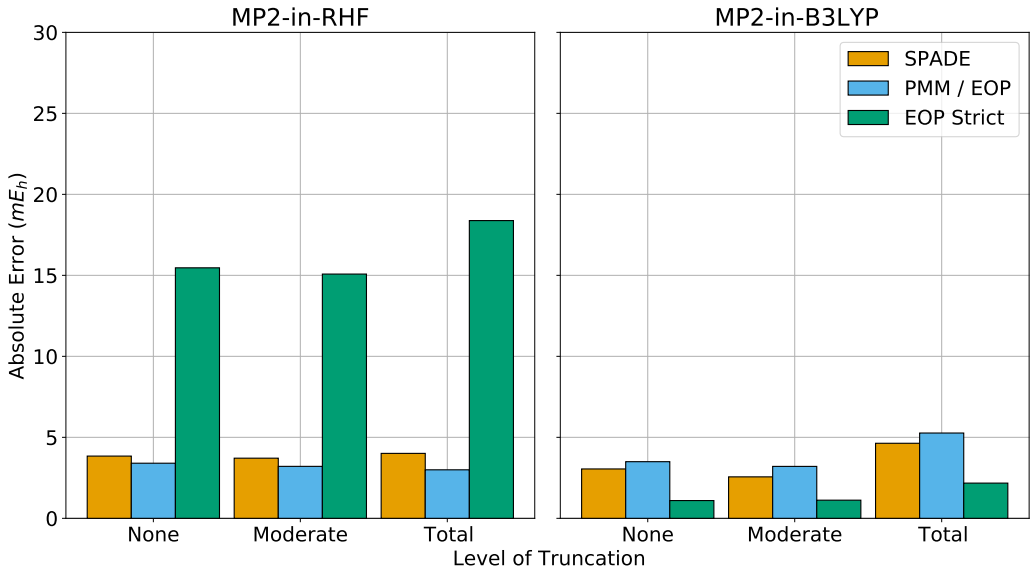


Figure 5: Unsigned errors (mE_h) for embedded calculations of Reaction 9 using Huzinaga projection.

Figure 5 shows the errors for Reaction 9, an acid hydrolysis reaction. The notable issue in this reaction is the high errors associated with the strict EOP in the RHF environment, as opposed to the lower errors for the same partitioning in the B3LYP environment. This difference is easily explained by the RHF case's complete lack of correlation for the C-C bond crossing between the subsystems. The same explanation can be applied to the similar trend

in the errors for Reaction 10 (Fig. S7), though the errors for that reaction are much smaller in general. Aside from the EOP Strict results in the RHF, the quality of results for both systems is relatively indifferent to the choice of background, partitioning, or truncation.

Though the partitioning accomplished by the PMM and EOP results explicitly overlap, they do show different levels of sensitivity in some cases. An example of this difference was seen in the calculations for the reactant and product of reaction 10 in the B3LYP environment. In the reactant case, the default threshold of 0.2 was slightly too tight, giving the same result as the EOP Strict threshold of 0.85. The desired number of active MOs was recovered by lowering the threshold to 0.15. For the product, the default threshold proved too loose, with the EOP partition including an additional MO into the active subsystem compared to the corresponding PMM or SPADE algorithms. Increasing the threshold to 0.3 proved to be too much, with the resulting partitioning being equivalent to that of EOP Strict. A threshold value of 0.29 was required to match the PMM selection.

Diels-Alder Reaction

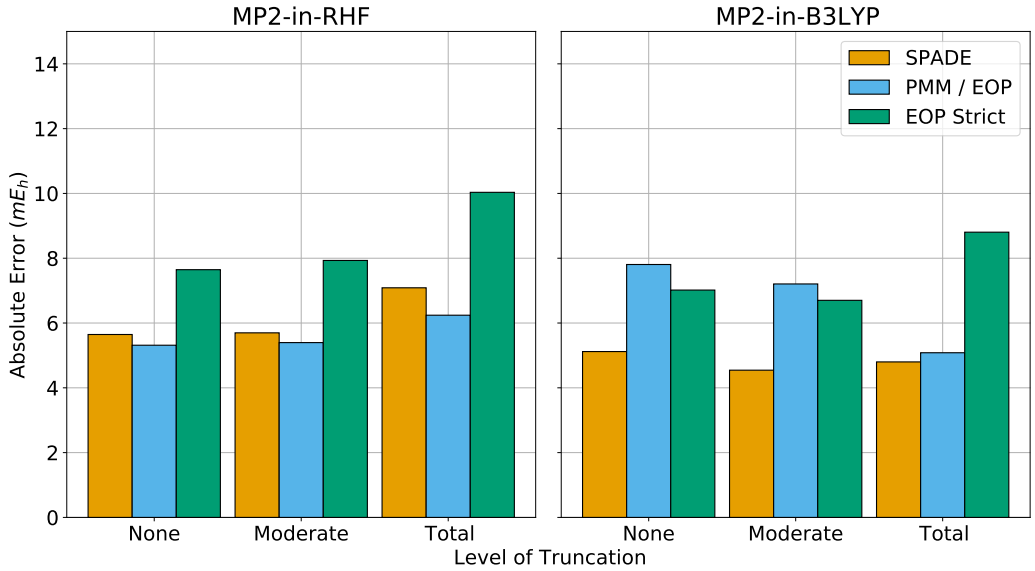


Figure 6: Unsigned errors (mE_h) for embedded calculations of Reaction 12 using Huzinaga projection.

Reactions 11 and 12 both concern the same Diels-Alder reaction of 2-methoxy-1,3,-butadiene and methyl vinyl ketone, and represent the largest molecular system in the collected reactions. Figure S8 provides the errors for the energy of the reaction, while Fig. 6 provides the errors for the barrier height. The reaction energy errors are below average for all categories, many of them close to or below 1 mE_h. On the other hand, the reaction barrier errors are all above average, with most near or outside of the standard deviation. The most likely explanation is that the changes of the electronic structure within the transition state reach beyond the active MOs, and that the treatment of those changes in the environment is not adequate. It is reasonable to assume that calculations involving transition states may require a broader selection for the active subsystem compared to equivalent calculations of reactants or products.

While PMM, EOP, and SPADE are usually quite similar in their active MO selection, they do display different edge case issues, as observed in the RHF environment calculations for this system. In calculations on the product, default PMM and EOP exclude an orbital corresponding to the C-O bond between the methoxy oxygen (an environment atom) and the ring carbon with which it is bonded (an active atom). The corresponding orbitals in the reactant and transition state are included in the active selection, resulting in an imbalance in its treatment between different calculations on the potential energy surface. The error in the reaction energy associated with these imbalanced results is around 50 mE_h in the Huzinaga projected case without AO truncation. The missing orbital is easily recovered in the product calculation by lowering the screening thresholds from 4.0 to 3.9 for PMM and 0.2 to 0.15 for for EOP, which implies that the difference between the densities is very subtle. SPADE has a similarly inconsistent treatment of this orbital, only it is excluded in the reactant and included in the transition state and product. The imbalance leads to errors in the untruncated results around 48 and 42 mE_h for the reaction energy and barrier height, respectively. As described above, this problem is overcome by overriding SPADE's selection criteria to include a predetermined number of the most significant orbitals (as determined

by their corresponding singular values).

Conclusion

The Huzinaga projector is shown to be more accurate than the level-shift based projector, while also more reliable with higher levels of AO truncation. The PMM, SPADE, and new EOP partition schemes all produce results of similar quality when utilized at a suitable level, with SPADE showing a slight edge. While both EOP and PMM rely on similar orbital localization techniques, the occupancy-based screening performed in EOP is computationally cheaper than the population screening in PMM. Surprisingly, the truncation of the AO space used in the embedded calculation does not appear to cause much deterioration in the result quality, often even when all AOs on non-active atoms have been removed.

As such, a cheap, black box procedure without undue error can be found in the combination of Huzinaga projection, SPADE partitioning, and total AO truncation. That said, the exact reliability of the black box nature of this combination is still questionable. The SPADE scheme shows good consistency in MO selection across a reaction pathway, but does have edge cases where it falters. These issues of inconsistent active MO selection can be alleviated by the application of methodologies that select MOs across the pathway (such as the "even-handed" or direct selection schemes[43,45]), instead of only selecting for a individual geometry. Similarly, total AO truncation is reliable in many cases but can encounter issues in MOs that are particularly dependent on AOs from the environment. As a note, the use of looser thresholds in the PMM or EOP partitions are likely to exacerbate these issues, as the resulting increase in active LMOs is not complimented by an increase in the number of AOs in the truncated basis set. Total AO truncation can only be expected to be reliable when the active subsystem density is reasonably representable within the AOs centered on the active atoms.

The application of projection-based embedding to truly large systems, where the active

subsystem is significantly smaller than that of the environment, has not yet been thoroughly considered. Many of the issues regarding proper partitioning and AO truncation can be expected to manifest differently in large-scale cases, where the localizability of the electronic structure of the system has to be considered,[58] while still allowing for sufficient isolation of the desired phenomena. Using NWChemEx to efficiently leverage exascale computation will open up the possibility to apply these embedding techniques to systems of considerable size, allowing for further investigation into the effects of subsystem partitioning and AO truncation.

Acknowledgement

This research was supported by the Exascale Computing Project (17-SC-20-SC), a collaborative effort of the U.S. Department of Energy Office of Science and the National Nuclear Security Administration.

Supporting Information Available

The supporting information includes individual unsigned error results not included in the main text.

References

(1) Jensen, F. *Introduction to Computational Chemistry*, 3rd ed.; John Wiley & Sons: West Sussex, England ; New York, 2017.

(2) Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, 1st ed.; Dover Publications: Mineola, New York, 1996.

(3) Bartlett, R. J.; Musiał, M. Coupled-cluster theory in quantum chemistry. *Rev. Mod. Phys* **2007**, *79*, 291.

(4) Roos, B. O. The complete active space self-consistent field method and its applications in electronic structure calculations. *Adv. Chem. Phys.* **1987**, *69*, 399–445.

(5) Hohenberg, P.; Kohn, W. Inhomogeneous electron gas. *Phys. Rev.* **1964**, *136*, B864.

(6) Kohn, W.; Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **1965**, *140*, A1133.

(7) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Challenges for density functional theory. *Chem. Rev.* **2012**, *112*, 289–320.

(8) Jones, R. O. Density functional theory: Its origins, rise to prominence, and future. *Rev. Mod. Phys* **2015**, *87*, 897.

(9) Pribram-Jones, A.; Gross, D. A.; Burke, K. Dft: A theory full of holes? *Annu. Rev. Phys. Chem.* **2015**, *66*, 283–304.

(10) Yu, H. S.; Li, S. L.; Truhlar, D. G. Perspective: Kohn-Sham density functional theory descending a staircase. *J. Chem. Phys.* **2016**, *145*, 130901.

(11) Seijo, L.; Barandiaran, Z. *Computational chemistry: reviews of current trends*; World Scientific, 1999; pp 55–152.

(12) Jacob, C. R.; Neugebauer, J. Subsystem density-functional theory. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 325–362.

(13) Libisch, F.; Huang, C.; Carter, E. A. Embedded correlated wavefunction schemes: Theory and applications. *Acc. Chem. Res.* **2014**, *47*, 2768–2775.

(14) Wesolowski, T. A.; Shedge, S.; Zhou, X. Frozen-density embedding strategy for multilevel simulations of electronic structure. *Chem. Rev.* **2015**, *115*, 5891–5928.

(15) Sun, Q.; Chan, G. K.-L. Quantum embedding theories. *Acc. Chem. Res.* **2016**, *49*, 2705–2712.

(16) Sushko, P. V.; Huang, C.; Govind, N.; Kowalski, K. Embedding Methods in Materials Discovery. *Computational Materials Discovery* **2018**, 87–116.

(17) Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976**, *103*, 227 – 249.

(18) Brunk, E.; Rothlisberger, U. Mixed quantum mechanical/molecular mechanical molecular dynamics simulations of biological systems in ground and electronically excited states. *Chem. Rev.* **2015**, *115*, 6217–6263.

(19) 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 + H2 Oxidative Addition. *J. Phys. Chem.* **1996**, *100*, 19357–19363.

(20) Chung, L. W.; Sameera, W.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z.; Liu, F. et al. The ONIOM method and its applications. *Chem. Rev.* **2015**, *115*, 5678–5796.

(21) Govind, N.; Wang, Y.; Da Silva, A.; Carter, E. Accurate ab initio energetics of extended systems via explicit correlation embedded in a density functional environment. *Chem. Phys. Lett.* **1998**, *295*, 129–134.

(22) Huang, C.; Pavone, M.; Carter, E. A. Quantum mechanical embedding theory based on a unique embedding potential. *J. Chem. Phys.* **2011**, *134*, 154110.

(23) Elliott, P.; Burke, K.; Cohen, M. H.; Wasserman, A. Partition density-functional theory. *Phys. Rev. A* **2010**, *82*, 024501.

(24) Nafziger, J.; Wasserman, A. Density-based partitioning methods for ground-state molecular calculations. *J. Phys. Chem. A* **2014**, *118*, 7623–7639.

(25) Huang, C.; Carter, E. A. Potential-functional embedding theory for molecules and materials. *J. Chem. Phys.* **2011**, *135*, 194104.

(26) Fornace, M. E.; Lee, J.; Miyamoto, K.; Manby, F. R.; Miller III, T. F. Embedded mean-field theory. *J. Chem. Theory Comput.* **2015**, *11*, 568–580.

(27) Inglesfield, J. A method of embedding. *J. Phys. C* **1981**, *14*, 3795.

(28) Chibani, W.; Ren, X.; Scheffler, M.; Rinke, P. Self-consistent Green's function embedding for advanced electronic structure methods based on a dynamical mean-field concept. *Phys. Rev. B* **2016**, *93*, 165106.

(29) Lan, T. N.; Kananenka, A. A.; Zgid, D. Communication: Towards ab initio self-energy embedding theory in quantum chemistry. *J. Chem. Phys.* **2015**, *143*, 241102.

(30) Rusakov, A. A.; Iskakov, S.; Tran, L. N.; Zgid, D. Self-energy embedding theory (SEET) for periodic systems. *J. Chem. Theory Comput.* **2018**, *15*, 229–240.

(31) Knizia, G.; Chan, G. K.-L. Density Matrix Embedding: A Strong-Coupling Quantum Embedding Theory. *J. Chem. Theory Comput.* **2013**, *9*, 1428–1432.

(32) Li, W.; Chen, M.; Rabani, E.; Baer, R.; Neuhauser, D. Stochastic embedding DFT: Theory and application to p-nitroaniline in water. *J. Chem. Phys.* **2019**, *151*, 174115.

(33) Manby, F. R.; Stella, M.; Goodpaster, J. D.; Miller, T. F. A Simple, Exact Density-Functional-Theory Embedding Scheme. *J. Chem. Theory Comput.* **2012**, *8*, 2564–2568.

(34) Lee, S. J. R.; Ding, F.; Manby, F. R.; Miller, T. F. Analytical gradients for projection-based wavefunction-in-DFT embedding. *J. Chem. Phys.* **2019**, *151*, 064112.

(35) Claudino, D.; Mayhall, N. J. Automatic Partition of Orbital Spaces Based on Singular Value Decomposition in the Context of Embedding Theories. *J. Chem. Theory Comput.* **2019**, *15*, 1053–1064.

(36) Huzinaga, S.; Cantu, A. A. Theory of Separability of Many-Electron Systems. *J. Chem. Phys.* **1971**, *55*, 5543–5549.

(37) Hégely, B.; Nagy, P. R.; Ferenczy, G. G.; Kállay, M. Exact density functional and wave function embedding schemes based on orbital localization. *J. Chem. Phys.* **2016**, *145*, 064107.

(38) Barnes, T. A.; Goodpaster, J. D.; Manby, F. R.; Miller, T. F. Accurate basis set truncation for wavefunction embedding. *J. Chem. Phys.* **2013**, *139*, 024103.

(39) Bennie, S. J.; Stella, M.; Miller, T. F.; Manby, F. R. Accelerating wavefunction in density-functional-theory embedding by truncating the active basis set. *J. Chem. Phys.* **2015**, *143*, 024105.

(40) Chulhai, D. V.; Goodpaster, J. D. Improved Accuracy and Efficiency in Quantum Embedding through Absolute Localization. *J. Chem. Theory Comput.* **2017**, *13*, 1503–1508.

(41) Graham, D. S.; Wen, X.; Chulhai, D. V.; Goodpaster, J. D. Robust, Accurate, and Efficient: Quantum Embedding Using the Huzinaga Level-Shift Projection Operator for Complex Systems. *J. Chem. Theory Comput.* **2020**, *16*, 2284–2295.

(42) Bensberg, M.; Neugebauer, J. Automatic basis-set adaptation in projection-based embedding. *J. Chem. Phys.* **2019**, *150*, 184104.

(43) Bensberg, M.; Neugebauer, J. Direct orbital selection for projection-based embedding. *J. Chem. Phys.* **2019**, *150*, 214106.

(44) Bensberg, M.; Neugebauer, J. Orbital Alignment for Accurate Projection-Based Embedding Calculations along Reaction Paths. *J. Chem. Theory Comput.* **2020**, *16*, 3607–3619.

(45) Welborn, M.; Manby, F. R.; Miller, T. F. Even-handed subsystem selection in projection-based embedding. *J. Chem. Phys.* **2018**, *149*, 144101.

(46) Govind, N.; Wang, Y. A.; Carter, E. A. Electronic-structure calculations by first-principles density-based embedding of explicitly correlated systems. *J. Chem. Phys.* **1999**, *110*, 7677–7688.

(47) Pipek, J.; Mezey, P. G. A fast intrinsic localization procedure applicable for ab initio and semiempirical linear combination of atomic orbital wave functions. *J. Chem. Phys.* **1989**, *90*, 4916–4926.

(48) Mulliken, R. S. Electronic Population Analysis on LCAO–MO Molecular Wave Functions. I. *J. Chem. Phys.* **1955**, *23*, 1833–1840.

(49) von Neumann, J. *Mathematical Foundations of Quantum Mechanics*, 1st ed.; Princeton University Press: Princeton, NJ, 1955.

(50) Golub, P.; Antalik, A.; Veis, L.; Brabec, J. Automatic selection of active spaces for strongly correlated systems using machine learning algorithms. 2020; `https://arxiv.org/abs/2011.14715`.

(51) McWeeny, R. Some Recent Advances in Density Matrix Theory. *Rev. Mod. Phys.* **1960**, *32*, 335–369.

(52) Goodpaster, J. D.; Barnes, T. A.; Manby, F. R.; Miller, T. F. Accurate and systematically improvable density functional theory embedding for correlated wavefunctions. *J. Chem. Phys.* **2014**, *140*, 18A507.

(53) Available at: https://github.com/NWChemEx-Project/PythonProjectorEmbedding.

(54) Sun, Q.; Berkelbach, T. C.; Blunt, N. S.; Booth, G. H.; Guo, S.; Li, Z.; Liu, J.; McClain, J. D.; Sayfutyarova, E. R.; Sharma, S. et al. PySCF: the Python-based simulations of chemistry framework. 2017; `https://onlinelibrary.wiley.com/doi/abs/10.1002/wcms.1340`.

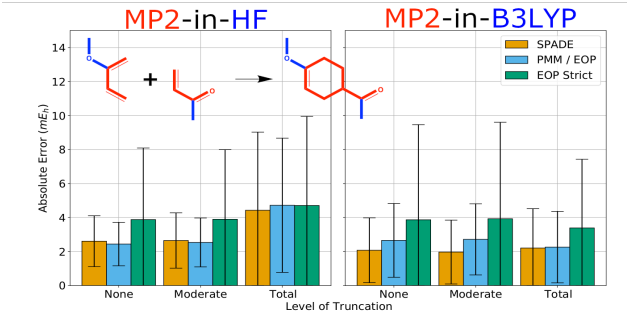
(55) Kowalski, K.; Bair, R.; Bauman, N. P.; Boschen, J. S.; Bylaska, E. J.; Daily, J.; de Jong, W. A.; Dunning, T.; Govind, N.; Harrison, R. J. et al. From NWChem to NWChemEx: Evolving with the Computational Chemistry Landscape. *Chem. Rev.* **2021**, *121*, 4962–4998.

(56) Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

(57) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96*, 6796–6806.

(58) Ratcliff, L. E.; Mohr, S.; Huhs, G.; Deutsch, T.; Masella, M.; Genovese, L. Challenges in large scale quantum mechanical calculations. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2017**, *7*, e1290.

TOC Graphic


MP2-in-HF
MP2-in-B3LYP
SPADE
PMM / EOP
EOP Strict
Absolute Error (mE_h)
0
2
4
6
8
10
12
14
None
Moderate
Total
None
Moderate
Total
Level of Truncation