Bootstrap Embedding for Molecules in Extended Basis Sets

Henry K. Tran,[†,‡] Leah P. Weisburn,[†,‡] Minsik Cho,[†] Shaun Weatherly,[†]

Hong-Zhou Ye,[¶] and Troy Van Voorhis[*,†]

†*Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

‡*Denotes Equal Contribution*

¶*Department of Chemistry, Columbia University, New York City, NY, 10027, USA*

E-mail: tvan@mit.edu

Abstract

Quantum embedding methods are powerful tools to exploit the locality of electron correlation, but thus far many wavefunction-in-wavefunction methods have focused on small (e.g. minimal) basis sets. One major challenge for extended basis sets lies in defining consistent atom- or fragment-localized orbitals in spite of the larger spatial extent of the underlying atomic orbitals. In this work, we modify a particular form of quantum embedding, Bootstrap Embedding (BE), to the case of extended basis sets. We find that using intrinsic atomic orbital (IAO) localization schemes alongside BE converges to ∼99.7% of the CCSD correlation energy in 3-21G, 6-311G, and cc-pVDZ basis sets for reasonably sized fragments. These results mark an important first step in extending the success of embedding methods to properly studying dynamic correlation.

1 Introduction

The coveted goal of chemically accurate quantum calculations hinges on developing methods capable of capturing at least 95% of the electronic correlation beyond the mean-field.[1] Methods such as coupled cluster (CC) singles and doubles (CCSD)[2] have been known to reach chemical accuracy, but CCSD in particular scales as $\mathcal{O}(N^6)$, quickly making it intractable beyond 10–30 atoms. Not to be deterred, many great advances have been made on the optimism that further method development can extend the great success of these proven methods on small systems to new methods equally as successful on large systems. This is an expansive and flourishing field, and a few successes will be highlighted here.

One approach is to restrict the correlated calculation to a subset of the full orbital space. A popular criterion is to pick only orbitals within an energy range lying around the valence orbitals, which intuitively is where the most important electronic effects should occur. Among these types of methods are active space methods such as complete active space self-consistent field[3–6] and restricted active space self-consistent field methods.[7,8] Rather than an energy metric, others have realized that electron correlation decays as r^{-3}. This inspires a class of local correlation methods that restrict the calculation to spatially close orbitals[9–15] such as the popular domain-based local pair natural orbital CC (DLPNO-CC).[16–19] These methods have shown that CCSD(T)-like accuracy can be achieved with near linear scaling in some cases.

With the same aim of linear scaling in mind, fragment embedding methods[20,21] partition the system spatially into fragments, each correlated locally and the total energy comprised as a sum over fragments. Unfortunately, the interaction of the fragment with the rest of the system, the bath, cannot be ignored. Nothing would be gained if the bath is treated at the same level of theory, so the bath is typically treated with cheaper theories. On the less expensive end of theories, the bath can be treated classically as point charges, popularized in QM/MM[22] and fragment molecular orbitals,[23,24] or as densities, used in frozen density embedding.[25,26] To obtain chemical accuracy for small fragment sizes, a more accurate quan-

tum treatment of the bath is essential. Being one of the most popular low-scaling quantum methods, DFT[27,28] has been used to treat the bath in subsystem DFT[29,30] and projection-based embedding methods.[31–35] Despite the practical scaling of DFT, the lack of systematic improvement along with various shortcomings[36] greatly limit quantum embedding methods based on DFT. The next step towards accurate, linear scaling, electron correlation in the framework of fragment embedding methods is to move beyond the density alone and instead approximate the bath wave function. Methods of this type include those based on the Schmidt Decomposition such as density matrix embedding theory (DMET),[37–45] local active space self-consistent field,[46–48] orbital partitioning schemes,[49] projected site-occupation embedding theory,[50] second order active space embedding theory,[51] and bootstrap embedding theory (BE).[52–60] These methods have shown great promise for low-scaling electron correlation but mostly for static correlation and minimal basis sets.

In this paper, we extend BE to work efficiently in large basis sets for molecular systems. We utilize an IAO-based localization scheme for multiple systems in extended basis sets, to remove the ambiguities that arise in defining fragments with spatially delocalized extended bases. For all systems, BE is able to capture >99.7% of the CCSD full system correlation energy with reasonably-sized fragments. Notably, the number of atoms per fragment that is required to achieve this level of accuracy appears to be roughly independent of both the system size and the density of the atomic orbital basis set. This paper is organized as follows. In Section 2, we review the theory behind BE and quantum embedding, and discuss how one generalizes this theory for extended basis sets. In Section 3, we discuss the computational details related to our results including the molecules of interest and the details of our BE calculation. In Section 4, we present the results of BE in extended basis sets using the theory presented and interpret the results. In Section 5, we wrap up the results and discuss future directions.

2 Theory

In this section, we review the established working equations of BE for minimal basis sets and discuss the modifications made for extended basis sets. BE follows three steps. First, a localization scheme is applied to a chemical system, split into disjoint atomic fragments, to define a set of localized fragment orbitals (FO) for each. Second, a set of delocalized bath orbitals (BOs) are defined for each fragment by choosing the BOs that are maximally entangled with that fragment's FOs. Finally, a correlated calculation is performed over each FO + BO space and matching conditions are applied to overlapping fragments.

2.1 BE In Minimal Basis Sets

In second quantized notation, the electronic Hamiltonian is

$$\hat{H} = \sum_{\mu\nu}^{N} h_{\mu\nu}\hat{c}_{\mu}^{\dagger}\hat{c}_{\nu} + \frac{1}{2}\sum_{\mu\nu\lambda\eta}^{N} V_{\mu\nu\lambda\eta}\hat{c}_{\mu}^{\dagger}\hat{c}_{\lambda}^{\dagger}\hat{c}_{\eta}\hat{c}_{\nu} \tag{1}$$

μ, ν, λ, η denote the working minimal basis set of atomic orbitals (AOs). $\hat{c}_{\mu}^{\dagger}$ ($\hat{c}_{\mu}$) are the creation (annihilation) operators for μ. $h_{\mu\nu}$ are the one electron integrals and $V_{\mu\nu\lambda\eta}$ are the two electron integrals.

2.1.1 Fragment Orbitals

We begin by describing the choice of FOs. FOs are a set of orthogonal orbitals in which each orbital is localized on one atom. We borrow from local correlation theories and obtain a set of orthogonal, localized orbitals (LOs) from a localization procedure such as Boys.[61]

In order to determine which LOs belong to which fragments, an "atom-based" partitioning scheme was developed for use with BE where the molecule is broken down into atomic groups (a non-hydrogen atom and all hydrogens attached to that atom).[56] Every LO is assigned to an atomic group. An example of this is in Figure 1. Fragments are then chosen as any set of atomic groups, and the set of FOs are all the LOs assigned to these atomic groups.

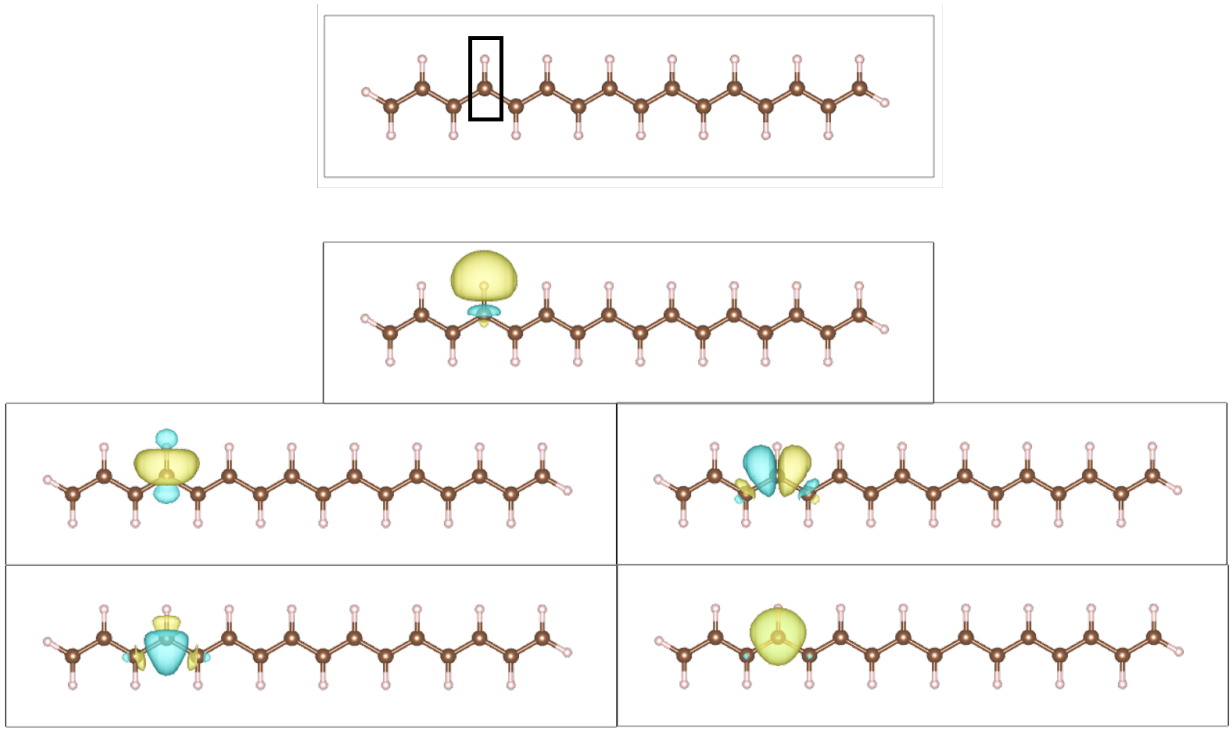

Figure 1: Sample LOs obtained through Boys localization for $C_{16}H_{18}$ E-polyacetylene in STO-3G. The top picture highlights an atomic group in this molecule and the bottom five pictures show sample LOs assigned to this atomic group.

2.1.2 Bath Orbitals

In quantum embedding approaches, orbitals outside the fragment that are entangled to the fragment are included in the calculation and obtained via the Schmidt Decomposition.[37,38,40] For a given fragment, call it F, with a defined set of FOs, the bath orbitals (BOs) are obtained as orbitals outside the fragment defined by the Schmidt Decomposition, and the environment orbitals (EOs) are the rest of the orbitals. The correlated fragment wavefunction can be written as

$$|\Psi\rangle = \left(\sum_I C_I^{\mathrm{fb},F} \left|\Psi_I^{\mathrm{fb},F}\right\rangle\right) \otimes \left|\Psi^{\mathrm{env},F}\right\rangle \tag{2}$$

The wavefunction takes a CASCI-like form consisting of the correlated fragment and bath (denoted fb) space and the uncorrelated environment (denoted env) space, playing the same role as the core in CASCI. The fb space consists of all N_f^F FOs and N_b^F BOs. Typically $N_b^F = N_f^F$, but $N_b^F < N_f^F$ is possible if there are linear dependencies among bath orbitals. $\left|\Psi_I^{\mathrm{fb},F}\right\rangle$ are all $\binom{2N_f^F+2N_b^F}{2N_f^F}$ configurations possible formed from distributing $2N_f^F$ electrons into $2N_f^F + 2N_b^F$ spin-orbitals in the fb space and $C_I^{\mathrm{fb},F}$ are the configuration interaction coefficients of each configuration. $\left|\Psi^{\mathrm{env},F}\right\rangle$ is the Slater determinant formed from the EOs, which are not correlated.

2.1.3 Bootstrap Embedding

To overcome the error introduced at the edges of the fragment, common in all embedding methods, BE utilizes overlapping fragments as illustrated in Figure 2. Each atom is used to uniquely determine the center of a fragment. The other atoms in the fragment surrounding the center make up its edge sites. With this partitioning, constraints can be applied between the fragments such that the wavefunction on edge sites matches the wavefunction of that site from a fragment for which it is the center. Fragments are systematically expanded by adding in neighboring atomic groups, up to the $n-1$ coordination shell. We denote this fragment

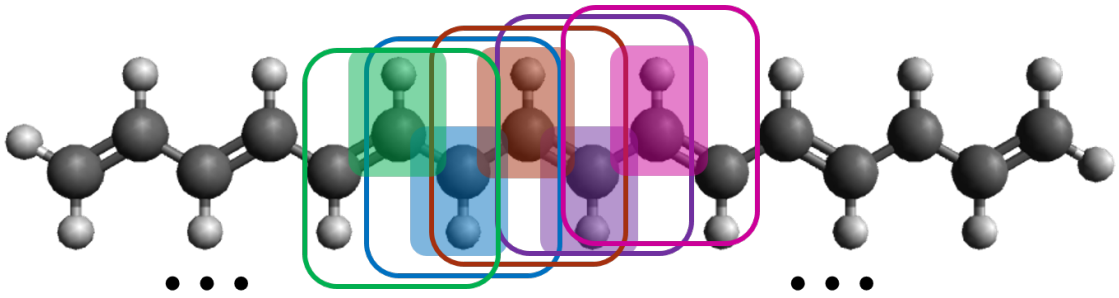

Figure 2: Illustrative example of atom-based BE. The fragments (boxed) pictured are chosen as three carbons and three hydrogens and computationally, the FO space is defined as all LOs lying on these atoms. The fragments overlap and their centers (highlighted) form a complete covering of the molecule. Take for example the red fragment in the center. The blue and purple atoms to the left and right are edge sites, and thus, the wavefunction of the red fragment must be constrained to match the blue fragment on the blue atoms, and the purple fragment on the purple atoms.

size as BE(n). Further details of the matching philosophy and fragmentation scheme can be found in prior literature.[56]

After the fragments are defined, the constraint discussed earlier is implemented. We match the wavefunctions through their one particle reduced density matrix (RDM1). For a fragment F, we can obtain the RDM1 as

$$P^F_{pq} = \left\langle \Psi^F \right| \hat{a}^{F\dagger}_p \hat{a}^F_q \left| \Psi^F \right\rangle \tag{3}$$

Let $\mathbb{C}^F$ denote the center and $\mathbb{E}^F$ denote the edge of fragment F. The matching conditions are

$$\left\langle \Psi^F \right| \hat{a}^{F\dagger}_p \hat{a}^F_q \left| \Psi^F \right\rangle = P^G_{pq} \text{ for } p, q \in \mathbb{E}^F \cap \mathbb{C}^G \tag{4}$$

We must also enforce a global constraint to ensure that the number of electrons remains consistent and correct for the full system across the fragments.

$$\sum_F^{N_{\text{frag}}} \sum_{p \in \mathbb{C}^F} \left\langle \Psi^F \right| \hat{a}^{F\dagger}_p \hat{a}^F_p \left| \Psi^F \right\rangle = N_e \tag{5}$$

Here, N_{frag} is the number of fragments and N_e is the total number of electrons in the system. In order to judge the impact of the matching conditions, we also present calculations in which density matching (Equation (4)) is not enforced, but the global number of electrons (Equation (5)) is still constrained. Calculations with (without) matching of the density are presented as BE(n) (BE(n)$_{no\ match}$) where n is the fragment size.

Minimizing the effective fragment Hamiltonian H^F, the Hamiltonian projected into the FO + BO basis, with the constraints in Equations (4) and (5) leads to the Lagrangian

$$\mathcal{L} = \sum_{F}^{N_{\text{frag}}} \left[\left\langle \Psi^F \right| \hat{H}^F \left| \Psi^F \right\rangle - \mathcal{E}^F \left(\left\langle \Psi^F \middle| \Psi^F \right\rangle - 1 \right) + \sum_{G \neq F}^{N_{\text{frag}}} \sum_{pq}^{\mathbb{E}_I^F \cap \mathbb{C}_I^G} \lambda_{pq}^F \left(\left\langle \Psi^F \right| \hat{a}_p^{F\dagger} \hat{a}_q^F \left| \Psi^F \right\rangle - P_{pq}^G \right) \right] + \mu \left[\sum_{F}^{N_{\text{frag}}} \sum_{p \in \mathbb{C}^F} \left\langle \Psi^F \right| \hat{a}_p^{F\dagger} \hat{a}_p^F \left| \Psi^F \right\rangle - N_e \right] \tag{6}$$

where λ_{pq}^F are the Lagrangian multipliers to enforce Equation (4) and μ is the Lagrangian multiplier to enforce Equation (5). Solving Equation (6) for the stationary point with respect to Ψ^F yields the following eigenvalue equation.

$$\left[\hat{H}^F + \sum_{pq}^{\mathbb{E}^F} \lambda_{pq}^F \hat{a}_p^{F\dagger} \hat{a}_q^F + \mu \sum_{p}^{\mathbb{C}^F} \hat{a}_p^{F\dagger} \hat{a}_p^F \right] \left| \Psi^F \right\rangle = \mathcal{E}^F \left| \Psi^F \right\rangle \tag{7}$$

This is the original fragment Hamiltonian in Equation (20) dressed with a local effective potential, λ_{pq}^F, and a global potential, μ. This dressed Hamiltonian exists in the whole FO and BO space, but the local potential acts on only the FO space.

The embedding energy is calculated with a cumulant-based expression summed across N_{frag} fragments, following ref. 62 and described in ref. 60. The working equations are presented in Appendix B.

2.2 BE in Extended Basis Sets

In principle, the same theory described for minimal basis sets can be applied to extended basis sets. Starting from a a larger AO basis set, the same localizations as before (e.g Boys, PM, or Löwdin) can be performed and the LOs will be assigned to fragments consistent with the size of the AO basis. The BOs can be obtained from the Schmidt Decomposition. The space of FOs and BOs are treated in a correlated calculation in the combined FO+BO space.

The results from this approach are somewhat unreliable in larger basis sets, as illustrated in Figure 3. In this figure, BE(2) in cc-pVDZ[63] with CCSD as the fragment solver is run for $C_{16}H_{18}$ E-polyacetylene (geometry in Supplementary Information). The data points labeled "Boys" indicate the convergence of BE(2) using Boys localization. It is clear that this scheme, as well as that labeled "Löwdin" referencing simple symmetric orthogonalization, fails to converge as the matching condition is never satisfied. These results motivate the exploration of other localization schemes for these more extended basis sets.

2.2.1 Fragment Orbitals

In order to understand why BE coupled with pure Boys or pure Löwdin orbitals struggles to converge, it is instructive to look at the FOs each method produces. The LOs of the same example system in the previous section are pictured in Figure 4. It is immediately apparent that the spatial extent of these orbitals is far greater than that of the LOs in Figure 1. The difficulty of localizing all the basis functions in extended basis sets is well-known.[64] Not only that, but there are some orbitals that cannot be clearly assigned to one atomic group, which is a challenge for the atom-based fragmentation scheme.

Given that the localization schemes work robustly in minimal, valence-like basis sets, one is naturally led to consider localization schemes that might leverage the valence-like functions that exist even in extended basis sets. The natural tool in this spirit is the construction of IAOs,[65] which have previously seen success in many other fragment embedding schemes.[40,66] In short, the IAOs involve projection of an extended basis set into a smaller "valence-like"

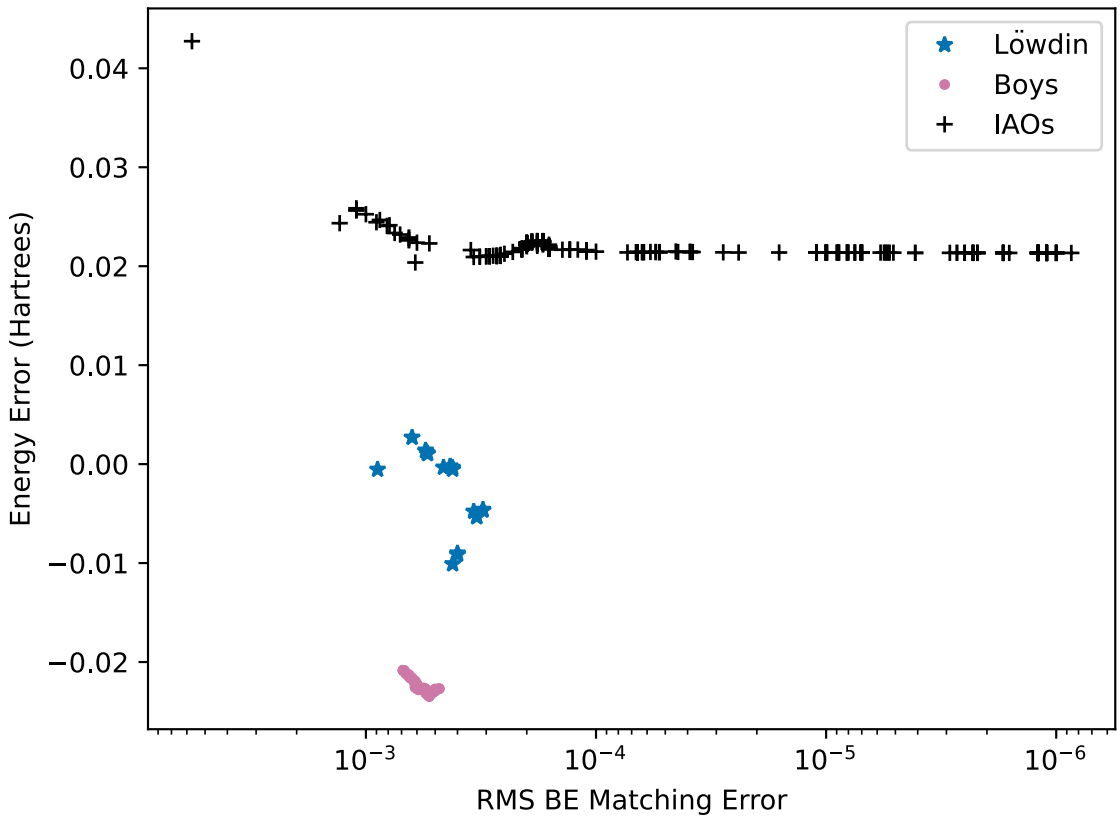

Figure 3: BE(2) error (Hartrees) versus RMS error in the BE matching condition. The system is $C_{16}H_{18}$ E-polyacetylene in cc-pVDZ. The three BE schemes listed are Löwdin (simply perfoming symmetric orthogonalization), Boys (raw Boys localization following the procedure in Section 2.1), and IAOs (the procedure in Section 2.2 matching only the IAOs). Further right indicates that the BE matching condition is closer to convergence. Each data point represents one Quasi-Newton iteration.

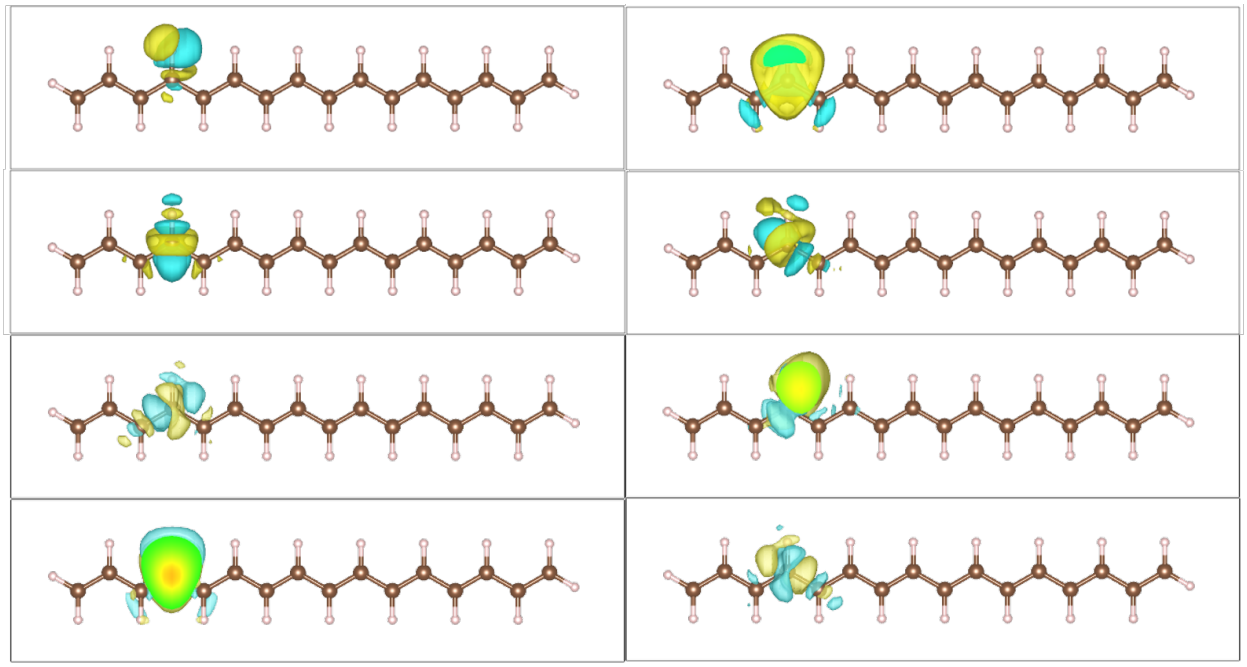

Figure 4: Sample LOs obtained through Boys localization for $C_{16}H_{18}$ E-polyacetylene in cc-pVDZ.

basis. These valence-like functions are designed to exactly span the occupied space and still resemble the minimal basis set that is easier to localize and assign to atomic fragments.

As shown in Appendix C, the IAOs are obtained by starting with a set of target valence functions $|\rho\rangle$ and defining three projection operators: one onto the full working basis set $\hat{P}_W$; one onto the valence part of the occupied space (O); and one onto the non-valence part of the occupied space ($\tilde{O}$). The IAOs are then obtained as:

$$|\bar{\rho}\rangle = \text{orth}\left(\left[\hat{\mathcal{O}}\hat{\tilde{\mathcal{O}}} + (1-\hat{\mathcal{O}})(1-\hat{\tilde{\mathcal{O}}})\right]\hat{P}_W\,|\rho\rangle\right) \tag{8}$$

The projection in the parenthesis yields the IAOs, which are further orthogonalized through Boys localization. The IAOs obtained this way are a set of orbitals the same size as and resembling the valence basis set as much as possible while still spanning the occupied space (and part of the virtual space).

Pictured in Figure 5 are two localized AOs for $C_{16}H_{18}$ E-polyacetylene using raw Boys localization of the orbital space (as seen in Figure 4) compared to their counterparts in

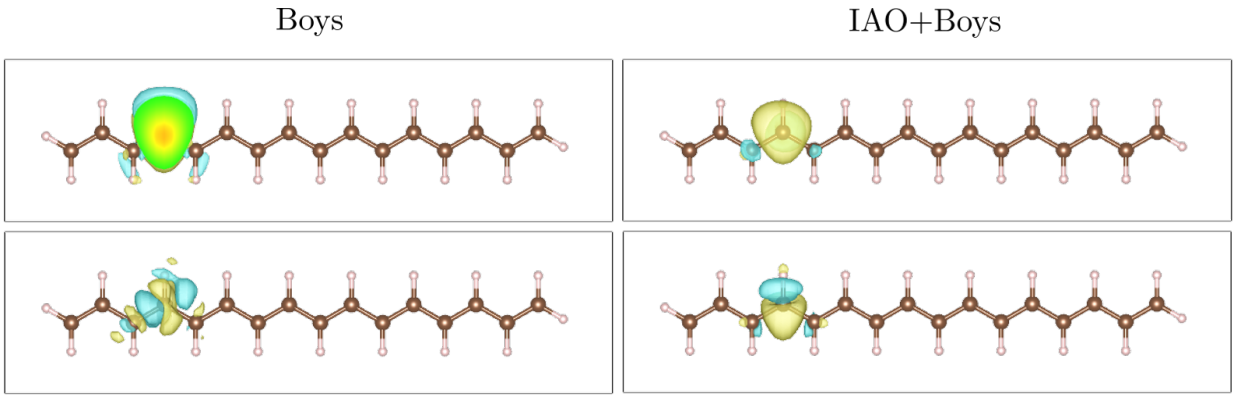

Figure 5: Two localized orbitals of E-polyacetylene ($C_{16}H_{18}$) in the cc-pVDZ basis. The two orbitals on the left are localized using Boys localization. The two orbitals on the right are obtained from using Boys localization on IAOs obtained from a STO-3G valence basis set.

the IAO space. Immediately, it can be seen that the IAO scheme yields orbitals far more localized and easier to assign to atomic groups compared to raw Boys localization.

The IAOs only span part of the orbital space. The rest of the orbital space, the part of the virtual space that is hard to localize, is included as PAOs. The PAOs[67] are simply the entire working basis with the IAOs projected out. This is accomplished by the following projections.

$$|\bar{\rho}^*\rangle = \text{orth}\left(\left[1 - \sum_{\rho} |\bar{\rho}\rangle \langle\bar{\rho}|\right]\left(1 - \hat{P}_V\right)|\mu\rangle\right) \tag{9}$$

Note that the first projection in Equation (9) will give us a redundant set of orbitals.[66] We have chosen to perform a canonical orthogonalization to get rid of redundancies, and then a Boys localization to localize the PAOs. Two example PAOs are pictured, in comparison to the corresponding naïve Boys localized orbitals discussed earlier in Figure 6. As is evident in the figure, the PAO scheme is no more effective than pure Boys at localizing these orbitals. This observation is consistent with the conventional wisdom from local electron correlation approaches, where it is known that the virtual space (an in particular the non-valence part of the virtual space) can be difficult to localize.[64] However, because we now only have difficulty in localizing these non-valence virtual orbitals (which play a smaller role in the wavefunction) one expects that the results will be somewhat more robust to the assignment of these orbitals

to different atoms.

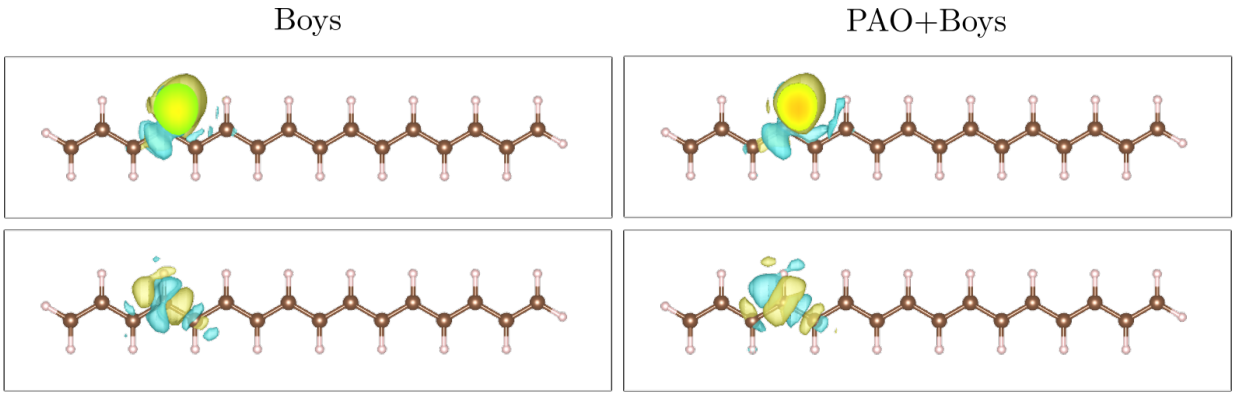

Figure 6: Two localized orbitals of E-polyacetylene ($C_{16}H_{18}$) in the cc-pVDZ basis. The two orbitals on the left are localized using Boys localization. The two orbitals on the right are obtained from using Boys localization on PAOs obtained from projecting out IAOs (with STO-3G as the valence basis set).

The set of working equations used to determine these IAOs and PAOs are detailed in Appendix C.

2.2.2 Bath Orbitals

With a set of FOs in hand, one proceeds to form the BOs using the Schmidt Decomposition as described above. We will note that, in principle, one ought to obtain an equal number of FOs and BOs out of the Schmidt Decomposition. Indeed, for minimal basis sets this almost always occurs. However, in more extended basis sets, the number of bath orbitals is almost always less than the number of fragment orbitals due to linear dependence of the BOs—an issue often referred to as bath disentanglement. In practice, we deal with this by keeping only the linearly independent BOs and performing the fragment calculation in the reduced FO+BO space. Future work will explore potential means of restoring the BOs that are "lost" in this fashion.

2.2.3 Bootstrap Embedding

The philosophy behind atom-based BE is to match atomic groups to atomic groups based on the idea that certain atomic groups are more accurately represented. This fails when orbitals cannot be definitively assigned to an atomic group, as it no longer becomes clear whether an orbital delocalized between a center and edge atomic group can be considered accurate. As demonstrated above, this is not a challenge for IAOs, but can be challenging for some of the virtual PAOs.

These concerns motivate a truncation of the matching conditions. Instead of matching all the LOs in the edge to the center, it may be helpful to only match the IAO blocks of the density matrix, as follows. Define the subset of the center and edge orbitals that are IAOs as $\mathbb{C}_I^F$ and $\mathbb{E}_I^F$ for fragment F. The new matching conditions are simply

$$\left\langle \Psi^F \right| \hat{a}_p^{F\dagger} \hat{a}_q^F \left| \Psi^F \right\rangle = P_{pq}^G \text{ for } p, q \in \mathbb{E}_I^F \cap \mathbb{C}_I^G \tag{10}$$

and yield the effective fragment Hamiltonian

$$\left[\hat{H}^F + \sum_{pq}^{\mathbb{E}_I^F} \lambda_{pq}^F \hat{a}_p^{F\dagger} \hat{a}_q^F + \mu \sum_{p}^{\mathbb{C}_I^F} \hat{a}_p^{F\dagger} \hat{a}_p^F \right] \left| \Psi^F \right\rangle = \mathcal{E}^F \left| \Psi^F \right\rangle \tag{11}$$

where the only modification is that the matching potential acts on only the IAO part of the FOs. As shown in Figure 3, this IAO scheme does indeed improve the convergence of the calculation in cases where simple Boys or Löwdin fail. We will further explore the utility of matching only the IAO part of the FO density matrix below.

3 Computational Details

All BE calculations were done using in-house software which utilizes `PySCF`[68] for electronic structure calculations and `Libint2`[69] for the integral evaluations. Any high level theory can

be used to solve each fragment Hamiltonian in Equation (7). In this work, we used CCSD as the fragment solver with the unrelaxed density matrices, and we compared the embedded results to the full-system CCSD solutions, also calculated using `PySCF`. Positive percent error indicates over-correlation and negative percent error indicates under-correlation. We note that although BE can be applied to quite large systems, our comparative results are limited by the size of systems on which we can run CCSD calculations.

BE calculations were done using the IAO and PAO scheme described in Section 2.2.1. The valence basis set used is the MinAO basis set, and Boys localization was used to orthogonalize the IAOs and PAOs. We hav e tried different choices of valence basis sets, finding that while some choices suit certain molecules, the MinAO/Boys scheme works universally well for the test molecules. Any other results using different choices for the valence basis set and localization can be found in the Supplementary Information. The working basis sets tested include STO-3G, 3-21G, 6-311G, and cc-pVDZ.

Calculations were conducted on five different classes of molecules: E-polyacetylene, fullerenes, polyglycine, nylon 6-6, and azane. The polymer molecules can be easily extended to larger sizes to test the error scaling. For E-polyacetylene, polyglycine, nylon 6-6, and azane, the geometry of each monomer unit is fixed and repeated to generate larger polymers. This ensures that larger polymers are consistent with the smaller polymers. The `xyz` files for all the geometries and visualizations of each molecule can be found in the Supplementary Information, as well as the converged BE results for all structures in each basis set.

4 Results

The BE errors compared to CCSD for the various molecules in this study are plotted in Figure 7. E-polyacetylene (Figure 7**a**) and fullerene (Figure 7**e**) are homogeneous carbon systems studied in previous works. Fullerenes present additional difficulty compared to the rest of the molecules in this study due to their two dimensional structure and their more

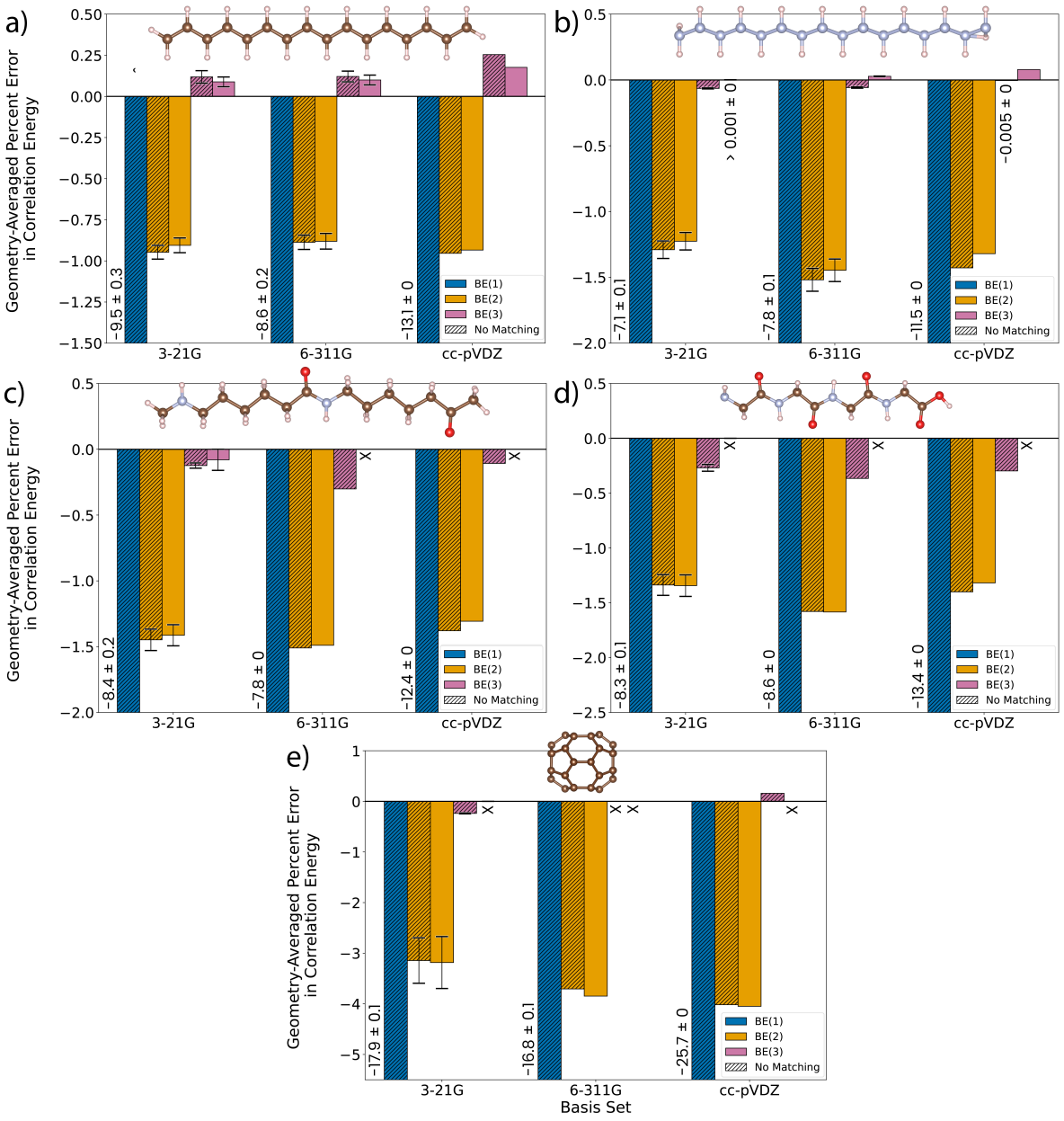

Figure 7: BE error versus CCSD, averaged across converged geometries in each basis set, for **a)** 16, 32, 48, and 64 carbon E-polyacetylene chains; **b)** 16, 32, and 48 nitrogen azane chains; **c)** 2, 4, and 6 monomer chains of nylon; **d)** 4, 8, and 12 monomer chains of polyglycine; and **e)** $C_{36} - D_{6H}$, C_{60}, and $C_{78} - D_{3H}$ fullerenes. All calculations used IAO-MINAO/Boys localization schemes.

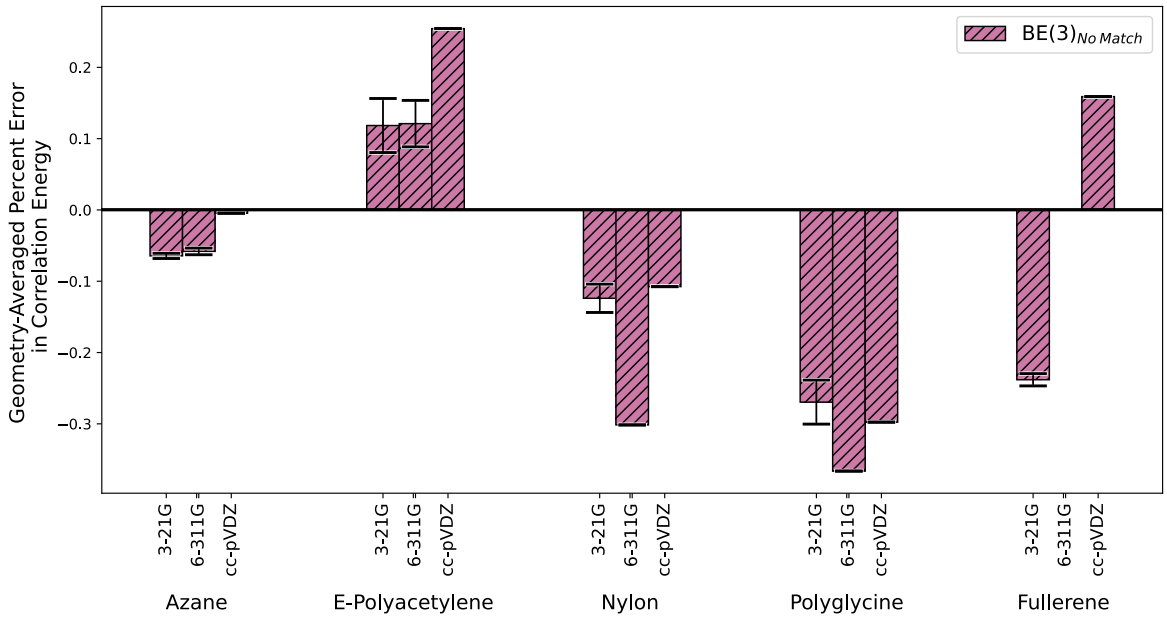

Figure 8: BE(3)$_{no\ match}$ errors versus CCSD for Azane, E-polyacetylene, Nylon, Polyglycine and Fullerene in the 3-21G, 6-311G, and cc-pVDZ basis sets. All are localized with the IAO-MINAO/Boys scheme. Errors are averaged where converged for each basis set over multiple geometries for each molecule type, and the standard deviation in this error is indicated.

complex connectivity. Azane (Figure 7**b**) is a similarly homogeneous nitrogen system. Nylon (Figure 7**c**) and polyglycine (Figure 7**d**) are two non-homogeneous systems, with less uniform fragments, used as more complex tests for BE. The 3-21G, 6-311G, and cc-pVDZ basis sets are tested.

As shown by the small variations in accuracy across the different examples, BE performs similarly across different sizes of polymeric chains and fullerenes. This demonstrates that BE is consistently reliable with increasing system size, as is the expectation of size-consistent embedding methods. The results presented here are averaged over all converged calculations for a range of sizes of each molecule type. Though each data point represents an averaged signed error (which can in principle mask large errors by cancellation of positive and negative errors), for a given molecule and basis set, the errors for a given BE(n) level are almost exclusively of the same sign. Because the sign is meaningful in differentiating over- and under-correlation, we include it here. Each individual molecular result is included in the Supplementary Information.

In the polymer test cases, BE(2)$_{no\ match}$ and BE(2) capture about 98.5% of the correlation energy, with a slightly higher error for the fullerenes. For the larger fragment BE(3)$_{no\ match}$ and BE(3), 99.7% of the CCSD correlation energy is captured for all systems in all basis sets, marking the success of BE for large basis sets, as seen in Figure 8. This combination of the IAO localization and BE framework succeeds in describing these systems with non-minimal basis sets in a systematically improvable way. In all cases, we see that the IAO-BE calculation improves with increasing fragment size. BE(3) improves the BE(2) results for all of the systems tested.

The effect of the matching condition can be seen by comparing BE(2)$_{no\ match}$ to BE(2) and BE(3)$_{no\ match}$ to BE(3). The matching conditions for these systems does not have a large impact on the accuracy of the calculated energy. For all systems and all basis sets, BE(2)$_{no\ match}$ and BE(2) perform almost indistinguishably. The matching with BE(3) sometimes improves the result over BE(3)$_{no\ match}$. However, the additional cost and difficulty converging the BE(3) matching conditions should be weighed in choosing the appropriate calculation. Matching has the greatest impact for the non-homogenous nylon and polyglycine systems, which can be rationalized with the irregularity of their fragments. Fragments in these compounds centered on oxygen, for instance, are chemically dissimilar from those centered on carbon; thus, the RDM1 description of oxygen in each fragment differs. For increasingly large and less atomically homogenous systems, matching may still prove beneficial and should be investigated more thoroughly. These results demonstrate that converging the matching conditions may not be necessary to chemical accuracy, allowing BE without matching to be useful for even these difficult cases.

Fullerenes stand out in this dataset as a particular challenge for BE. Fullerenes are pseudo two-dimensional systems, making the BE(3) prescription expensive and difficult to converge due to the rapid fragment size growth with respect to coordination shell. This problem is exacerbated in large basis sets, leading to a steep computational cost. Such larger fragment sizes, however, may be necessary to achieve desired accuracy, perhaps due

to the two-dimensional nature of these more locally correlated systems. Moving forward, an alternative scheme to increasing fragment sizes may be necessary to describe these systems at a lower computational cost.

5 Conclusion

In this paper, we extend the methodology of Schmidt Decomposition-based embedding methods, BE in particular, to the realm of extended basis sets. Our initial success in studying molecular properties with BE[56,57] can now be expanded into the regime of dynamic correlation. The main challenge for embedding methods in large basis sets is the difficulty of localizing the basis functions and defining fragments. Our remedy involves the localization of the IAO subspace and restricting all fragment matching to the IAOs. This modified BE successfully converges and captures 99.7% of the CCSD correlation energy for all systems in this study and for basis sets including 6-311G and cc-pVDZ.

While the prescription used in this paper has been successful, the questions of whether there are more effective localization schemes or ways to include more matching conditions beyond matching the IAOs could still be explored. The results in this work indicate that matching conditions are not as important as fragment size for converging to the correct correlation energy. It is possible that certain systems will require matching conditions for this level of accuracy, but we did not encounter them in this study. We also find that for highly connected systems (like the 2D fullerene systems studied here) or for basis sets that are even more dense than the ones studied here (e.g. TZ2P or QZPP basis sets), the BE(3) calculations that we would expect to give chemical accuracy are computationally prohibitive.

To improve BE calculations shown in this work, a scheme for improving BE(2) at a lower cost than full BE(3) is desirable. One potential means of doing so would be to use the cluster natural orbitals[70] that have been successful in other embedding strategies. Alternative localization schemes and perhaps alternative choices of valence basis sets for the IAO projection

could yield better results, or even improve the convergence issues with BE(3). Although much regarding BE in extended basis sets has yet to be explored, the results presented here are extraordinarily promising and suggest that with further development, embedding can become a benchmark-quality low-scaling electron correlation method making worthwhile impacts in predictive chemistry.

Acknowledgement

This work was supported by a grant from the NSF (Grant No. CHE-2154938). HKT is an NSF Graduate Research Fellow.

Supporting Information Available

The Supplementary Information contains all the geometries of the molecules in this study, convergence plots for different basis sets, and BE results on the same systems studied here using different valence orbital choices and localization techniques.

A Schmidt Decomposition for Mean-Field Wavefunctions

In this section, we detail how one obtains the fragment, bath, and environment orbitals used in the embedding calculation. The following expressions are assumed to be in some orthogonal basis of local orbitals (LOs). Define the fragment, F, and identify all LOs that belong to F. In our case, F is defined by a set of atoms and LOs are assigned to atoms based on their maximum Löwdin population.[71] Denoting the list of LOs belonging to F as I_F and the complement as J_F, we partition the rows of the occupied coefficient matrix in the LO basis ($\mathbf{C}^{\text{occ}}$) as

$$\mathbf{C}^{\text{occ}} = \begin{bmatrix} \mathbf{C}^{\text{occ}}_{I_F} \\ \mathbf{C}^{\text{occ}}_{J_F} \end{bmatrix} \tag{12}$$

In order to obtain the maximally entangled elements from the complement of F, we perform a singular value decomposition on $\mathbf{C}^{\text{occ}}_{J_F}$

$$\mathbf{C}^{\text{occ}}_{J_F} = \mathbf{U}^F \Sigma^F \mathbf{V}^{FT} \tag{13}$$

where Σ is the core tensor, which is diagonal. The diagonal elements of Σ lie between 0 and 1. 0 means fully virtually, 1 means fully occupied, and between 0 and 1 means entangled. We divide $\mathbf{U}$ in columns corresponding to orbitals that are entangled and those that are fully occupied.

$$U^F_{B;ij} = U^F_{ij} \text{ for } j \text{ such that } \Sigma^F_{jj} \neq 0, 1 \tag{14}$$

$$U^F_{E;ij} = U^F_{ij} \text{ for } j \text{ such that } \Sigma^F_{jj} = 1 \tag{15}$$

Now, the fragment, bath, and environment orbitals (FOs, BOs, EOs) can be defined. The FOs are simply the LOs in F. The BOs and EOs are the columns of $\mathbf{U}^F_B$ and $\mathbf{U}^F_E$, respectively. Hence, denoting our LO basis as $|\phi_i\rangle$, we have

$$\left|f^F_p\right\rangle = \sum_i \delta_{ip} \left|\phi_i\right\rangle \tag{16}$$

$$\left|b^F_q\right\rangle = \sum_i U^F_{B;iq} \left|\phi_i\right\rangle \tag{17}$$

$$\left|e^F_r\right\rangle = \sum_i U^F_{E;ir} \left|\phi_i\right\rangle \tag{18}$$

where $p \in I_F$. Denote the transformation from the LOs into the FOs and BOs as $\mathbf{T}^F$

$$\mathbf{T}^F = \begin{bmatrix} \mathbf{I} & 0 \\ 0 & \mathbf{U}^F_B \end{bmatrix} \tag{19}$$

With these orbital definitions in mind, we can form an effective Hamiltonian for fragment

F.

$$\hat{H}^F = \sum_{pq}^{2N_F} h_{pq}^F \hat{a}_p^{F\dagger} \hat{a}_q^F + \frac{1}{2} \sum_{pqrs}^{2N_F} V_{pqrs}^F \hat{a}_p^{F\dagger} \hat{a}_r^{F\dagger} \hat{a}_s^F \hat{a}_q^F + E^{env;F} \tag{20}$$

where

$$\mathbf{P}^{env;F} \equiv \sum_{pq} \left| e_p^F \right\rangle \left\langle e_q^F \right| \tag{21}$$

$$h_{pq}^F = \sum_{\mu\nu}^{N} T_{\mu p}^F F_{\mu\nu}^{env,F} T_{\nu q}^F \tag{22}$$

$$V_{pqrs}^F = \sum_{\mu\nu\lambda\eta}^{N} T_{\mu p}^F T_{\nu q}^F V_{\mu\nu\lambda\eta} T_{\lambda r}^F T_{\eta s}^F \tag{23}$$

$$E^{env;F} = \mathrm{Tr}\big(\boldsymbol{h} + \boldsymbol{F}^{env,F}\big)\boldsymbol{P}^{env,F} \tag{24}$$

and F^{env} is the environment Fock matrix.

B Cumulant Energy

The cumulant energy can be expressed as

$$E = E_{HF}^{[0]} + \sum_{F}^{N_{frag}} \sum_{p \in \mathbb{C}_F} \left[\sum_{q}^{2N_F} F_{pq}^{[0],F} \Delta P_{pq}^F + \frac{1}{2} \sum_{qrs}^{2N_F} V_{pqrs}^F K_{pqrs}^F \right] \tag{25}$$

where $E_{HF}^{[0]}$ is the reference system HF energy, $F^{[0],F}$ is the Fock matrix corresponding to that reference HF density in each FO + BO basis, and ΔP^F is the difference between the correlated RDM1, P^F, and the corresponding reference HF density of the fragment, $P^{HF,F}$. The approximate two-body cumulant term, K^F, is defined relative to the true two-body fragment cumulant $\tilde{K}^F$:

$$K_{pqrs}^F = \tilde{K}_{pqrs}^F + \Delta P_{pq}^F \Delta P_{rs}^F - \frac{1}{2} \Delta P_{pr}^F \Delta P_{sq}^F \tag{26}$$

The true two-body cumulant $\tilde{K}^F$ is related to the RDM2 Γ^F by

$$\Gamma^F_{pqrs} = P^F_{pq} P^F_{rs} - \frac{1}{2} P^F_{pr} P^F_{sq} + \tilde{K}^F_{pqrs} \tag{27}$$

As demonstrated in periodic-BE,[60] we note that the cumulant-based energy expression in Equation 25 provides significantly better correlation energies for our embedded calculations compared to the previously used density-matrix based energy expression[54–57] . This is also observed in ref. 62.

C IAO and PAO Working Equations

Here, we detail the working equations to obtaining the IAOs and PAOs.

We start with the working AO basis set $|\mu\rangle$ (with span B_W) and choose a "valence basis" denoted by $|\rho\rangle$ (with span B_V). This valence basis set, in our case, is some minimal basis set that we want the final IAOs to resemble.

All matrices are in the working basis set. Define the overlap matrix between the working basis and the valence basis as $\mathbf{S}_{WV}$. The projection operators are

$$\mathbf{P}_W = \mathbf{S}_W^{-1} \tag{28}$$

$$\mathbf{P}_V = \mathbf{S}_{WV} \mathbf{S}_V^{-1} \mathbf{S}_{WV}^T \tag{29}$$

Let $\mathbf{C}_{\text{occ}}$ be the occupied coefficient matrix. Begin by obtaining the depolarized occupied orbitals.

$$\tilde{\mathbf{C}}_{\text{occ}} = \mathbf{P}_W \mathbf{P}_V \mathbf{C}_{\text{occ}} \tag{30}$$

The projection into the occupied orbitals and the depolarized occupied orbitals are

$$\mathbf{P}_{\text{occ}} = \mathbf{C}_{\text{occ}}\mathbf{C}_{\text{occ}}^T \tag{31}$$

$$\tilde{\mathbf{P}}_{\text{occ}} = \tilde{\mathbf{C}}_{\text{occ}}\tilde{\mathbf{S}}^{-1}\tilde{\mathbf{C}}_{\text{occ}}^T \tag{32}$$

where

$$\tilde{\mathbf{S}} = \tilde{\mathbf{C}}_{\text{occ}}^T\mathbf{S}_W\tilde{\mathbf{C}}_{\text{occ}} \tag{33}$$

The IAOs are (Equation (8))

$$\bar{\mathbf{C}} = \text{orth}\left[(\mathbf{I} - (\mathbf{P}_{\text{occ}} + \tilde{\mathbf{P}}_{\text{occ}} - 2\mathbf{P}_{\text{occ}}\mathbf{S}_W\tilde{\mathbf{P}}_{\text{occ}})\mathbf{S}_W)\tilde{\mathbf{R}}\right] \tag{34}$$

where orth indicates Boys localization. Next we solve for the PAOs. Define

$$\mathbf{D} = \mathbf{S}_W^{-1}\mathbf{S}_{WV} \tag{35}$$

and the non-valence part of the working basis is

$$\mathbf{M}^* = \mathbf{I} - \mathbf{D}\mathbf{D}^T \tag{36}$$

The projection into the IAOs is

$$\bar{\mathbf{P}} = \bar{\mathbf{C}}\bar{\mathbf{C}}^T\mathbf{S}_W \tag{37}$$

The PAOs are (Equation (9))

$$\bar{\mathbf{C}}^* = \text{orth}\left[(\mathbf{I} - \bar{\mathbf{P}})\mathbf{M}^*\right] \tag{38}$$

These PAOs are canonically orthogonalized and then Boys localized.

References

(1) Fock, V. Z. Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems. *Zeitschrift für Physik* **1930**, *61*, 126–148.

(2) Bartlett, R. J. Many-Body Perturbation Theory and Coupled Cluster Theory for Electron Correlation in Molecules. *Ann. Rev. Phys. Chem.* **1981**, *32*, 359–401.

(3) Roos, B. O.; Taylor, P. R.; Siegbahn, P. E. M. A Complete Active Space SCF Method (CASSCF) Using a Density Matrix Formulated Super-CI Approach. *Chem. Phys.* **1980**, *48*, 157–173.

(4) Roos, B. O.; Taylor, P. R.; Sigbahn, P. E. A complete active space SCF method (CASSCF) using a density matrix formulated super-CI approach. *Chemical Physics* **1980**, *48*, 157 – 173.

(5) Siegbahn, P.; Heiberg, A.; Roos, B.; Levy, B. A Comparison of the Super-CI and the Newton-Raphson Scheme in the Complete Active Space SCF Method. *Physica Scripta* **1980**, *21*, 323–327.

(6) Siegbahn, P. E. M.; Almlöf, J.; Heiberg, A.; Roos, B. O. The complete active space SCF (CASSCF) method in a Newton–Raphson formulation with application to the HNO molecule. *The Journal of Chemical Physics* **1981**, *74*, 2384–2396.

(7) Olsen, J.; Roos, B. O.; Jo/rgensen, P.; Jensen, H. J. A. Determinant based configuration interaction algorithms for complete and restricted configuration interaction spaces. *The Journal of Chemical Physics* **1988**, *89*, 2185–2192.

(8) Malmqvist, P. A.; Rendell, A.; Roos, B. O. The Restricted Active Space Self-Consistent-Field Method, Implemented with a Split Unitary Group Approach. *J. Phys. Chem.* **1990**, *94*, 5477–5482.

(9) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618–622.

(10) Schütz, M.; Hetzer, G.; Werner, H.-J. Low-order scaling local electron correlation methods. I. Linear scaling local MP2. *The Journal of Chemical Physics* **1999**, *111*, 5691–5705.

(11) Saebo/, S.; Pulay, P. Fourth-order Mo/ller–Plessett perturbation theory in the local correlation treatment. I. Method. *The Journal of Chemical Physics* **1987**, *86*, 914–922.

(12) Schütz, M.; Werner, H.-J. Low-order scaling local electron correlation methods. IV. Linear scaling local coupled-cluster (LCCSD). *The Journal of Chemical Physics* **2001**, *114*, 661–681.

(13) Sæbø, S.; Pulay, P. Local configuration interaction: An efficient approach for larger molecules. *Chemical Physics Letters* **1985**, *113*, 13–18.

(14) Nagy, P. R.; Kállay, M. Approaching the Basis Set Limit of CCSD(T) Energies for Large Molecules with Local Natural Orbital Coupled-Cluster Methods. *Journal of Chemical Theory and Computation* **2019**, *15*, 5275–5298, PMID: 31465219.

(15) Li, W.; Li, S. Divide-and-conquer local correlation approach to the correlation energy of large molecules. *The Journal of Chemical Physics* **2004**, *121*, 6649–6657.

(16) Neese, F.; Wennmohs, F.; Hansen, A. Efficient and accurate local approximations to coupled-electron pair approaches: An attempt to revive the pair natural orbital method. *The Journal of Chemical Physics* **2009**, *130*, 114108.

(17) Saitow, M.; Becker, U.; Riplinger, C.; Valeev, E. F.; Neese, F. A new near-linear scaling, efficient and accurate, open-shell domain-based local pair natural orbital coupled cluster singles and doubles theory. *The Journal of Chemical Physics* **2017**, *146*, 164105.

(18) Sandler, I.; Chen, J.; Taylor, M.; Sharma, S.; Ho, J. Accuracy of DLPNO-CCSD(T): Effect of Basis Set and System Size. *The Journal of Physical Chemistry A* **2021**, *125*, 1553–1563, PMID: 33560853.

(19) Adler, T. B.; Werner, H.-J. Local explicitly correlated coupled-cluster methods: Efficient removal of the basis set incompleteness and domain errors. *The Journal of Chemical Physics* **2009**, *130*, 241101.

(20) Gordon, M. S.; Mullin, J. M.; Pruitt, S. R.; Roskop, L. B.; Slipchenko, L. V.; Boatz, J. A. Accurate Methods for Large Molecular Systems. *The Journal of Physical Chemistry B* **2009**, *113*, 9646–9663, PMID: 19368406.

(21) Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. Fragmentation Methods: A Route to Accurate Calculations on Large Systems. *Chemical Reviews* **2012**, *112*, 632–672, PMID: 21866983.

(22) Warshel, A.; Levitt, M. Theoretical Studies of Enzymitic Reactions - Dielectric, Electrostatic and Steric Stabilization of Carbonium-Ion in Reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103*, 227–249.

(23) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. Fragment molecular orbital method: an approximate computational method for large molecules. *Chemical Physics Letters* **1999**, *313*, 701–706.

(24) Fedorov, D. G.; Nagata, T.; Kitaura, K. Exploring chemistry with the fragment molecular orbital method. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7562–7577.

(25) Wesolowski, T. A.; Warshel, A. Frozen density functional approach for ab initio calculations of solvated molecules. *The Journal of Physical Chemistry* **1993**, *97*, 8050–8053.

(26) Wesolowski, T. A.; Shedge, S.; Zhou, X. Frozen-Density Embedding Strategy for Mul-

tilevel Simulations of Electronic Structure. *Chemical Reviews* **2015**, *115*, 5891–5928, PMID: 25923542.

(27) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864.

(28) Kohn, W.; Sham, L. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133.

(29) Severo Pereira Gomes, A.; Jacob, C. R. Quantum-chemical embedding methods for treating local electronic excitations in complex chemical systems. *Annu. Rep. Prog. Chem., Sect. C: Phys. Chem.* **2012**, *108*, 222–277.

(30) Jacob, C. R.; Neugebauer, J. Subsystem density-functional theory. *WIREs Computational Molecular Science* **2014**, *4*, 325–362.

(31) Lee, S. J. R.; Welborn, M.; Manby, F. R.; Miller, T. F. Projection-Based Wavefunction-in-DFT Embedding. *Accounts of Chemical Research* **2019**, *52*, 1359–1368.

(32) Chulhai, D. V.; Goodpaster, J. D. Projection-Based Correlated Wave Function in Density Functional Theory Embedding for Periodic Systems. *Journal of Chemical Theory and Computation* **2018**, *14*, 1928–1942, PMID: 29494155.

(33) Wen, X.; Graham, D. S.; Chulhai, D. V.; Goodpaster, J. D. Absolutely Localized Projection-Based Embedding for Excited States. *Journal of Chemical Theory and Computation* **2020**, *16*, 385–398, PMID: 31769981.

(34) Chulhai, D. V.; Goodpaster, J. D. Projection-Based Correlated Wave Function in Density Functional Theory Embedding for Periodic Systems. *Journal of Chemical Theory and Computation* **2018**, *14*, 1928–1942, PMID: 29494155.

(35) 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. *Journal of Chemical Theory and Computation* **2020**, *16*, 2284–2295, PMID: 32105469.

(36) Cohen, A. J.; Mori-Sanchez, P.; Yang, W. Insights into Current Limitations of Density Functional Theory. *Science* **2008**, *321*, 792–794.

(37) Knizia, G.; Chan, G. K.-L. Density Matrix Embedding: A Simple Alternative to Dynamical Mean-Field Theory. *Phys. Rev. Lett.* **2012**, *109*, 186404.

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

(39) Zheng, B.-X.; Chan, G. K.-L. Ground-state phase diagram of the square lattice Hubbard model from density matrix embedding theory. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2016**, *93*, 035126.

(40) Wouters, S.; Jimenez-Hoyos, C. A.; Sun, Q.; Chan, G. K.-L. A Practical Guide to Density Matrix Embedding Theory in Quantum Chemistry. *J. Chem. Theory Comput.* **2016**, *12*, 2706–2719.

(41) Tsuchimochi, T.; Welborn, M.; Van Voorhis, T. Density matrix embedding in an antisymmetrized genimal power bath. *J. Chem. Phys.* **2015**, *143*, 024107.

(42) Chen, Q.; Booth, G. H.; Sharma, S.; Knizia, G.; Chan, G. K.-L. Intermediate and spin-liquid phase of the half-filled honeycomb Hubbard model. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2014**, *89*, 165134.

(43) Booth, G. H.; Chan, G. K.-L. Spectra Functions of strongly correlated extended systems via an exact quantum embedding. *Phys. Rev. B* **2015**, *91*, 155107.

(44) Tran, H. K.; Van Voorhis, T.; Thom, A. J. W. Using SCF metadynamics to extend density matrix embedding theory to excited states. *The Journal of Chemical Physics* **2019**, *151*, 034112.

(45) Bulik, I. W.; Chen, W.; Scuseria, G. E. Intermediate and spin-liquid phase of the half-filled honeycomb Hubbard model. *J. Chem. Phys.* **2014**, *141*, 054113.

(46) Hermes, M. R.; Gagliardi, L. Multiconfigurational Self-Consistent Field Theory with Density Matrix Embedding: The Localized Active Space Self-Consistent Field Method. *Journal of Chemical Theory and Computation* **2019**, *15*, 972–986, PMID: 30620876.

(47) Hermes, M. R.; Gagliardi, L. Multiconfigurational Self-Consistent Field Theory with Density Matrix Embedding: The Localized Active Space Self-Consistent Field Method. *Journal of Chemical Theory and Computation* **2019**, *15*, 972–986, PMID: 30620876.

(48) Hermes, M. R.; Pandharkar, R.; Gagliardi, L. Variational Localized Active Space Self-Consistent Field Method. *Journal of Chemical Theory and Computation* **2020**, *16*, 4923–4937, PMID: 32491849.

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

(50) Senjean, B. Projected site-occupation embedding theory. *Phys. Rev. B* **2019**, *100*, 035136.

(51) He, N.; Li, C.; Evangelista, F. A. Second-Order Active-Space Embedding Theory. *Journal of Chemical Theory and Computation* **0**, *0*, null, PMID: 35175028.

(52) Welborn, M.; Tsuchimochi, T.; Van Voorhis, T. Bootstrap embedding: An internally consistent fragment-based method. *J. Chem. Phys.* **2016**, *145*, 074102.

(53) Ricke, N.; Welborn, M.; Ye, H.-Z.; Van Voorhis, T. Performance of Bootstrap Embedding for long-range iinteraction and 2D systems. *Mol. Phys.* **2017**, 1–12.

(54) Ye, H.-Z.; Welborn, M.; Ricke, N. D.; Van Voorhis, T. Incremental embedding: A density matrix embedding scheme for molecules. *The Journal of Chemical Physics* **2018**, *149*, 194108.

(55) Ye, H.-Z.; Ricke, N. D.; Tran, H. K.; Van Voorhis, T. Bootstrap Embedding for Molecules. *Journal of Chemical Theory and Computation* **2019**, *15*, 4497–4506, PMID: 31343878.

(56) Ye, H.-Z.; Van Voorhis, T. Atom-Based Bootstrap Embedding For Molecules. *The Journal of Physical Chemistry Letters* **2019**, *10*, 6368–6374, PMID: 31578867.

(57) Ye, H.-Z.; Tran, H. K.; Van Voorhis, T. Bootstrap Embedding For Large Molecular Systems. *Journal of Chemical Theory and Computation* **2020**, *16*, 5035–5046, PMID: 32589842.

(58) Tran, H. K.; Ye, H.-Z.; Van Voorhis, T. Bootstrap embedding with an unrestricted mean-field bath. *The Journal of Chemical Physics* **2020**, *153*, 214101.

(59) Ye, H.-Z.; Tran, H. K.; Van Voorhis, T. Accurate Electronic Excitation Energies in Full-Valence Active Space via Bootstrap Embedding. *Journal of Chemical Theory and Computation* **2021**, *17*, 3335–3347, PMID: 33957050.

(60) Meitei, O. R.; Van Voorhis, T. Periodic Bootstrap Embedding. *Journal of Chemical Theory and Computation* **2023**, *19*, 3123–3130.

(61) Foster, J. M.; Boys, S. F. Quantum Variational Calculations for a Range of CH_2 Configurations. *Rev. Mod. Phys.* **1960**, *32*, 305.

(62) Nusspickel, M.; Ibrahim, B.; Booth, G. H. Effective Reconstruction of Expectation Values from Ab Initio Quantum Embedding. *Journal of Chemical Theory and Computation* **2023**, *19*, 2769–2791.

(63) 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.

(64) Subotnik, J. E.; Dutoi, A. D.; Head-Gordon, M. Fast localized orthonormal virtual orbitals which depend smoothly on nuclear coordinates. *The Journal of Chemical Physics* **2005**, *123*, 114108.

(65) Knizia, G. Intrinsic Atomic Orbitals: An Unbiased Bridge between Quantum Theory and Chemical Concepts. *Journal of Chemical Theory and Computation* **2013**, *9*, 4834–4843, PMID: 26583402.

(66) Cui, Z.-H.; Zhu, T.; Chan, G. K.-L. Efficient Implementation of Ab Initio Quantum Embedding in Periodic Systems: Density Matrix Embedding Theory. *Journal of Chemical Theory and Computation* **2020**, *16*, 119–129, PMID: 31815466.

(67) Pulay, P. Localizability of dynamic electron correlation. *Chemical Physics Letters* **1983**, *100*, 151–154.

(68) 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.; Wouters, S.; Chan, G. K.-L. PySCF: the Python-based simulations of chemistry framework. *WIREs Computational Molecular Science* **2018**, *8*, e1340.

(69) Valeev, E. F. Libint: A library for the evaluation of molecular integrals of many-body operators over Gaussian functions. http://libint.valeyev.net/, 2021; version 2.7.1.

(70) Nusspickel, M.; Booth, G. H. Systematic improvability in quantum embedding for real materials. 2022.

(71) Löwdin, P. On the Non-Orthogonality Problem Connected with the Use of Atomic Wave

Functions in the Theory of Molecules and Crystals. *The Journal of Chemical Physics* **1950**, *18*, 365–375.