The Journal of Chemical Physics

Even-handed subsystem selection in projection-based embedding

Cite as: J. Chem. Phys. **149**, 144101 (2018); https://doi.org/10.1063/1.5050533
Submitted: 31 July 2018 . Accepted: 07 September 2018 . Published Online: 08 October 2018

Matthew Welborn, Frederick R. Manby, and Thomas F. Miller

COLLECTIONS

This paper was selected as Featured

View Online Export Citation CrossMark

ARTICLES YOU MAY BE INTERESTED IN

A universal density matrix functional from molecular orbital-based machine learning: Transferability across organic molecules
The Journal of Chemical Physics **150**, 131103 (2019); https://doi.org/10.1063/1.5088393

Perspective: Computational chemistry software and its advancement as illustrated through three grand challenge cases for molecular science
The Journal of Chemical Physics **149**, 180901 (2018); https://doi.org/10.1063/1.5052551

Communication: A mean field platform for excited state quantum chemistry
The Journal of Chemical Physics **149**, 081101 (2018); https://doi.org/10.1063/1.5045056

Lock-in Amplifiers up to 600 MHz
Watch

© 2018 Author(s).

Even-handed subsystem selection in projection-based embedding

Matthew Welborn,[1] Frederick R. Manby,[2] and Thomas F. Miller III[1]
[1]*Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, USA*
[2]*Centre for Computational Chemistry, School of Chemistry, University of Bristol, Bristol BS8 1TS, United Kingdom*

(Received 31 July 2018; accepted 7 September 2018; published online 8 October 2018)

Projection-based embedding offers a simple framework for embedding correlated wavefunction methods in density functional theory. Partitioning between the correlated wavefunction and density functional subsystems is performed in the space of localized molecular orbitals. However, during a large geometry change—such as a chemical reaction—the nature of these localized molecular orbitals, as well as their partitioning into the two subsystems, can change dramatically. This can lead to unphysical cusps and even discontinuities in the potential energy surface. In this work, we present an even-handed framework for localized orbital partitioning that ensures consistent subsystems across a set of molecular geometries. We illustrate this problem and the even-handed solution with a simple example of an S_N2 reaction. Applications to a nitrogen umbrella flip in a cobalt-based CO_2 reduction catalyst and to the binding of CO to Cu clusters are presented. In both cases, we find that even-handed partitioning enables chemically accurate embedding with modestly sized embedded regions for systems in which previous partitioning strategies are problematic. *Published by AIP Publishing.*
https://doi.org/10.1063/1.5050533

I. INTRODUCTION

Embedding methods for electronic structure offer the possibility of accurate description of chemical reactions at greatly reduced computational cost by treating different parts of a chemical system with different levels of theory. Many versions of embedding exist, including implicit solvent,[1–3] QM/MM[4–6] and ONIOM,[7] subsystem density functional theory (DFT) and frozen density embedding,[8–24] active space methods,[25,26] local correlation treatments,[27,28] Green's function embedding,[29–31] and density matrix embedding theory.[32,33] Central to all of these methods is the question of how the system is to be partitioned into a number of subsystems to be treated at different levels of theory.

Prescriptions for partitioning the system usually consider a single geometry at a time. However, this approach may fail when considering a chemical reaction in which the nature of the embedded subsystem changes across the reaction coordinate. For example, this problem can arise in QM/MM when MM atoms wander into the QM region.[6] In active space methods, this issue can manifest as the intruder state problem.[34] Recently, in the context of active space methods, progress has been made in the automated selection of active spaces based on a user-specified set of relevant atomic orbitals (AO)[35] or based on a correlated wavefunction ansatz less expensive than the full active space method.[36–39] These methods have demonstrated a robust ability to select consistent active spaces across reaction coordinates.

In this work, we present a method for handling the subsystem inconsistency problem in the context of wavefunction-in-DFT embedding, where a subset of localized occupied molecular orbitals (LMOs) is selected for embedded wavefunction treatment. We begin with a previously established charge-based criterion for the automated selection of these embedded orbitals. We next demonstrate how—even for a simple S_N2 reaction—this procedure can result in a set of embedded LMOs which is inconsistent with respect to the reactant and product. Drawing inspiration from the domain merging method for local correlation,[40] we propose an "even-handed" LMO selection procedure. Our method seeks to form a consensus set of LMOs which contains for every geometry every orbital that is important for any geometry. This results in an automatic procedure which uses information available at the DFT level and requires no user input beyond the set of atoms to be embedded. Although we present this method in the context of projection-based embedding specifically,[20,22] the methodology is sufficiently general to be applied in other embedding contexts.

II. THEORY

A. Projection-based embedding

We begin by reviewing the projection-based embedding method, which is a rigorous framework for embedding a correlated wavefunction theory in a mean field (MF) theory such as Hartree Fock (HF) theory or DFT, with interactions between the embedded and embedding densities treated at the MF level. The system is partitioned into two subsystems: subsystem A is generally treated at the correlated wavefunction level and subsystem B is treated at the MF level. Projection-based embedding belongs to a class of methods which ensure orthogonality between the MF description of subsystems A and B.[20,22] This class includes frozen-core approximations,[41] the region

method for local correlation,[42] Henderson's coupled cluster in the DFT embedding method,[43] Khait and Hoffman's modified Kohn-Sham equations with Constrained Electron Density (KSCED),[44] the Huzinaga projection operator method,[45] and more recently the multi-level Hartree Fock[46] and orthogonality constrained basis set expansion (OCBSE) methods.[47]

As inputs to a projection-based calculation, we specify a low-level MF method, a high-level correlated wavefunction method, and a set of atoms to be embedded, $\{X\}_A$. The algorithm begins by performing a MF calculation on the full system, resulting in a set of occupied molecular orbitals, which are then rotated to form a set of LMOs, $\{\psi\}$. The LMOs are partitioned into two sets, $\{\psi\}_A$ and $\{\psi\}_B$, corresponding to subsystems A and B, respectively. This partitioning is usually performed by choosing $\{\psi\}_A$ to be the set of LMOs with sufficient population on $\{X\}_A$ (detailed in Sec. II B). The wave function calculation is then performed with the number of electrons necessary to occupy the $\{\psi\}_A$ and a modified core Hamiltonian[20,22] written in the atomic orbital (AO) basis as

$$\mathbf{h}^{A\,\text{in}\,B} = \mathbf{h} + \mathbf{g}\left[\boldsymbol{\gamma}^A + \boldsymbol{\gamma}^B\right] - \mathbf{g}\left[\boldsymbol{\gamma}^A\right] + \mu\mathbf{P}^B, \tag{1}$$

where $\mathbf{h}$ is the core Hamiltonian of the full system, $\boldsymbol{\gamma}^A$ and $\boldsymbol{\gamma}^B$ are the one-particle reduced density matrices (1RDMs) corresponding to $\{\psi\}_A$ and $\{\psi\}_B$, respectively, and $\mathbf{g}$ includes all mean-field two-electron interactions. The last term enforces orthogonality between the subsystems using a level shift, μ, and a projector onto subsystem B,

$$\mathbf{P}^B = \mathbf{S}\boldsymbol{\gamma}^B\mathbf{S}, \tag{2}$$

where $\mathbf{S}$ is the AO overlap matrix. In the limit $\mu \to \infty$, subsystems A and B are exactly orthogonal; in practice, this procedure is numerically robust for μ between 10^4 and 10^7 hartree, and $\mu = 10^6$ hartree is usually sufficient to converge the embedding to a type-in-type error of less than a microhartree.[20,22]

B. Charge method for LMO selection

In this subsection, we present the current "charge" algorithm that has previously proven effective for partitioning the LMOs into $\{\psi\}_A$ and $\{\psi\}_B$.[20,48] We then construct a simple example where this method breaks down, motivating the "even-handed" solution presented in Sec. II C.

The chemically intuitive notion of embedded atoms must be translated into a specific set of LMOs, which do not fully reside on specific atoms. The charge method includes in $\{\psi\}_A$ all occupied LMOs with significant population on the $\{X\}_A$. This selection is performed using a charge threshold, q,

$$\begin{aligned}\{\psi\}_A &= \{\psi_i | Q_A(\psi_i) > q\},\\ \{\psi\}_B &= \{\psi_i | Q_A(\psi_i) \leq q\},\end{aligned} \tag{3}$$

where $Q_A(\psi_i)$ is the charge of LMO ψ_i assigned to the atoms in $\{X\}_A$ and q is typically chosen to be 0.4. For example, charges may be computed using the gross Mulliken population[49]

$$Q_A(\psi_i) = \sum_{X_j \in \{X\}_A} \sum_{\lambda\,\text{on}\,X_j} \sum_{\kappa} \gamma^{\psi_i}_{\lambda\kappa} S_{\kappa\lambda}, \tag{4}$$

where κ indexes all AOs, λ on X_j indexes AOs λ centered on atom X_j, $\mathbf{S}$ is the AO overlap matrix, and $\boldsymbol{\gamma}^{\psi_i}$ is the 1RDM corresponding to ψ_i. [Alternative population schemes, such as intrinsic bond orbital (IBO),[50] can be applied similarly.] For a single-point energy calculation, this algorithm generally selects a reasonable $\{\psi\}_A$. However, because it considers only a single geometry at a time, problems can arise during large molecular geometry changes.

One problem is that the number of orbitals in $\{\psi\}_A$ selected by the charge method can differ at different molecular geometries. Because of the difference in electron chemical potentials between the high-level method of subsystem A and the low-level method of subsystem B, a change in the number of LMOs in $\{\psi\}_A$ versus $\{\psi\}_B$ manifests as a discontinuity in the potential energy. This problem is easily remedied by choosing $\{\psi\}_A$ to be the N orbitals corresponding to the N largest values of Q_A at each geometry; N is chosen to be the largest size of $\{\psi\}_A$ selected by Eq. (3) for any geometry along the reaction coordinate. In the following, all results presented for the charge method include this simple fix.

A second—and less trivial—problem occurs when LMO populations on the embedded atoms change qualitatively between geometries. Consider, for example, the S_N2 reaction of I^- with bromomethane with $\{X\}_A$ chosen to be the carbon atom. Figure 1 shows the reaction profile for coupled cluster singles and doubles with perturbative triples, CCSD(T),[51] embedded in the B3LYP density functional,[52–55] using LMOs selected by the charge method. The result of this selection can be seen in the embedding potential energy surface for this reaction. Near either the reactant or the product, the CCSD(T)-in-B3LYP energy profile follows the CCSD(T) result. However, as the transition state is crossed, an unphysically high barrier appears in the energy profile.

Figure 2 shows the embedded density, ρ_A, that results from the charge selection method (top panel, highlighted in yellow). At geometries near the reactant, ρ_A resembles the density of bromomethane; near the product, ρ_A resembles iodomethane. In the transition region, ρ_A switches abruptly between these two qualitatively different densities. Examining the orbitals $\{\psi\}_A$ that comprise ρ_A at each geometry (bottom panel, yellow), we see that the reactant, transition state, and product differ on whether the LMOs associated with the C–Br and C–I σ bonds are included in $\{\psi\}_A$, with reactant-like geometries including the C–Br bond LMO and excluding the C–I

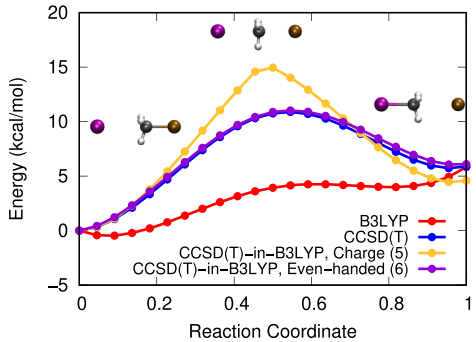

FIG. 1. Energy profiles for the S_N2 reaction of I^- with bromomethane, comparing B3LYP, CCSD(T), and embedded CCSD(T)-in-B3LYP using charge and even-handed LMO selection. The set of embedded atoms, $\{X\}_A$, contains only the carbon atom. All curves are referenced to an energy of zero at the reactant geometry. The number of occupied LMOs in subsystem A is given in parentheses.

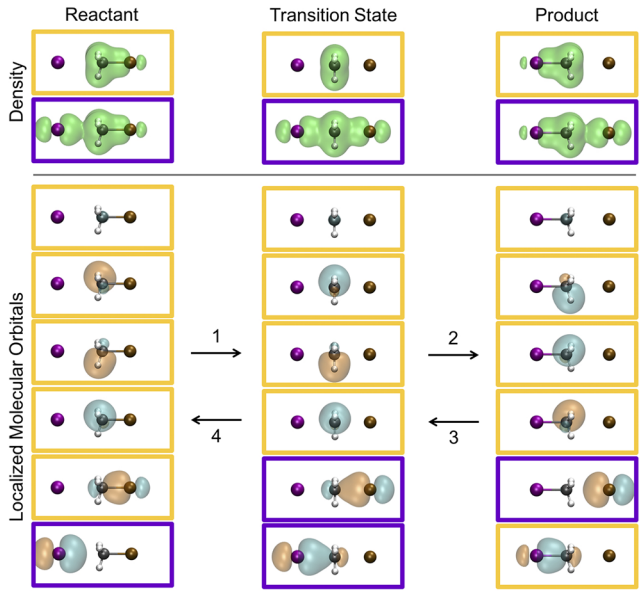

FIG. 2. Illustration of the even-handed selection method applied to the S_N2 reaction of I^- with bromomethane, where $\{X\}_A$ is the carbon atom. Above: the subsystem A densities, ρ^A, resulting from the charge (even-handed) method are boxed in yellow (purple). Below: the set of LMOs selected for inclusion in subsystem A, $\{\psi\}_A$, which comprise the corresponding ρ^A above. In the reactant state, the charge method $\{\psi\}_A$ includes the LMO corresponding to the C–Br bond, but not that corresponding to the C–I bond. In the product state, the latter is included in $\{\psi\}_A$ while the former is excluded. Near the transition state, neither orbital is included. Sweeping from the reactant to product and back and taking the union of charge method LMOs between neighboring geometries results in the even-handed set of subsystem A orbitals.

bond LMO, and vice versa for product-like geometries. In the transition region, the abrupt change in the orbital character from C–Br to C–I results in an incorrectly high reaction barrier. Although this problem is presented in the context of projection-based embedding, we note that it is common to any embedding method which relies on atomic populations to partition orbitals.

C. Even-handed LMO selection

This problem can be solved by ensuring that $\{\psi\}_A$ spans a consistent volume of Hilbert space as a function of geometry and thus provides a consistent description of γ_A throughout a reaction profile. The solution requires specification of (i) a strategy for determining what the $\{\psi\}_A$ should be and (ii) a method for quantitatively comparing $\{\psi\}_A$ between geometries.

To address the first point, we propose an "even-handed" procedure. Starting with a reaction coordinate composed of an ordered set of geometries, the charge selection procedure is performed at every geometry. We then form the union of these charge-selected orbitals. To do this, we sweep from the reactant to product and back. For each geometry k along this sweep, the corresponding $\{\psi\}_A^k$ is augmented to include any orbitals that both match a member of the $\{\psi\}_A^{k-1}$ corresponding to the previous geometry (on the basis of the method for quantitatively comparing orbitals between geometries, described later) and that are not already included in $\{\psi\}_A^{k-1}$.

Figure 2 illustrates the application of this procedure to the aforementioned example S_N2 reaction. For visual simplicity, we consider three geometries: the reactant R, the transition state T, and the product P. Four charge-selected LMOs are common to $\{\psi\}_A^R$, $\{\psi\}_A^T$, and $\{\psi\}_A^P$. Sweeping along the reaction coordinate from R to P, we begin by comparing $\{\psi\}_A^R$ to $\{\psi\}_A^T$. $\{\psi\}_A^R$ contains a C–Br σ bond LMO that $\{\psi\}_A^T$ lacks; the matching LMO in $\{\psi\}^T$ is thus added to $\{\psi\}_A^T$, resulting in a five orbital set (labeled in the figure as step 1). Comparing this updated $\{\psi\}_A^T$ to $\{\psi\}_A^P$, we see that $\{\psi\}_A^P$ lacks a C–Br σ bond LMO and contains an additional C–I σ bond LMO absent in $\{\psi\}^T$; the matching C–Br σ bond LMO is thus added to $\{\psi\}_A^P$, resulting in a six orbital set (labeled step 2). We now sweep backwards from P to R. $\{\psi\}_A^P$ is compared $\{\psi\}_A^T$, and the latter is found to lack the C–I σ bond LMO, which is added to form a new $\{\psi\}_A^T$ (labeled step 3). Finally, $\{\psi\}_A^T$ is compared $\{\psi\}_A^R$ and the C–I σ bond LMO is added to the latter (labeled step 4). The result is that for all three geometries, $\{\psi\}_A$ contains the four methyl LMOs, the C–Br σ bond LMO, and the C–I σ bond LMO. Examining the ρ_A resulting from even-handed LMO selection (Fig. 2, top panel, purple), we see that it remains consistent from the reactant to the transition state to the product, especially in comparison to the ρ_A obtained from the charge selection method (yellow). Figure 1 shows that the even-handed scheme corrects the error in the CCSD(T)-in-B3LYP energy profile generated using the charge selection method, yielding a CCSD(T)-in-B3LYP energy profile in quantitative agreement with a CCSD(T) calculation on the full system. In Sec. III, we will examine the relative performance of even-handed selection versus charge selection with regard to a given increase in the size of $\{\psi\}_A$.

To complete the even-handed LMO selection algorithm, we must define a way to quantitatively compare LMOs $\{\psi\}_A^k$ and $\{\psi\}_A^{k+1}$ associated with neighboring geometries along the reaction coordinate. The LMOs for geometry k do not directly correspond to those of $k+1$; they have different basis functions (due to motion of the atom-centered gaussians) and different LMO coefficients. The latter problem is addressed with a maximum overlap formalism, similar to that used in non-Aufbau MF calculations.[56] Specifically, for geometry $k+1$, we form a new subset of $\{\psi\}^{k+1}$ which contains the M orbitals that best overlap with the span of the M orbitals so far included in subsystem A for geometry k, $\{\psi\}_A^k$. For each LMO $\psi_i \in \{\psi\}^{k+1}$, we assign an overlap with the span of $\{\psi\}_A^k$,

$$o_i^{k+1} = \sum_j^{N_{occ}} \left(\sum_{\lambda\kappa}^{N_{AO}} L_{\lambda i}^k S_{\lambda\kappa}^{k,k+1} L_{\kappa j}^{k+1} \right)^2, \tag{5}$$

where N_{AO} is the number of basis functions, N_{occ} is the number of occupied orbitals, $\mathbf{L}^k$ is the LMO coefficient matrix corresponding to geometry k, and $\mathbf{S}^{k,k+1}$ is an overlap matrix between the AO bases corresponding to the two geometries.

Because the two geometries do not share the same AO basis, the precise definition of $\mathbf{S}^{k,k+1}$ is not immediately obvious. We might choose the spatial overlap of the AOs;

however, due to the exponential tails of gaussian basis functions, geometries must be very closely spaced to ensure numerical stability. Instead, we compare the orbitals between two different geometries as though the basis had not moved. In order to correct for changes in the AO overlap, we introduce Löwdin orthogonalization, resulting in

$$\mathbf{S}^{k,k+1} = \left(\mathbf{S}^{k}\right)^{\frac{1}{2}} \left(\mathbf{S}^{k+1}\right)^{\frac{1}{2}}, \tag{6}$$

where $\mathbf{S}^k$ and $\mathbf{S}^{k+1}$ are the AO overlap matrices for geometries k and $k+1$. As we sweep back and forth across the reaction coordinate, at each geometry $k+1$ we form a new set of LMOs $\{\psi\}_A^{k+1}$ which contains the M LMOs in $\{\psi\}^{k+1}$ corresponding to the M largest o_i as well as any charge-selected orbitals missed by this criterion. The even-handed selection algorithm is summarized in Algorithm 1.

A potential concern is that the overlap of Eq. (6) could become inaccurate if neighboring geometries are spatially distant. A useful diagnostic of whether two neighboring geometries are too distant can be formulated by comparing the Mth largest o_i to the $(M+1)$th largest o_i of Eq. (5) after the even-handed procedure is complete. This corresponds to the difference in overlap between the least overlapping LMO included in $\{\psi\}_A$ and most overlapping LMO excluded from $\{\psi\}_A$. If this difference becomes small, additional geometry points may be added along the reaction coordinate. These interpolating geometries require only MF calculations and thus have negligible impact on the overall cost of the calculation. For the systems studied in this work, this problem does not arise; for the energy profiles presented, we choose the density of interpolating geometries to illustrate the smoothness of energy profiles rather than to minimize the number of interpolating geometries.

It is worth emphasizing that the even-handed LMO selection procedure only uses information from MF calculations and only involves computation of orbital overlaps; even-handed LMO selection thus has a negligible effect on the overall cost of a projection-based wavefunction-in-DFT embedding calculation. It is also worth noting that even-handed LMO selection introduces no additional parameters beyond that appearing in the original charge method for LMO selection and the choice of the reaction pathway.

Algorithm 1. Even-handed LMO selection.

1: **Input**: LMOs $\{\psi\}^k$ for each geometry k, $k = 1, \ldots, N$.
2: For each geometry, select $\{\psi\}_A^k$ using the charge method [Eq. (3)].
3: **for** $k = 1, \ldots, N-1$ **do**
4: Compute o_i^{k+1} using Eq. (5).
5: Let M be the number of LMOs in $\{\psi\}_A^k$.
6: Construct $\{\psi\}_{\text{EH}}^{k+1}$ as the M LMOs in $\{\psi\}^{k+1}$ corresponding to the M largest o_i^{k+1}.
7: Set $\{\psi\}_A^{k+1} = \{\psi\}_A^{k+1} \cup \{\psi\}_{\text{EH}}^{k+1}$.
8: **end for**
9: Repeat the above "for loop" with the order of the geometries reversed.
10: Repeat the forward and reverse "for loops" until the number of selected LMOs converges, typically after one cycle.

D. Even-handed AO truncation

Up to this point, we have provided a framework for reducing the number of occupied orbitals in the embedded calculation while retaining all virtual orbitals. However, correlated wavefunction calculations usually scale more strongly with the number of virtual orbitals than occupied. Noting that the number of virtual orbitals increases with the size of the AO basis set, a useful approach is to remove the AOs least necessary to represent $\boldsymbol{\gamma}^A$.[48,57] A charge-based heuristic, similar in spirit to Eq. (4), is used to select which AOs are necessary to represent $\boldsymbol{\gamma}^A$ and therefore retained. And similar to charge selection above, this heuristic can lead to different basis sets for different geometries. To correct this error, for all geometries, we retain the union of all AOs heuristically retained at each geometry. That is, at every geometry, we use the same truncated basis set, $\{\lambda\}$,

$$\{\lambda\} = \bigcup_k \{\lambda\}^k, \tag{7}$$

where $\{\lambda\}^k$ is the set of AOs retained at geometry k. In the systems presented in this work which contain up to 1772 basis functions, this procedure results in retention of no more than 20 additional basis functions compared to standard AO truncation.[48]

III. RESULTS AND DISCUSSION

A. Intramolecular hydrogen bonding in a CO_2 reduction complex

As a first example of even-handed partitioning, we consider a cobalt aminopyridine complex that has been shown to catalyze the two-electron two-proton reduction of carbon dioxide to carbon monoxide.[58,59] Experimental and DFT studies on analogous Co, Fe, and Ni complexes suggest that CO_2 is stabilized by intramolecular hydrogen bonds from pendant protons and that these protons may transfer to CO_2 during the reduction process.[60–62] However, crystallography on this particular aminopyridine complex suggests that, prior to CO_2 binding, the pendant protons point away from the binding pocket, necessitating a conformational rearrangement in order for intramolecular hydrogen bonds to form. The energetics of this rearrangement provide mechanistic insight into the catalytic role of hydrogen-bonding and intramolecular proton transfer in this system. However, the energetics of this rearrangement presents challenges for DFT; although widely used for Co-based systems,[63–72] DFT can be inaccurate for reaction barriers, particularly those involving multireference character.[73–77] Furthermore, the geometry and strength of hydrogen bonds can be strongly functional dependent.[78,79]

Here, we employ projection-based wavefunction-in-DFT embedding to better capture these physical effects with a correlated wavefunction. This approach has previously been used to study a similar Co-centered hydrogen evolution catalyst.[80] We specifically consider the conformational change associated with forming one intramolecular hydrogen bond. This rearrangement requires umbrella flipping of a pendant nitrogen center and rotation of the CO_2 ligand about the dihedral

axis of the Co–C bond. A guess minimum energy path is computed at the B3LYP/6-31+G*[81] level with a Lebedev (75 302) exchange-correlation grid[82] using the freezing string method[83] as implemented in the QCHEM 4.1 software package[84] with 15 nodes and 3 gradient descent steps. The system is a doublet and treated with an unrestricted reference. The path is refined by relaxing each image independently, holding fixed the hydrogen bond length and the Co–C bond dihedral.

The resulting B3LYP energy profile is shown in Fig. 3(b), with stationary-point geometries illustrated [Fig. 3(a)]. These profiles are computed with a restricted open-shell reference for comparison with embedding calculations (below); as a result, the B3LYP profile does not completely reach the product minimum. For comparison, a second profile is shown corresponding to Perdew–Burke–Ernzerhof (PBE)[85] energy calculations performed at the same geometries. Interestingly, the barrier predicted by PBE is larger than that predicted by B3LYP, running contrary to the general trend that hybrid functionals predict larger barriers than generalized gradient approximation (GGA) functionals,[86] due to the latter over-stabilizing the relatively delocalized electronic structure of the transition state;[87] consistently, HF predicts a still lower barrier of 3.5 kcal/mol. In this system, the reactant and product feature electronic structures with a greater degree of delocalization than in that of the transition state. Thus, the self-interaction error present in GGA functionals may be over-stabilizing the reactant and product relative to the transition state, resulting in a barrier that is too high. The exact exchange present in B3LYP partially corrects this error, reducing the barrier.

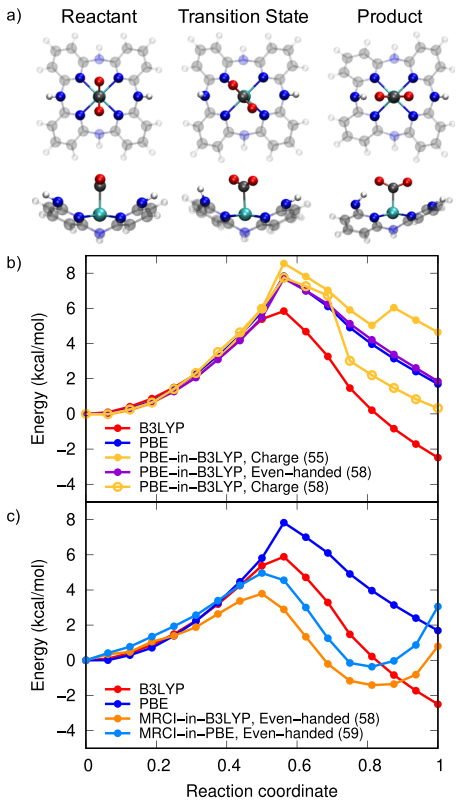

FIG. 3. Embedding calculations on the formation of a hydrogen bond between a cobalt aminopyridine complex and a bound CO_2 molecule. (a) Reactant, transition state, and product geometries illustrated with opaque subsystem A atoms and transparent subsystem B atoms. Two views are shown for each geometry. (b) B3LYP, PBE, and PBE-in-B3LYP energy profiles, with three methods of LMO selection for the last. (c) Energy profiles from MRCI embedded in B3LYP and in PBE using even-handed LMO selection. All curves are referenced to an energy of zero at the reactant geometry. The number of occupied LMOs in subsystem A is given in parentheses (out of 121 total).

With this reaction coordinate in hand, the even-handed procedure is used to select $\{\psi\}_A$ corresponding to the choice of $\{X\}_A$ shown as opaque atoms in Fig. 3(a). Embedding calculations are performed in the MOLPRO 2018.0 software package[88,89] with the def2-TZVP basis set, an exchange-correlation grid threshold of 10^{-10}, and with density fitting for both Coulomb and exchange integrals[90] evaluated with the def2-TZVPP/JKFIT basis.[91] A restricted open-shell reference is employed for all embedding calculations. The intrinsic bond orbital (IBO) procedure[50] is employed for orbital localization.

We first embed PBE in B3LYP to test the smoothness of the embedded reaction coordinate and to test the convergence of the embedded energy with respect to the size of subsystem A; a correlated wavefunction calculation on this large system would prove computationally infeasible and so PBE is used as a proxy. Figure 3(b) presents the results for PBE-in-B3LYP embedding for various LMO selection procedures. Selection of LMOs using the charge method (Sec. II B) results in subsystem A containing 55 occupied LMOs (of 121 total) and a discontinuous PBE-in-B3LYP energy profile (yellow circles). Even-handed LMO selection (Sec. II C) yields a continuous PBE-in-B3LYP energy profile in quantitative agreement with the whole-system PBE energy profile (purple circles). Even-handed selection adds only 3 LMOs to $\{\psi\}_A$ beyond those selected by the charge method, for a total of 58. As a further comparison, the charge method is performed with a modified threshold in Eq. (3) chosen such that every geometry has 58 selected LMOs in $\{\psi\}_A$. As for the charge method with the default threshold (0.4 electrons), this modified treatment of the charge threshold also results in a discontinuous energy profile (yellow open circles). Taken together, these results demonstrate that (i) even-handed LMO selection successfully improves upon the charge method to yield a continuous energy profile; (ii) the results of the even-handed LMO selection procedure cannot be replicated by simply performing the charge method with a different threshold; and (iii) the resulting embedded energy profile is in qualitative agreement with PBE calculations performed on the whole system, suggesting that the embedding calculation is converged with respect to the choice of $\{X\}_A$ and that the even-handed $\{\psi\}_A$ is an appropriate set of LMOs for embedding a correlated wavefunction method (see below).

We next embed correlated wavefunction methods using the even-handed set of LMOs determined above. To reduce the size of the virtual space, even-handed AO truncation is employed as discussed in Sec. II D with a charge threshold of 0.001 electrons, resulting in the retention of 713 out of 1102 AO basis functions in this compact complex. Comparison of MP2 with MP2-in-B3LYP calculations confirms that this AO truncation alters the reaction energy by less than 0.3 kcal/mol. Initial CCSD-in-B3LYP calculations performed at each geometry along the reaction coordinate yield T1 diagnostics[92] of at least 0.07; this value exceeds both the standard

threshold of 0.02 and the threshold of 0.05 proposed for first-row transition metals in Ref. 93, suggesting the presence of strong multireference character that CCSD cannot reliably describe.

To treat this potential multireference character, embedded multireference configuration interaction singles and doubles (MRCI)[94,95] in DFT calculations with a complete active space self-consistent field (CASSCF)[96,97] reference are performed. CASSCF calculations are performed without density fitting. The active space is chosen to comprise nine electrons in nine orbitals. Two factors suggest this active space is sufficiently large. First, the unrestricted natural orbital complete active space method of Bofill and Pulay[98] produces a guess active space that is at most three electrons in three orbitals at any geometry. Second, the CAS(9,9) canonical orbital populations are at least 1.98 for the lowest-energy active orbital and at most 0.03 for the highest-energy active orbital across all geometries. Multireference character is confirmed by substantial deviations from integer occupations within these active canonical orbitals. All reference configurations with the norm larger than 0.01 are included in the MRCI dynamical correlation calculation. This parameter is sufficient to converge the resulting energy profile and results in a maximum of 52 reference configurations. Relaxed reference Davidson inextensivity corrections are applied.[99]

Figure 3(c) shows the results of embedding MRCI in both B3LYP and PBE. (An even-handed $\{\psi\}_A$ was determined for the case of PBE in the same manner as for the case of B3LYP.) The resulting reaction profiles are continuous and smooth like those of the test PBE-in-B3LYP calculations. Under the MRCI-in-DFT embedding, the hydrogen bond is weaker and has a longer bond length (corresponding to a smaller optimal value of the reaction coordinate) compared to the underlying DFT calculations. The embedded profiles also show a further decrease in the barrier height. Comparing embedding in B3LYP versus in PBE, both result in the same qualitative shape of the reaction profile. Moving along the reaction coordinate, the MRCI-in-B3LYP and MRCI-in-PBE energy profiles diverge, resulting in a disagreement of 2 kcal/mol at the geometry corresponding to a reaction coordinate value of 1. This disagreement is smaller than the 4 kcal/mol disagreement between B3LYP and PBE at the same geometry, but is still significant. This residual difference is due to different descriptions by the two functionals of the geometric changes in subsystem *B*. Across this reaction coordinate, significant strain forms in the ligand backbone which lies largely in subsystem *B*. Thus, we see that embedding largely removes the dependence of the description of subsystem *A* on the density functional employed for subsystem *B*.

In this example, charge selection results in a qualitatively incorrect discontinuous energy profile, while even-handed selection results in a quantitatively accurate and continuous profile. This improvement comes at the cost of merely 3 additional LMOs included in $\{\psi\}_A$.

B. Binding of CO to Cu clusters

As a second demonstration of even-handed selection, we consider the binding of carbon monoxide to copper clusters. This class of systems has proven difficult for DFT, with different functionals predicting a wide range of binding energies as well as different preferred binding sites. For the example of CO binding to a Cu(111) surface, PBE, which is often accurate for metals, incorrectly predicts a hollow site preference for CO binding.[100] Meanwhile B3LYP, which is often accurate for molecules, predicts the correct on-top binding site preference, but with a binding energy of −2.2 kcal/mol,[101] compared to experimental measurements which range between −10.4 and −12.0 kcal/mol.[102–106] Previous work employing frozen-density embedding with approximate non-additive kinetic energy potentials has shown that embedded MRCI-in-LDA calculations can reproduce the measured CO–Cu(111) binding-site preference and binding energy.[100]

We first consider the binding of CO to a Cu_{38} cluster, constructed as a hemispheric cut of a Cu(111) surface centered around a top site, corresponding to two surface neighbor rings around this site. A reaction path for the binding of CO is constructed by varying the C–Cu bond length in steps of 0.3 Å and relaxing the C–O bond length at the PBE/def2-TZVP level of theory, holding the Cu atoms fixed in their bulk lattice positions. The Cu atoms are treated with the 10 electron Stuttgart/Cologne effective core potential ECP10MDF.[107] As in Sec. III A, calculations are performed in the MOLPRO 2018.0 software package with an exchange-correlation grid threshold of 10^{-10} and with density fitting for Coulomb and exchange integrals evaluated with the def2-TZVPP/JKFIT basis. The binding energy is found to be −17.3 kcal/mol with a C–Cu bond length of 1.90 Å. At the same geometry, the binding energy from B3LYP is −9.4 kcal/mol.

Figure 4 examines the convergence of the B3LYP-in-PBE binding energy as a function of subsystem *A*. Four choices of subsystem *A* are considered, corresponding to embedding CO and 1, 4, 7, and 10 copper atoms [Fig. 4(c)]. Even-handed AO truncation is employed with a charge threshold of 0.001 electrons and is found to introduce negligible errors. Figure 4(a) shows the convergence of the embedded binding energy using charge (yellow) versus even-handed LMO selection (purple). Charge selection results in an inconsistent set of LMOs for subsystem *A*, $\{\psi\}_A$, between the bound and unbound CO geometries, leading to large errors even for the larger sizes of subsystem *A* considered. By contrast, even-handed selection results in rapid convergence to the full B3LYP result. Even for the smallest choice of subsystem *A* tested ($n_{Cu} = 1$), even-handed selection reproduces the full B3LYP result within 1.5 kcal/mol, despite treating only three atoms (8 occupied LMOs) at the B3LYP level. Notably, the charge and even-handed methods select the same number of LMOs for each choice of $\{X\}_A$. Equation (3) selects the most LMOs at bound geometries, and LMOs must be added to the $\{\psi\}_A$ corresponding to unbound geometries in order to keep the number of occupied LMOs constant throughout the binding coordinate. Due to different polarization of the Cu atoms in the bound versus unbound geometry, the charge method selects qualitatively different LMOs than in the $\{\psi\}_A$ corresponding to the bound geometry. Figure 4(b) shows CO–Cu_{38} binding curves computed with PBE, B3LYP, and B3LYP-in-PBE embedding. These B3LYP-in-PBE energy profiles are smooth and converge rapidly to the reference B3LYP

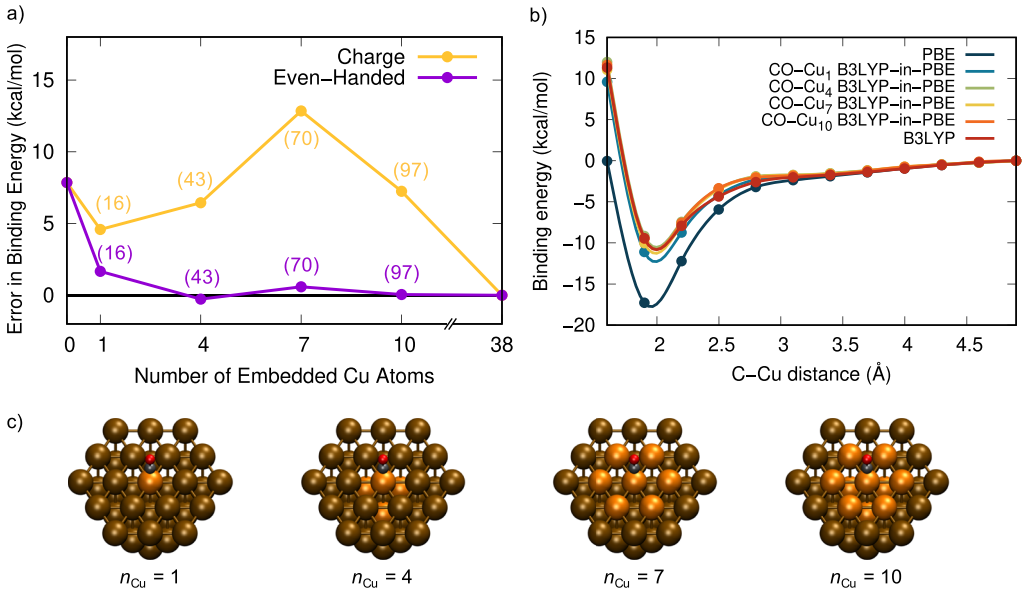

FIG. 4. Embedded calculations on CO–Cu_{38} top site binding. (a) Convergence of the binding energy error with respect to subsystem A's size for B3LYP-in-PBE embedding using charge and even-handed LMO selection. $n_{Cu} = 0$ corresponds to a PBE calculation on the whole system and $n_{Cu} = 38$ corresponds to a B3LYP calculation on the whole system. The number of occupied LMOs in subsystem A is given in parentheses (out of 368 total). As discussed is Sec. II B, the set of LMOs selected by the charge method with a given threshold is augmented so that a consistent number of LMOs is selected across all geometries in the reaction coordinate. (b) CO–Cu_{38} binding potential energy surfaces calculated using B3LYP-in-PBE embedding with even-handed LMO selection. (c) CO–Cu_{38} geometry. Subsystem A contains CO as well as the Cu atoms indicated in orange.

result. Furthermore, these results include ligand-induced relaxation of the electronic density on the metal cluster that was neglected in previous frozen-density embedding calculations.[100] Taken together, Figs. 4(a) and 4(b) demonstrate that even-handed LMO selection enables rapid convergence with respect to the size of subsystem A and yields smooth energy profiles.

B3LYP-in-PBE embedding provides some computational savings compared to a B3LYP calculation on the whole system. Averaged over the twelve geometries considered, B3LYP-in-PBE embedding results in speedups[108] versus B3LYP ranging from 5.8× for $n_{Cu} = 1$ to 3.4× for $n_{Cu} = 10$. (PBE is on average 7.2× faster than B3LYP for this system in the implementation used here.) This speedup results both from a reduction in the number of AOs for which exact exchange is evaluated and from a reduction of the number of SCF cycles necessary to converge the embedded B3LYP calculation.[109]

The metallic nature of this system (PBE HOMO-LUMO gap of 0.2 eV) results in poor orbital localization, particularly for those MOs near the Fermi level with primarily 4s character.[110] Consequently, AO truncation, which depends on the spatial locality of $\boldsymbol{\gamma}^A$, results in a large fraction of retained AO basis functions. For $n_{Cu} = 1$, the smallest subsystem A considered, 789 AOs are retained (out of 1772 total); 1413 AOs are retained for the case of $n_{Cu} = 10$. As a result, wavefunction-in-DFT embedding with this approach is computationally infeasible without further method development regarding virtual space reduction or occupied MO localization; for the purposes of the current work, we instead focus on a smaller cluster with a larger bandgap.

A Cu_{10} cluster is constructed similarly to the Cu_{38} cluster by a hemispheric cut of a Cu(111) surface centered around a top site, corresponding to one nearest-neighbor shell around this site. The bound CO–Cu_{10} geometry is optimized at the PBE/def2-SVP level of theory with the Cu atoms held fixed in their bulk lattice positions. The smaller size of this cluster results in an increase of the PBE HOMO-LUMO gap to 0.7 eV.

CCSD-in-PBE and CCSD-in-HF embedding calculations are performed in a def2-SVP basis. Three choices of subsystem A are constructed in the same manner as in the Cu_{38} cluster (see Fig. 5). $\{\psi\}_A$ is chosen with even-handed LMO selection. (Although for these three choices of $\{X\}_A$, the charge

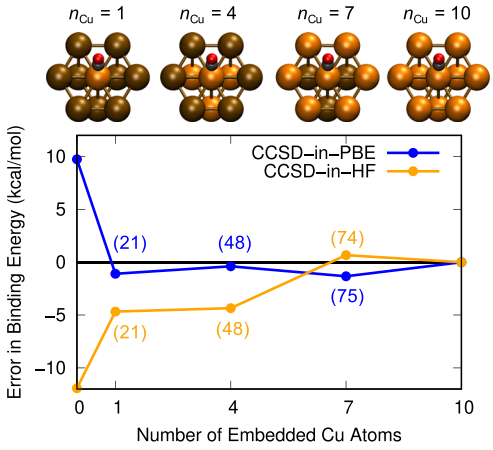

FIG. 5. Error in the embedded CCSD CO–Cu_{10} binding energy with respect to subsystem A size for the case of embedding in PBE versus embedding in HF. Above: CO–Cu_{10} geometry with subsystem A Cu atoms shown in orange for the three sizes of subsystem A considered. (CO is always included in subsystem A.) The number of occupied LMOs in subsystem A is given in parentheses (out of 102 total). $n_{Cu} = 0$ corresponds to a MF calculation on the whole system, while $n_{Cu} = 10$ corresponds to a CCSD calculation on the whole system.

method selects the same LMOs.) Figure 5 presents the binding energy error associated with CCSD-in-PBE versus CCSD-in-HF embedding as a function of the choice of subsystem *A*. Errors are computed relative to reference full-system CCSD calculation, available due to the smaller size of this cluster. The T1 diagnostic is found to be less that 0.02 at all geometries. CCSD-in-PBE embedding converges rapidly to the full CCSD result, with an error of 1 kcal/mol at the smallest subsystem *A* size considered. By contrast, CCSD-in-HF embedding does not reach the same level of accuracy until subsystem *A* contains all atoms in the system except the three subsurface Cu atoms. As in Fig. 3, projection-based embedding removes dependence on the MF method used for subsystem *B*; however, a larger subsystem *A* is required when the MF poorly describes the electronic structure of subsystem *B* (as in HF for this Cu cluster).

Compared to the charge selection method, which is fragile to changes in polarization upon CO binding, we see that even-handed LMO selection gives a consistent $\{\psi\}_A$ with respect to all geometries along the binding coordinate (Fig. 4). Using even-handed LMO selection leads to rapid convergence of the energy calculation with respect to the size of subsystem A.

IV. CONCLUSION

Projection-based embedding provides a framework for rigorous wavefunction-in-DFT embedding. In this paper, we demonstrate that the standard charge-based criterion for selecting the integer number of occupied LMOs that comprise the embedded subsystem can lead to discontinuous energy profiles and slow convergence of the calculated energy with respect to the size of subsystem *A*. We have introduced an even-handed selection procedure that ensures a consistent set of embedded occupied LMOs throughout a reaction coordinate. This algorithm only uses mean-field quantities that have already been computed as part of the projection-based embedding wavefunction-in-DFT procedure; thus, it adds negligible cost.

This method has been applied in several situations including for the cases of an organometallic catalyst and the binding of a molecule to a metal cluster, considering both DFT-in-DFT embedding and wavefunction-in-DFT embedding. In all cases, the even-handed method has been shown to be superior to the original charge method for selecting LMOs, resulting in smooth potential embedded energy surfaces and more rapid convergence of the embedded energy with respect to the size of subsystem *A*. Furthermore, embedding calculations performed using even-handed selection are largely insensitive to the underlying mean-field theory used to treat subsystem *B*, although it is worth noting that a sufficiently poor choice of mean-field theory, such as HF for a metal cluster, can lead to slower convergence with respect to the size of subsystem *A* than when starting from a better mean-field theory (Fig. 5).

Projection-based embedding has recently been extended to include periodic boundary conditions,[109,111] and the ideas of even-handed selection can be naturally applied in such a framework; however, caution should be taken in the case of wavefunction-in-DFT embedding for metals due to slow convergence of the AO truncation in small-bandgap systems for which poorly localized occupied LMOs contribute substantially to the correlation energy. Regardless, we show that even-handed selection results in highly accurate B3LYP-in-PBE embedding for Cu binding to a Cu_{38} cluster despite substantial changes in polarization of the metal (Fig. 4), and accurate CCSD-in-PBE embedding was demonstrated for CO binding to the smaller Cu_{10} cluster (Fig. 5).

The even-handed selection method introduced here provides a consistent set of occupied LMOs for an embedding calculation across a reaction coordinate. Because its input comes entirely from mean-field calculations at each reaction coordinate geometry, the method has negligible cost compared to the original charge method for LMO selection. In cases where even-handed and charge selection yield the same embedded occupied LMOs, the even-handed procedure provides a measure of confidence that the embedded density is consistent across geometries. When the two procedures differ, even-handed selection has in all cases resulted in qualitatively superior results at the cost of a modest number of additional embedded occupied LMOs. We therefore recommend the use of the even-handed LMO selection procedure whenever considering a specified reaction pathway.

ACKNOWLEDGMENTS

We thank Feizhi Ding and Sebastian Lee for helpful discussions. M.W. thanks the Resnick Sustainability Institute for a postdoctoral fellowship. T.F.M. acknowledges support in part from the NSF under Award No. CHE-1611581; additionally, this material is based upon work performed by the Joint Center for Artificial Photosynthesis, a DOE Energy Innovation Hub, supported through the Office of Science of the U.S. Department of Energy under Award No. DE-SC0004993. F.R.M. is grateful for funding from EPSRC (Grant No. EP/M013111/1).

[1] M. Born, Z. Phys. **1**, 45 (1920).
[2] S. Miertuš, E. Scrocco, and J. Tomasi, Chem. Phys. **55**, 117 (1981).
[3] C. J. Cramer and D. G. Truhlar, Chem. Rev. **99**, 2161 (1999).
[4] A. Warshel and M. Levitt, J. Mol. Biol. **103**, 227 (1976).
[5] M. J. Field, P. A. Bash, and M. Karplus, J. Comput. Chem. **11**, 700 (1990).
[6] H. Lin and D. G. Truhlar, Theor. Chem. Acc. **117**, 185 (2006).
[7] T. Vreven and K. Morokuma, *Annual Reports in Computational Chemistry* (Elsevier, Amsterdam, Netherlands, 2006), pp. 35–51.
[8] P. Cortona, Phys. Rev. B **44**, 8454 (1991).
[9] T. A. Wesolowski and A. Warshel, J. Phys. Chem. **97**, 8050 (1993).
[10] T. Wesolowski and A. Warshel, J. Phys. Chem. **98**, 5183 (1994).
[11] M. Iannuzzi, B. Kirchner, and J. Hutter, Chem. Phys. Lett. **421**, 16 (2006).
[12] O. Roncero, M. P. de Lara-Castells, P. Villarreal, F. Flores, J. Ortega, M. Paniagua, and A. Aguado, J. Chem. Phys. **129**, 184104 (2008).
[13] P. Elliott, K. Burke, M. H. Cohen, and A. Wasserman, Phys. Rev. A **82**, 024501 (2010).
[14] S. Fux, C. R. Jacob, J. Neugebauer, L. Visscher, and M. Reiher, J. Chem. Phys. **132**, 164101 (2010).
[15] C. Huang, M. Pavone, and E. A. Carter, J. Chem. Phys. **134**, 154110 (2011).
[16] J. Nafziger, Q. Wu, and A. Wasserman, J. Chem. Phys. **135**, 234101 (2011).
[17] J. D. Goodpaster, N. Ananth, F. R. Manby, and T. F. Miller, J. Chem. Phys. **133**, 084103 (2010).
[18] J. D. Goodpaster, T. A. Barnes, and T. F. Miller, J. Chem. Phys. **134**, 164108 (2011).
[19] J. D. Goodpaster, T. A. Barnes, F. R. Manby, and T. F. Miller, J. Chem. Phys. **137**, 224113 (2012).

[20]F. R. Manby, M. Stella, J. D. Goodpaster, and T. F. Miller III, J. Chem. Theory Comput. **8**, 2564 (2012).
[21]A. S. P. Gomes and C. R. Jacob, Annu. Rep. Prog. Chem., Sect. C: Phys. Chem. **108**, 222 (2012).
[22]J. D. Goodpaster, T. A. Barnes, F. R. Manby, and T. F. Miller III, J. Chem. Phys. **140**, 18A507 (2014).
[23]F. Libisch, C. Huang, and E. A. Carter, Acc. Chem. Res. **47**, 2768 (2014).
[24]T. A. Wesolowski, S. Shedge, and X. Zhou, Chem. Rev. **115**, 5891 (2015).
[25]B. O. Roos, P. R. Taylor, and P. E. Siegbahn, Chem. Phys. **48**, 157 (1980).
[26]P. Å. Malmqvist, A. Rendell, and B. O. Roos, J. Phys. Chem. **94**, 5477 (1990).
[27]P. Pulay, Chem. Phys. Lett. **100**, 151 (1983).
[28]H.-J. Werner and K. Pflüger, *Annual Reports in Computational Chemistry* (Elsevier, Amsterdam, Netherlands, 2006), Vol. 2, pp. 53–80.
[29]J. E. Inglesfield, J. Phys. C: Solid State Phys. **14**, 3795 (1981).
[30]D. Zgid and G. K.-L. Chan, J. Chem. Phys. **134**, 094115 (2011).
[31]J. J. Phillips and D. Zgid, J. Chem. Phys. **140**, 241101 (2014).
[32]G. Knizia and G. K.-L. Chan, Phys. Rev. Lett. **109**, 186404 (2012).
[33]G. Knizia and G. K.-L. Chan, J. Chem. Theory Comput. **9**, 1428 (2013).
[34]K. R. Glaesemann, M. S. Gordon, and H. Nakano, Phys. Chem. Chem. Phys. **1**, 967 (1999).
[35]E. R. Sayfutyarova, Q. Sun, G. K.-L. Chan, and G. Knizia, J. Chem. Theory Comput. **13**, 4063 (2017).
[36]C. J. Stein and M. Reiher, J. Chem. Theory Comput. **12**, 1760 (2016).
[37]C. Stein and M. Reiher, Chimia Int. J. Chem. **71**, 170 (2017).
[38]A. K. Dutta, M. Nooijen, F. Neese, and R. Izsák, J. Chem. Phys. **146**, 074103 (2017).
[39]J. J. Bao, S. S. Dong, L. Gagliardi, and D. G. Truhlar, J. Chem. Theory Comput. **14**, 2017 (2018).
[40]R. A. Mata and H.-J. Werner, J. Chem. Phys. **125**, 184110 (2006).
[41]P. G. Lykos and R. G. Parr, J. Chem. Phys. **25**, 1301 (1956).
[42]R. A. Mata, H.-J. Werner, and M. Schütz, J. Chem. Phys. **128**, 144106 (2008).
[43]T. M. Henderson, J. Chem. Phys. **125**, 014105 (2006).
[44]Y. G. Khait and M. R. Hoffmann, *Annual Reports in Computational Chemistry* (Elsevier, Amsterdam, Netherlands, 2012), Vol. 8, pp. 53–70.
[45]B. Hégely, P. R. Nagy, G. G. Ferenczy, and M. Kállay, J. Chem. Phys. **145**, 064107 (2016).
[46]S. Sæther, T. Kjærgaard, H. Koch, and I.-M. Høyvik, J. Chem. Theory Comput. **13**, 5282 (2017).
[47]T. Culpitt, K. R. Brorsen, and S. Hammes-Schiffer, J. Chem. Phys. **146**, 211101 (2017).
[48]S. J. Bennie, M. Stella, T. F. Miller III, and F. R. Manby, J. Chem. Phys. **143**, 024105 (2015).
[49]I. Mayer, *Simple Theorems, Proofs, and Derivations in Quantum Chemistry* (Springer, New York, 2003).
[50]G. Knizia, J. Chem. Theory Comput. **9**, 4834 (2013).
[51]R. J. Bartlett, J. D. Watts, S. A. Kucharski, and J. Noga, Chem. Phys. Lett. **165**, 513 (1990).
[52]S. H. Vosko, L. Wilk, and M. Nusair, Can. J. Phys. **58**, 1200 (1980).
[53]C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
[54]A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).
[55]P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, J. Phys. Chem. **98**, 11623 (1994).
[56]A. T. B. Gilbert, N. A. Besley, and P. M. W. Gill, J. Phys. Chem. A **112**, 13164 (2008).
[57]T. A. Barnes, J. D. Goodpaster, F. R. Manby, and T. F. Miller III, J. Chem. Phys. **139**, 024103 (2013).
[58]A. Chapovetsky, T. H. Do, R. Haiges, M. K. Takase, and S. C. Marinescu, J. Am. Chem. Soc. **138**, 5765 (2016).
[59]A. Chapovetsky, M. Welborn, J. M. Luna, R. Haiges, T. F. Miller III, and S. C. Marinescu, ACS Cent. Sci. **4**, 397 (2018).
[60]E. Fujita, C. Creutz, N. Sutin, and B. S. Brunschwig, Inorg. Chem. **32**, 2657 (1993).
[61]J. D. Froehlich and C. P. Kubiak, Inorg. Chem. **51**, 3932 (2012).
[62]C. Costentin, G. Passard, M. Robert, and J.-M. Savéant, J. Am. Chem. Soc. **136**, 11821 (2014).
[63]K. Leung, I. M. B. Nielsen, N. Sai, C. Medforth, and J. A. Shelnutt, J. Phys. Chem. A **114**, 10174 (2010).
[64]I. M. B. Nielsen and K. Leung, J. Phys. Chem. A **114**, 10166 (2010).
[65]C. Liu, T. R. Cundari, and A. K. Wilson, J. Phys. Chem. C **116**, 5681 (2012).
[66]A. A. Peterson and J. K. Nørskov, J. Phys. Chem. Lett. **3**, 251 (2012).
[67]D. Cheng, F. R. Negreiros, E. Aprà, and A. Fortunelli, ChemSusChem **6**, 944 (2013).
[68]V. Tripkovic, M. Vanin, M. Karamad, M. E. Björketun, K. W. Jacobsen, K. S. Thygesen, and J. Rossmeisl, J. Phys. Chem. C **117**, 9187 (2013).
[69]Y. J. Zhang, V. Sethuraman, R. Michalsky, and A. A. Peterson, ACS Catal. **4**, 3742 (2014).
[70]M. J. Cheng, Y. Kwon, M. Head-Gordon, and A. T. Bell, J. Phys. Chem. C **119**, 21345 (2015).
[71]A. J. Göttle and M. T. M. Koper, Chem. Sci. **8**, 458 (2017).
[72]J. Shen, M. J. Kolb, A. J. Göttle, and M. T. Koper, J. Phys. Chem. C **120**, 15714 (2016).
[73]C. J. Barden, J. C. Rienstra-Kiracofe, and H. F. Schaefer, J. Chem. Phys. **113**, 690 (2000).
[74]C. Diedrich, A. Luchow, and S. Grimme, J. Chem. Phys. **122**, 021101 (2005).
[75]K. P. Jensen, B. O. Roos, and U. Ryde, J. Chem. Phys. **126**, 014103 (2007).
[76]S. Goel and A. E. Masunov, J. Chem. Phys. **129**, 214302 (2008).
[77]C. J. Cramer and D. G. Truhlar, Phys. Chem. Chem. Phys. **11**, 10757 (2009).
[78]B. Santra, A. Michaelides, and M. Scheffler, J. Chem. Phys. **127**, 184104 (2007).
[79]G. A. DiLabio, E. R. Johnson, and A. Otero-de-la Roza, Phys. Chem. Chem. Phys. **15**, 12821 (2013).
[80]P. Huo, C. Uyeda, J. D. Goodpaster, J. C. Peters, and T. F. Miller III, ACS Catal. **6**, 6114 (2016).
[81]V. A. Rassolov, M. A. Ratner, J. A. Pople, P. C. Redfern, and L. A. Curtiss, J. Comput. Chem. **22**, 976 (2001).
[82]V. I. Lebedev, Sib. Math. J. **18**, 99 (1977).
[83]A. Behn, P. M. Zimmerman, A. T. Bell, and M. Head-Gordon, J. Chem. Phys. **135**, 224108 (2011).
[84]Y. Shao, Z. Gan, E. Epifanovsky, A. T. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng, D. Ghosh, M. Goldey, P. R. Horn, L. D. Jacobson, I. Kaliman, R. Z. Khaliullin, T. Kuś, A. Landau, J. Liu, E. I. Proynov, Y. M. Rhee, R. M. Richard, M. A. Rohrdanz, R. P. Steele, E. J. Sundstrom, H. L. Woodcock, P. M. Zimmerman, D. Zuev, B. Albrecht, E. Alguire, B. Austin, G. J. O. Beran, Y. A. Bernard, E. Berquist, K. Brandhorst, K. B. Bravaya, S. T. Brown, D. Casanova, C.-M. Chang, Y. Chen, S. H. Chien, K. D. Closser, D. L. Crittenden, M. Diedenhofen, R. A. DiStasio, H. Do, A. D. Dutoi, R. G. Edgar, S. Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-Zadorozhnaya, J. Gomes, M. W. Hanson-Heine, P. H. Harbach, A. W. Hauser, E. G. Hohenstein, Z. C. Holden, T.-C. Jagau, H. Ji, B. Kaduk, K. Khistyaev, J. Kim, J. Kim, R. A. King, P. Klunzinger, D. Kosenkov, T. Kowalczyk, C. M. Krauter, K. U. Lao, A. D. Laurent, K. V. Lawler, S. V. Levchenko, C. Y. Lin, F. Liu, E. Livshits, R. C. Lochan, A. Luenser, P. Manohar, S. F. Manzer, S.-P. Mao, N. Mardirossian, A. V. Marenich, S. A. Maurer, N. J. Mayhall, E. Neuscamman, C. M. Oana, R. Olivares-Amaya, D. P. O'Neill, J. A. Parkhill, T. M. Perrine, R. Peverati, A. Prociuk, D. R. Rehn, E. Rosta, N. J. Russ, S. M. Sharada, S. Sharma, D. W. Small, A. Sodt, T. Stein, D. Stück, Y.-C. Su, A. J. Thom, T. Tsuchimochi, V. Vanovschi, L. Vogt, O. Vydrov, T. Wang, M. A. Watson, J. Wenzel, A. White, C. F. Williams, J. Yang, S. Yeganeh, S. R. Yost, Z.-Q. You, I. Y. Zhang, X. Zhang, Y. Zhao, B. R. Brooks, G. K. Chan, D. M. Chipman, C. J. Cramer, W. A. Goddard, M. S. Gordon, W. J. Hehre, A. Klamt, H. F. Schaefer, M. W. Schmidt, C. D. Sherrill, D. G. Truhlar, A. Warshel, X. Xu, A. Aspuru-Guzik, R. Baer, A. T. Bell, N. A. Besley, J.-D. Chai, A. Dreuw, B. D. Dunietz, T. R. Furlani, S. R. Gwaltney, C.-P. Hsu, Y. Jung, J. Kong, D. S. Lambrecht, W. Liang, C. Ochsenfeld, V. A. Rassolov, L. V. Slipchenko, J. E. Subotnik, T. Van Voorhis, J. M. Herbert, A. I. Krylov, P. M. Gill, and M. Head-Gordon, Mol. Phys. **113**, 184 (2015).
[85]J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[86]A. Mahler, B. G. Janesko, S. Moncho, and E. N. Brothers, J. Chem. Phys. **148**, 244106 (2018).
[87]S. Patchkovskii and T. Ziegler, J. Chem. Phys. **116**, 7806 (2002).
[88]H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, and M. Schütz, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **2**, 242 (2012).
[89]H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, W. Györffy, D. Kats, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, S. Bennie, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Köppl, S. J. R. Lee, Y. Liu, A. W. Lloyd, Q. Ma, R. A. Mata, A. J. May,

S. J. McNicholas, W. Meyer, T. F. Miller III, M. E. Mura, A. Nicklass, D. P. O'Neill, P. Palmieri, D. Peng, K. Pflüger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, M. Wang, and M. Welborn, MOLPRO, version 2018.1, a package of *ab initio* programs, 2018, see http://www.molpro.net.
[90]R. Polly, H.-J. Werner, F. R. Manby, and P. J. Knowles, Mol. Phys. **102**, 2311 (2004).
[91]F. Weigend, J. Comput. Chem. **29**, 167 (2007).
[92]T. J. Lee, Chem. Phys. Lett. **372**, 362 (2003).
[93]W. Jiang, N. J. DeYonker, and A. K. Wilson, J. Chem. Theory Comput. **8**, 460 (2012).
[94]P. J. Knowles and H.-J. Werner, Chem. Phys. Lett. **145**, 514 (1988).
[95]H.-J. Werner and P. J. Knowles, J. Chem. Phys. **89**, 5803 (1988).
[96]P. J. Knowles and H.-J. Werner, Chem. Phys. Lett. **115**, 259 (1985).
[97]H.-J. Werner and P. J. Knowles, J. Chem. Phys. **82**, 5053 (1985).
[98]J. M. Bofill and P. Pulay, J. Chem. Phys. **90**, 3637 (1989).
[99]E. R. Davidson and D. W. Silver, Chem. Phys. Lett. **52**, 403 (1977).
[100]S. Sharifzadeh, P. Huang, and E. Carter, J. Phys. Chem. C **112**, 4649 (2008).
[101]M. Neef and K. Doll, Surf. Sci. **600**, 1085 (2006).
[102]J. Kessler and F. Thieme, Surf. Sci. **67**, 405 (1977).
[103]W. Kirstein, B. Krüger, and F. Thieme, Surf. Sci. **176**, 505 (1986).
[104]L. Bartels, G. Meyer, and K.-H. Rieder, Surf. Sci. **432**, L621 (1999).
[105]S. Vollmer, G. Witte, and C. Wöll, Catal. Lett. **77**, 97 (2001).
[106]A. Patra, J. Sun, and J. P. Perdew, e-print arXiv:1807.05450.
[107]D. Figgen, G. Rauhut, M. Dolg, and H. Stoll, Chem. Phys. **311**, 227 (2005).
[108]B3LYP reference timing calculations are started from both the Superposition of Atomic Densities (SAD) guess [J. H. Van Lenthe *et al.*, J. Comput. Chem. **27**, 926 (2006)] and from the converged PBE orbitals. The faster of these two timings is used for speedup calculations. PBE and B3LYP-in-PBE timings are computed starting from the SAD guess.
[109]F. Libisch, M. Marsman, J. Burgdorfer, and G. Kresse, J. Chem. Phys. **147**, 034110 (2017).
[110]The Pipek-Mezey [J. Pipek and P. G. Mezey, J. Chem. Phys. **90**, 4916 (1989)] and Boys [S. F. Boys, Rev. Mod. Phys. **32**, 296 (1960)] localization schemes result in a greater number of poorly localized LMOs, as does changing the IBO localization exponent from 4 to 2.
[111]D. V. Chulhai and J. D. Goodpaster, J. Chem. Theory Comput. **14**, 1928 (2018).