AIP | The Journal of Chemical Physics

Accelerating wavefunction in density-functional-theory embedding by truncating the active basis set

Simon J. Bennie, Martina Stella, Thomas F. Miller III, and Frederick R. Manby

Citation: The Journal of Chemical Physics **143**, 024105 (2015); doi: 10.1063/1.4923367
View online: http://dx.doi.org/10.1063/1.4923367
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/143/2?ver=pdfcov
Published by the AIP Publishing

Accelerating wavefunction in density-functional-theory embedding by truncating the active basis set

Simon J. Bennie,[1,a)] Martina Stella,[1] Thomas F. Miller III,[2] and Frederick R. Manby[1,b)]
[1]*Centre for Computational Chemistry, School of Chemistry, University of Bristol, Bristol BS8 1TS, United Kingdom*
[2]*Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, USA*

(Received 2 April 2015; accepted 22 June 2015; published online 10 July 2015)

Methods where an accurate wavefunction is embedded in a density-functional description of the surrounding environment have recently been simplified through the use of a projection operator to ensure orthogonality of orbital subspaces. Projector embedding already offers significant performance gains over conventional post-Hartree–Fock methods by reducing the number of correlated occupied orbitals. However, in our first applications of the method, we used the atomic-orbital basis for the full system, even for the correlated wavefunction calculation in a small, active subsystem. Here, we further develop our method for truncating the atomic-orbital basis to include only functions within or close to the active subsystem. The number of atomic orbitals in a calculation on a fixed active subsystem becomes asymptotically independent of the size of the environment, producing the required $O(N^0)$ scaling of cost of the calculation in the active subsystem, and accuracy is controlled by a single parameter. The applicability of this approach is demonstrated for the embedded many-body expansion of binding energies of water hexamers and calculation of reaction barriers of S_N2 substitution of fluorine by chlorine in α-fluoroalkanes. © *2015 AIP Publishing LLC.* [http://dx.doi.org/10.1063/1.4923367]

I. INTRODUCTION

Density functional theory (DFT)[1,2] can be used as a flexible framework to build multiscale electronic structure methods in which an accurate wavefunction (WF) calculation on a chemically important subsystem is embedded in a lower-level model of the environment.[3–9] The electronic density of a system can be partitioned into active (A) and environment (B) subsystems

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

and the energy

$$E[\rho] = E[\rho^{\mathrm{A}}] + E[\rho^{\mathrm{B}}] + \delta E[\rho^{\mathrm{A}}, \rho^{\mathrm{B}}] \tag{2}$$

partitions along similar lines, with the last term accounting for all sources of nonadditivity. Amongst these, the Coulomb term is simple, and exchange-correlation is readily handled through existing approximate functionals or other low-cost methods.[10]

The calculation of the remaining contribution to $\delta E[\rho^{\mathrm{A}}, \rho^{\mathrm{B}}]$, the nonadditive kinetic energy

$$\delta T_s[\rho^{\mathrm{A}}, \rho^{\mathrm{B}}] = T_s[\rho] - T_s[\rho^{\mathrm{A}}] - T_s[\rho^{\mathrm{B}}], \tag{3}$$

constitutes one of the central challenges. This component of the energy accounts for exchange-repulsion effects between subsystems. It has been treated approximately using a range of kinetic energy density functionals[6,8,11,12] or exactly through Kohn-Sham (KS) potential reconstruction schemes.[13–18]

Neither approach is completely satisfactory: despite much progress,[19–22] currently available approximate kinetic energy functionals lack the accuracy required to partition across covalent bonds,[16,23,24] and potential reconstruction schemes, like the optimized effective potential,[25,26] are fraught with numerical difficulties when applied to problems in finite atomic-orbital basis sets.[27,28] It was recently shown, though, that the complication of nonadditive kinetic energy terms can be completely bypassed by use of a projection operator to constrain subsystem densities to be formed from orthogonal subsets of orbitals.[29] This leads to a nonadditive kinetic energy of exactly zero, so further approximations can be avoided.

The approach taken in our group is to first perform a KS calculation on the entire system; then unitarily rotate the occupied KS orbitals, for example, to localise them; and then to pick a partition of orbitals, implying a partition of the electronic density. The key point is that then the nonadditive kinetic energy vanishes, and the energy can simply be expressed in terms of the fragment density matrices $\boldsymbol{\gamma}^{\mathrm{A}}$ and $\boldsymbol{\gamma}^{\mathrm{B}}$,

$$E[\boldsymbol{\gamma}^{\mathrm{A}}, \boldsymbol{\gamma}^{\mathrm{B}}] = \mathrm{tr}(\boldsymbol{\gamma}^{\mathrm{A}} + \boldsymbol{\gamma}^{\mathrm{B}})\mathbf{h} + J[\boldsymbol{\gamma}^{\mathrm{A}} + \boldsymbol{\gamma}^{\mathrm{B}}] + E_{\mathrm{xc}}[\boldsymbol{\gamma}^{\mathrm{A}} + \boldsymbol{\gamma}^{\mathrm{B}}], \tag{4}$$

where $\mathbf{h}$ is the core Hamiltonian for the whole system.[29]

It would be perfectly reasonable to perform a wavefunction calculation on subsystem A by transforming the atomic-orbital basis set to the orthogonal complement of the occupied orbitals in subsystem B. But the lower part of the spectrum of the operator $(1 - \hat{P})\hat{H}(1 - \hat{P})$ coincides with that of $\hat{H} + \mu\hat{P}$ for sufficiently large, positive μ,[29] and we elected to use the latter expression so that we could interface our embed-

a)Electronic mail: simonbennie@gmail.com
b)Electronic mail: fred.manby@bristol.ac.uk

ding scheme to any wavefunction-based electronic structure method implemented in Molpro[30,31] without further modifications.

The Fock matrix in the atomic-orbital basis for a mean-field calculation on subsystem A, taking account of embedding in the environment B, can be written as

$$f^{\mathrm{A}}_{\alpha\beta} = \frac{\partial}{\partial \gamma^{\mathrm{A}}_{\alpha\beta}} E[\boldsymbol{\gamma}^{\mathrm{A}}, \boldsymbol{\gamma}^{\mathrm{B}}] + \mu P^{\mathrm{B}}_{\alpha\beta}, \tag{5}$$

where

$$P^{\mathrm{B}}_{\alpha\beta} = \langle \alpha | \Big\{ \sum_{i \in B} |\psi^{\mathrm{B}}_i\rangle\langle\psi^{\mathrm{B}}_i| \Big\} | \beta \rangle = [\mathbf{S}\boldsymbol{\gamma}^{\mathrm{B}}\mathbf{S}]_{\alpha\beta} \tag{6}$$

and where α and β label atomic orbitals, and i labels an occupied molecular orbital. From the Fock matrix in Eq. (5), it is possible to extract the embedded core Hamiltonian

$$\begin{aligned}\mathbf{h}^{\mathrm{A\ in\ B}} = \mathbf{h} &+ \mathbf{J}[\boldsymbol{\gamma}^{\mathrm{A}} + \boldsymbol{\gamma}^{\mathrm{B}}] - \mathbf{J}[\boldsymbol{\gamma}^{\mathrm{A}}] \\ &+ \mathbf{v}_{\mathrm{xc}}[\boldsymbol{\gamma}^{\mathrm{A}} + \boldsymbol{\gamma}^{\mathrm{B}}] - \mathbf{v}_{\mathrm{xc}}[\boldsymbol{\gamma}^{\mathrm{A}}] + \mu\mathbf{P}^{\mathrm{B}}.\end{aligned} \tag{7}$$

This modified matrix $\mathbf{h}^{\mathrm{A\ in\ B}}$ can easily be used with any method that reads in an arbitrary core Hamiltonian. The level-shift projector prevents the KS orbitals from the environment from being occupied during any subsequent calculations and accounts (at the DFT level) for the Pauli exclusion effect of the state in subsystem B on the wavefunction in A. The energy of subsystem B is added in the same way as the internuclear repulsion, to give the total energy expression

$$\begin{aligned}E[\Psi^{\mathrm{A}}; \boldsymbol{\gamma}^{\mathrm{B}}] = &\langle \Psi^{\mathrm{A}} | \hat{H}^{\mathrm{A\ in\ B}} | \Psi^{\mathrm{A}} \rangle \\ &- \mathrm{tr}\, \boldsymbol{\gamma}^{\mathrm{A}} (\mathbf{v}_{\mathrm{xc}}[\boldsymbol{\gamma}^{A} + \boldsymbol{\gamma}^{B}] - \mathbf{v}_{\mathrm{xc}}[\boldsymbol{\gamma}^{\mathrm{A}}]) \\ &+ E_{\mathrm{xc}}[\boldsymbol{\gamma}^{\mathrm{A}} + \boldsymbol{\gamma}^{\mathrm{B}}] - E_{\mathrm{xc}}[\boldsymbol{\gamma}^{\mathrm{A}}] + E[0; \boldsymbol{\gamma}^{\mathrm{B}}].\end{aligned} \tag{8}$$

The first term is the conventional wavefunction energy expectation value using a Hamiltonian containing the embedding terms in $\mathbf{h}^{\mathrm{A\ in\ B}}$; the subsequent terms provide the standard correction for the exchange-correlation energy and the energy of subsystem B.

II. THEORY

In most applications of WF-in-DFT embedding, the wavefunction calculation is performed in a localised region of space, but in our original formulation of projector-based embedding, the wavefunction calculation was performed using the atomic-orbital basis set of the entire system. In previous work, Barnes *et al.* introduced a method for truncating the active-subsystem basis set.[32] This approach obtained $O(N^0)$ scaling but involved the specification of both a truncation threshold parameter and a geometry-dependent set of atoms that border the active subsystem. Here, we achieve the same aim with only a single truncation threshold parameter, avoiding the need to specify any geometry-dependent information.

When our embedding scheme is used with localised orbitals, subsystem A is specified by a list of atoms (and the list of active local orbitals is generated from this, via Mulliken population of the localised orbitals). In the basis truncation scheme, we will always include atomic orbitals centred on these atoms, so the truncation scheme that we will construct will smoothly interpolate between the full atomic-orbital basis set and the subset of functions centred on subsystem A atoms.

We determine the functions that make an important contribution to the density matrix of the active subsystem through their net Mulliken populations. The more commonly used gross population,

$$q_\alpha = \gamma^{\mathrm{A}}_{\alpha\alpha} S_{\alpha\alpha} + \sum_{\alpha \neq \beta} \gamma^{\mathrm{A}}_{\alpha\beta} S_{\beta\alpha}, \tag{9}$$

is composed of two parts. The first term is the net population,[33] while the second results from the nonorthogonality of the basis and determines the way in which the contributions from off-diagonal elements of the density matrix are shared between atomic orbitals.

A disadvantage in using (gross) Mulliken populations to determine the importance of the atomic orbital α is that the charges are not always positive. We also investigated a scheme using Löwdin populations but found it retained fewer energetically important subsystem B functions, whilst retaining functions that did not contribute strongly to the energy.

Of the methods we tried, we found net populations ($q^{\mathrm{net}}_\alpha = \gamma^{\mathrm{A}}_{\alpha\alpha} S_{\alpha\alpha}$) to offer the most reliable truncation. A given function α is discarded if $q^{\mathrm{net}}_\alpha < \lambda$. The retained atomic orbitals are then grouped so that complete shells of basis functions are kept together (e.g., if a particular p_x function is retained, then so too are the other two components, and if one component of a generalized contraction is retained, so too are the other components). With this definition of the basis-set truncation, we find each particular value of λ to provide a similar level of accuracy across a range of chemical problems, enabling us to recommend a default value of around $\lambda = 10^{-4}$ for typical applications.

An important consideration for a practical implementation of embedding is the generation of an accurate orbital guess for the post-embedding SCF. In the full basis, this is straightforward, because the natural orbitals of $\boldsymbol{\gamma}^{\mathrm{A}}$ constitute an exact initial guess (to within numerical considerations). The simplest estimate of the subsystem density matrix in the truncated basis $\tilde{\boldsymbol{\gamma}}^{\mathrm{A}}$ is simply the corresponding submatrix of $\boldsymbol{\gamma}^{\mathrm{A}}$. However, the molecular orbitals implied by this truncation are no longer perfectly orthogonal to the subsystem B orbitals, so the truncated projector contribution to the Fock matrix can be very large. Furthermore, the idempotency of the full subsystem density matrix $\boldsymbol{\gamma}^{\mathrm{A}}$ is not preserved in $\tilde{\boldsymbol{\gamma}}^{\mathrm{A}}$, or in other words, the orthonormality of the localised molecular orbitals is broken when atomic orbitals with small coefficients are eliminated.

To generate an accurate starting guess for the embedded SCF calculation, the truncated subsystem density matrix is subjected to canonical purification.[34] We then construct an effective Hamiltonian using this purified density matrix $\tilde{\boldsymbol{\gamma}}^{\mathrm{A}}$ together with the truncated-basis environment density $\tilde{\boldsymbol{\gamma}}^{\mathrm{B}}$. The Hamiltonian contains the core one-electron terms and the level shift projector, as in Equation (7), together with total Coulomb and exchange-correlation contributions in the truncated basis:

$$\tilde{\mathbf{h}}^{\mathrm{total}} = \tilde{\mathbf{h}} + \tilde{\mathbf{J}}[\tilde{\boldsymbol{\gamma}}^{\mathrm{A}} + \tilde{\boldsymbol{\gamma}}^{\mathrm{B}}] + \tilde{\mathbf{v}}_{\mathrm{xc}}[\tilde{\boldsymbol{\gamma}}^{\mathrm{A}} + \tilde{\boldsymbol{\gamma}}^{\mathrm{B}}] + \mu\tilde{\mathbf{P}}^{\mathrm{B}}. \tag{10}$$

Upon diagonalising this Hamiltonian, we obtain a set of starting orbitals that are influenced by the retained portion of the level shift projector. At the full supermolecular basis limit,

the orbitals that are generated with this method are the same (to within a unitary rotation) as those obtained by diagonalising the active density. The number of post-embedding SCF cycles required to converge is reduced as more functions are retained, and we find that this method avoids convergence artifacts that can arise due to the change in basis.

In summary, a typical calculation in which coupled-cluster theory with singles, doubles and perturbative triples (CCSD(T)) is embedded in DFT proceeds as follows:

1. Run a KS calculation on the entire system.
2. Form localised molecular orbitals and select a subset to define subsystem A.
3. Use truncation parameter λ to determine an atomic-orbital subset.
4. Perform Hartree-Fock (HF) and then CCSD(T) calculations in subsystem A.

In practice, we also correct the results by adding the difference between truncated DFT-in-DFT and the full basis DFT energy, so that the total energies presented below are computed as

$$E^{\text{WF-in-DFT}} = E^{\text{WF-in-DFT}}_{\text{truncated}} + E^{\text{DFT}}_{\text{full}} - E^{\text{DFT-in-DFT}}_{\text{truncated}}. \quad (11)$$

We note that in this work the only calculations to use this correction are the S_N2 reaction barriers.

III. RESULTS

A. Pentacene

Projector-based embedding without truncation is exact in the sense that DFT-in-DFT is equivalent to DFT on the whole system. The minute error introduced by the finiteness of the parameter μ was discussed elsewhere[29] and is negligible on the scale of the data presented. In the extreme of truncation where only a few environment functions are retained, μ dependence can be observed, but for chemically reasonable values of λ that are used in this work, this dependence is weak. We use the same value of $\mu = 10^6$ throughout. A simple test of basis-set truncation is to assess the degree of error in such a DFT-in-DFT calculation when truncation is used.

The accuracy of any truncated scheme is determined by the degree of orbital locality. The delocalised electronic structure of polyaromatic molecules is therefore a challenging test case. We present results for PBE-in-PBE[36] calculations on pentacene where the active region is spanned by the 22 local bond orbitals that have significant contributions on a terminal ring.

During our investigations, we found that the performance of the truncation method depended quite strongly on the choice of localisation scheme. For example, we compared standard Pipek-Mezey (PM) localisation[37] with the new intrinsic bond orbital (IBO) localisation scheme of Knizia.[38] From the left panel of Figure 1, it can be seen that basis-set truncation leads to much smaller errors using IBOs than PM orbitals. The errors that arise in these PBE-in-PBE calculations highlight the orbital tails that "leak" out of subsystem A and into the environment.

Knizia's IBO implementation in Molpro[30,31] allows flexibility in the power used for the locality criterion and options for the orbital orthogonaliser. It has been shown that for conjugated systems, the 4th-moment localisation criterion leads to reduced orbital tails;[39] however for our purposes, we find it to have a minimal effect on the PBE-in-PBE error (right panel of Figure 1). Changing from symmetric orthogonalisation to the zero bond-dipole (ZBD) scheme,[40] on the other hand, does seem to offer a significant improvement. We found similar trends for all the other molecules tested in this paper. Overall we find that with high λ (many functions discarded), IBOs always have a smaller DFT-in-DFT error than PM orbitals. For intermediate truncation, the IBOs were closer to the PM errors and no real difference was found with low orbital truncation thresholds. For the remainder of the results in this paper, we use the IBO method with ZBD and 4th-moment localisation.

As shown in the left panel of Figure 2, the error due to truncation decreases systematically as the number of functions retained increases. The truncation works better in large basis sets; this is hardly surprising, but the rate at which the improvement takes place is reassuring and this is demonstrated for the pentacene example (right panel of Figure 2) using the STO-3G, 3-21G, 6-31G, SVP, cc-pVDZ, TZVP, cc-pVTZ, and cc-pVQZ basis sets.[35,41–43]

B. Embedded many-body expansions

Embedded many-body expansions of energies of molecular aggregates and materials provide a promising new avenue for systematically improving accuracy, not by including

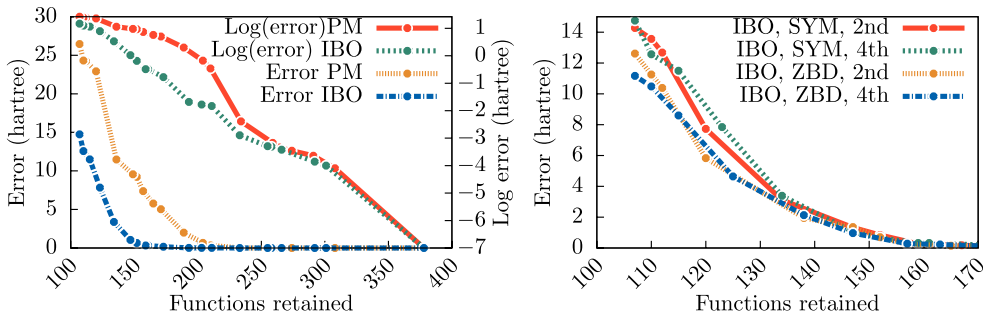

FIG. 1. Errors in the total energy of pentacene using PBE-in-PBE with subsystem A defined by 22 orbitals associated with a terminal ring. Left panel: comparison of the errors using Pipek-Mezey and standard IBOs (using symmetric orthogonalisation and the fourth moment criterion). Right panel: the error in the total energy using four variations of IBOs. All data are computed using PBE/cc-pVDZ. Note that although the absolute errors can be very large when many functions are discarded, they can be brought under control by decreasing the value of λ, hence increasing the number of retained atomic orbitals.

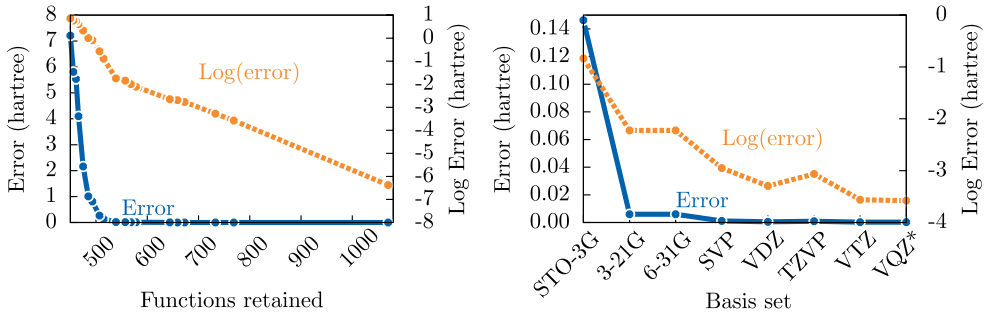

FIG. 2. Left: the error and log error in the total energy of pentacene using PBE-in-PBE with respect to the number of retained basis functions with the cc-pVQZ basis for atoms in subsystem A (C_6H_4) and cc-pVTZ for the remaining atoms.[35] Right: the error in the total energy versus the basis set used. All calculations were performed with approximately the same fraction of functions retained (between 50% and 55%). The orange/green lines show the log error (right hand axis). The blue/red lines show the absolute error (left hand axis). VQZ∗ denotes cc-pVQZ in subsystem A and cc-pVTZ in B.

3-body and higher-order effects explicitly (which is time consuming and in polar systems may not work at all[44]) but by modelling their effect on monomer and dimer energetics with increasingly sophisticated embedding methods.[45,46]

In our previous work,[29] it was shown that the CCSD(T)-in-HF embedded many-body expansion for a model water trimer could accurately reproduce CCSD(T) calculations on the whole system. Using our atom-centred basis-set truncation scheme, we considered the energetics of water hexamers.[32] Here, we do so again, but the advantage of the present work is that the truncation is controlled by the single, geometry-independent truncation threshold λ. The four structures we study are shown in Figure 3. These geometries were optimised at the MP2 level by Santra and co-workers, who report an energetic ordering of prism < cage < book < cyclic.[47]

The counterpoise-corrected coupled-cluster binding energy of a molecular cluster

$$E_{\text{bind}}^{\text{CC}} = E^{\text{CC}} - \sum_i E^{\text{CC}}(i) \tag{12}$$

is the difference between the supermolecular energy (E^{CC}) and the sum of the monomer energies in the supermolecular basis ($E^{\text{CC}}(i)$). To approximate this binding energy without calculating the full hexamer using CCSD(T), one can use the many-body expansion truncated at second order (MBE2), including only the pairwise binding energies

$$E_{\text{bind}}^{\text{MBE2}} = \sum_{i<j} \delta E^{\text{CC}}(i,j) \tag{13}$$

in which the two-body contributions are given by $\delta E^{\text{CC}}(i,j) = E^{\text{CC}}(i,j) - E^{\text{CC}}(i) - E^{\text{CC}}(j)$.

Use of projector embedding for an embedded MBE implies availability of a mean-field calculation on the whole system, so here (as before[29]) we use the MBE only for the correlation contribution to the binding energy and add this quantity to the HF binding energy of the cluster.

In the simplest approximation, the correlation contribution could be treated at the one-body level, to give

$$E_{\text{bind}}^{\text{EMBE1}} = E_{\text{bind}}^{\text{HF}} + \sum_i E_{\text{emb}}^{\text{corr}}(i) - \sum_i E^{\text{corr}}(i), \tag{14}$$

where $E_{\text{emb}}^{\text{corr}}(i)$ is the correlation energy from a CC calculation on monomer i computed in the supermolecular basis and embedded in the environment produced by the other water molecules. The second-order term gives the correlation contribution of the embedded dimers

$$E_{\text{bind}}^{\text{EMBE2}} = E_{\text{bind}}^{\text{EMBE1}} + \sum_{i<j} \delta E_{\text{emb}}^{\text{corr}}(i,j). \tag{15}$$

To reduce computational scaling with system size, each correlation energy is calculated using the basis-set truncation scheme.

The CCSD(T) binding energies for the four hexamer structures are shown in Table I together with MBE2 and EMBE2 approximations, without basis-set truncation. All approaches reproduce the energy ordering found by Santra *et al.*;[47] however, there is a ∼10 mhartree discrepancy between MBE2

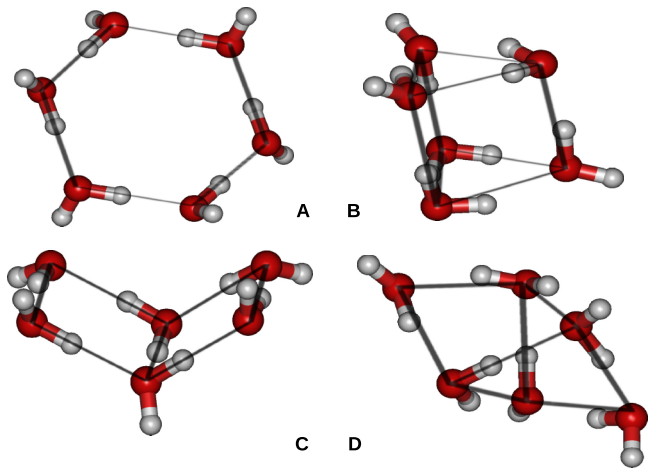

FIG. 3. The four water hexamer structures: (a) cyclic; (b) prism; (c) book; (d) cage.

TABLE I. Binding energies of four water hexamer structures using MBE2, embedded EMBE2 (without basis-set truncation) and full CCSD(T) calculations, and the aug-cc-pVDZ basis set.

	E_{bind}/mhartree		
	MBE2	EMBE2	Full
Prism	−50.7	−65.6	−65.6
Cage	−49.9	−65.1	−65.0
Book	−46.1	−64.4	−64.6
Cyclic	−41.1	−62.9	−63.1

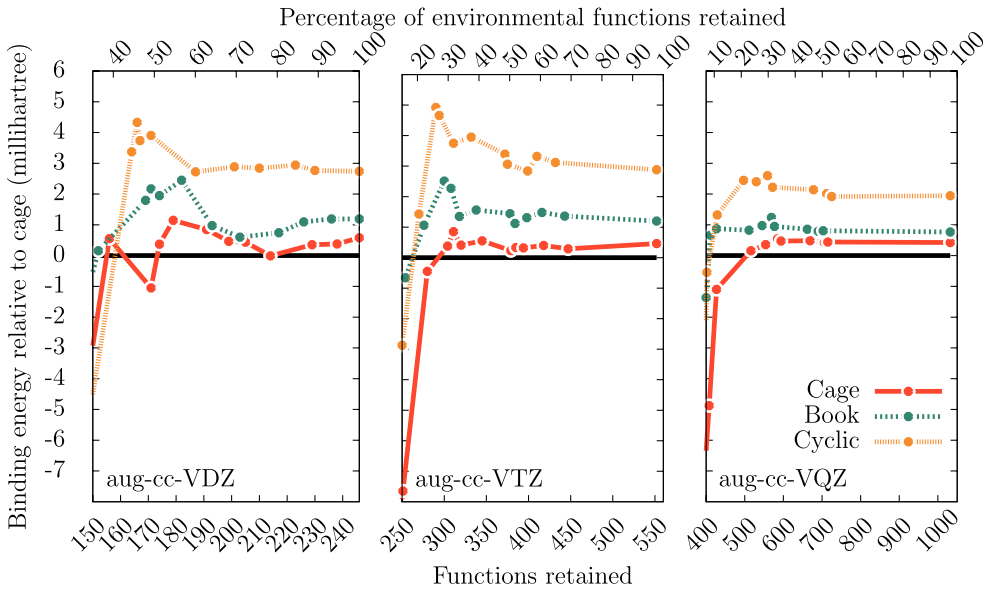

FIG. 4. Binding energies for water hexamer structures relative to those of the prism structure as a function of the average number of basis functions in the correlated calculation on dimers for various choices of the truncation parameter λ. The top scale indicates the percentage of atomic orbitals retained in the environment. Correlation energies were calculated at the CCSD(T) level with the aug-cc-pVDZ (left) and aug-cc-pVTZ (center) bases, and at the CCSD level for the aug-cc-pVQZ basis (right).

and the full CCSD(T) binding energies, due to neglect of three-body effects. The EMBE2 results essentially match full CCSD(T).

Figure 4 shows the hexamer binding energies relative to the prism structure for the aug-cc-pVDZ, aug-cc-pVTZ, and aug-cc-pVQZ basis sets, as a function of number of atomic orbitals retained in the truncation scheme. In all basis sets, EMBE2 returns the correct energetic ordering for the structures in the full basis set. The calculations using only the active-atom subset of atomic orbitals (large λ) fail to model the energetics of these clusters; however, as λ is reduced, and atomic orbitals from neighbouring atoms are included, the results quickly converge to the correct orderings and correct relative energies. In particular, the convergence is stronger in more complete atomic-orbital basis sets. For example, in aug-cc-pVDZ, at least 80% of the functions in the environment are required to converge the energy ordering, whereas in aug-cc-pVQZ, this figure is ~25%. The number of environment atomic orbitals retained for a given degree of convergence is thus similar for both the small and the large basis, and the relative computational savings for the larger basis set are considerably greater.

C. Reaction barriers

Truncated embedding makes it possible to perform wavefunction-in-DFT calculations where the cost of the wavefunction calculation does not scale with the size of the environment.

To demonstrate this, we consider an S_N2 reaction of chloride with alkyl fluorides with increasing chain length. We primarily use a truncation threshold of λ = 0.0005 which we found to offer a reasonable compromise between accuracy and number of functions retained. Subsystem A is spanned by the localised orbitals on the terminal CH_2CH_2F moiety. Using the TZVP basis, 117 atomic orbitals are centred on the atoms of this subsystem and 154 for the corresponding atoms of the transition state. Figure 5 shows that the number of functions remains reasonably constant when the chain length increases with a fixed truncation threshold (0.0005).

The two plots in Figure 6 give the errors of various quantum chemical approaches, relative to the full CCSD(T) reaction barrier. The DFT (or HF) calculations produce various values for the reaction barrier, with PBE resulting in a barrier much lower in energy than CCSD(T). The truncated CCSD(T)-in-DFT calculations largely remove this variability, with all barriers sitting close to that obtained with canonical CCSD(T) on the whole system. The truncation threshold used (0.0005) results in a maximum deviation from the full answer of 3 kcal/mol (M06 on C_8 chain length). Increasing the active region to encompass an extra CH_2 group reduces the error compared to the full CCSD(T) answer to 0.5 kcal/mol for M06 with the same truncation threshold, showing that functional variation can be eliminated through the selection of a sufficiently large active subsystem.

Deeper examination of the errors of truncated embedding is explored in Figure 7, where we compare energy barriers computed using CCSD(T)-in-DFT with and without basis-set truncation. The right-hand plot shows that reducing the trun-

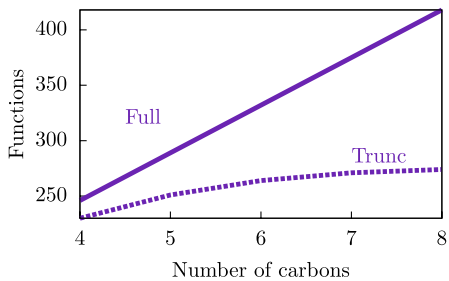

FIG. 5. The number of functions used in the CCSD(T) part of WF-in-DFT calculations on the transition state of S_N2 reactions of chloride with $CH_3(CH_2)_{n-2}CH_2F$ with and without basis-set truncation. The horizontal axis indicates the total number of carbon atoms in the system (n) and a truncation threshold λ = 0.0005 was used.

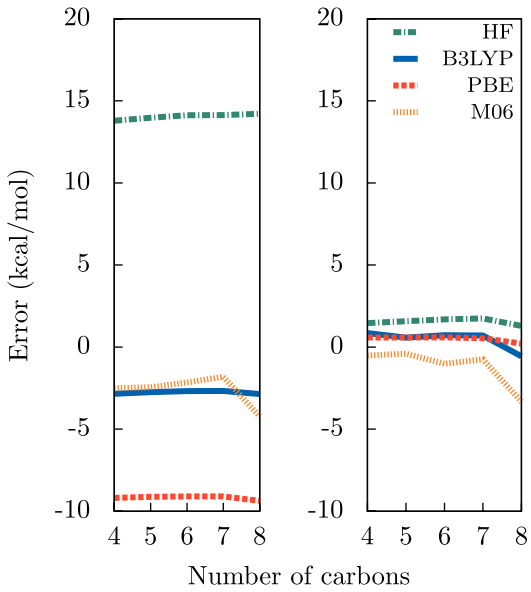

FIG. 6. Errors in barrier heights relative to canonical CCSD(T)/TZVP for S_N2 reactions of chloride with $CH_3(CH_2)_{n-2}CH_2F$, where the horizontal axis indicates the total number of carbon atoms in the system (n). Left: errors in four typical mean-field methods. Right: errors in CCSD(T)-in-DFT with basis-set truncation ($\lambda = 0.0005$).

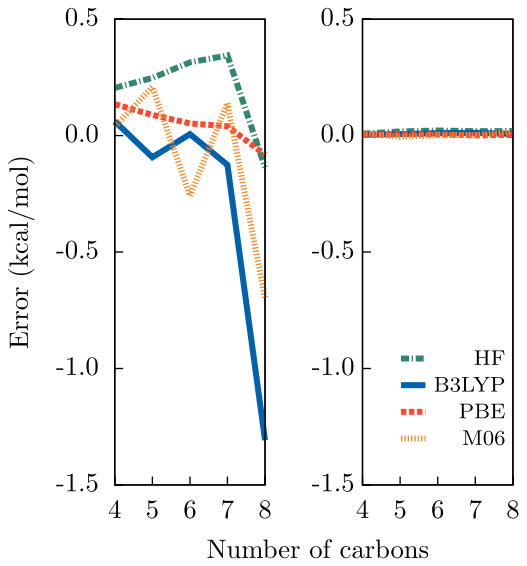

FIG. 7. Errors in reaction barriers arising only from basis-set truncation, for two different truncation thresholds. Each line is the error in truncated CCSD(T)-in-DFT relative to full-basis CCSD(T)-in-DFT. For the C_8 molecule, 21% of environment functions are retained using $\lambda = 0.0005$ and 45% with $\lambda = 0.000\,025$.

cation threshold to 0.000 025 dramatically reduces the error, bringing it below 0.02 kcal/mol in all cases. By reducing λ, the fraction of functions retained in the C_8 chain is increased from 21% for λ = 0.0005 to 45% of environment functions for λ = 0.000 025, but this leads to a reduction of the truncation error of two orders of magnitude.

D. Ketene dissociation

So far it has been demonstrated that truncated embedding is accurate (for a reasonable λ) across a range of static molecular structures. To show that truncation is robust for changes in geometry, we present the dissociation of the alkene bond in 1-penten-1-one. It was previously shown[10] that projector-based embedding accurately reproduces the full CCSD(T) curve and avoids artifacts sometimes seen in local correlation methods for such dissociations.[48] Here, we present results for a 22-electron active subsystem encompassing the —CH═C═O moiety. The geometry was optimized using B3LYP/TZVP. Figure 8 shows CCSD-in-B3LYP energies using both the full TZVP basis and a truncated basis. We initially selected the number of functions for the equilibrium geometry (181 out of 234 for λ = 0.0001, 61% of environmental functions) and lock this selection for all subsequent geometries. We find that the truncated calculation at all points has an error of around 1 mhartree, shows no discontinuities, and is highly parallel to the full-basis curve.

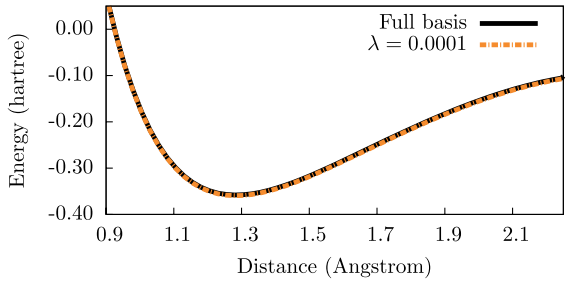

FIG. 8. Binding energy relative to isolated molecular fragments for CCSD-in-B3LYP for the full basis projector embedding and for an atomic-orbital truncation threshold of $\lambda = 0.0001$.

IV. CONCLUSIONS

Projector-based embedding allows for simple, general, and robust embedding of wavefunction methods in DFT models of the chemical environment. Truncation of the atomic-orbital basis set used in the wavefunction calculation makes the cost of what is typically the most expensive part of the calculation independent of the size of the environment, opening the possibility of applications on much more complex systems.

Although this was achieved previously,[32] here we have introduced and tested a simpler basis-set truncation scheme in which a single truncation parameter is used to select the subset of atomic orbitals retained, allowing for smooth tuning between the subset strictly localised on the atoms in subsystem A and the full basis. The method automatically adjusts the selection of retained functions depending on the type of bonding that connects subsystems A and B by using net Mulliken populations of the atomic orbitals to determine their importance. Truncation thresholds that give chemically acceptable errors lead to vast reduction of computational time, and the gains only increase as the size and complexity of the environment increase.

ACKNOWLEDGMENTS

The authors would like to thank David Tew, Gerald Knizia, Taylor Barnes, and Chris Taylor for helpful discussions. Analysis in this work was greatly helped through the use of the Gabedit suite.[49] Some of this work was performed while F.R.M. was on sabbatical at Caltech; support for that visit from

the Institute for Advanced Studies at the University of Bristol (University Research Fellowship) is gratefully acknowledged. We are grateful to EPSRC for funding for this work through research grants (Grant Nos. EP/K018965/1 and EP/J012742/1) and the Doctoral Training Grant. T.F.M. acknowledges support from the National Science Foundation CAREER Award under Grant No. CHE-1057112. All data from this work are held in an open-access repository.[50]

[1]P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[2]W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[3]P. Sherwood, A. H. de Vries, S. J. Collins, S. P. Greatbanks, N. A. Burton, M. A. Vincent, and I. H. Hillier, Faraday Discuss. **106**, 79 (1997).
[4]J. Gao, P. Amara, C. Alhambra, and M. J. Field, J. Phys. Chem. A **102**, 4714 (1998).
[5]H. Lin and D. G. Truhlar, Theor. Chem. Acc. **117**, 185 (2007).
[6]T. A. Wesolowski and A. Warshel, J. Phys. Chem. **97**, 8050 (1993).
[7]G. Senatore and K. R. Subbaswamy, Phys. Rev. B **34**, 5754 (1986).
[8]P. Cortona, Phys. Rev. B **44**, 8454 (1991).
[9]S. R. Pruitt, M. A. Addicoat, M. A. Collins, and M. S. Gordon, Phys. Chem. Chem. Phys. **14**, 7752 (2012).
[10]J. D. Goodpaster, T. A. Barnes, F. R. Manby, and T. F. Miller, J. Chem. Phys. **140**, 18A507 (2014).
[11]N. Govind, Y. A. Wang, A. J. R. da Silva, and E. A. Carter, Chem. Phys. Lett. **295**, 129 (1998).
[12]N. Govind, Y. A. Wang, and E. A. Carter, J. Chem. Phys. **110**, 7677 (1999).
[13]J. D. Goodpaster, N. Ananth, F. R. Manby, and T. F. Miller, J. Chem. Phys. **133**, 084103 (2010).
[14]J. D. Goodpaster, T. A. Barnes, and T. F. Miller, J. Chem. Phys. **134**, 164108 (2011).
[15]J. D. Goodpaster, T. A. Barnes, F. R. Manby, and T. F. Miller, J .Chem. Phys. **137**, 224113 (2012).
[16]S. Fux, C. R. Jacob, J. Neugebauer, L. Visscher, and M. Reiher, J. Chem. Phys. **132**, 164101 (2010).
[17]J. Nafziger, O. Wu, and A. Wasserman, J. Chem. Phys. **135**, 234101 (2011).
[18]C. Huang, M. Pavone, and E. A. Carter, J. Chem. Phys. **134**, 154110 (2011).
[19]L. Wang and M. P. Teter, Phys. Rev. B **45**, 13196 (1992).
[20]M. J. G. Peach, D. G. J. Griffiths, and D. J. Tozer, J. Chem. Phys. **136**, 144101 (2012).
[21]A. Borgoo, A. M. Teale, and D. J. Tozer, Phys. Chem. Chem. Phys. **16**, 14578 (2014).
[22]V. V. Karasiev and S. B. Trickey, Comput. Phys. Commun. **183**, 2519 (2012).
[23]A. W. Götz, S. M. Beyhan, and L. Visscher, J. Chem. Theory Comput. **5**, 3161 (2009).
[24]J. Xia, C. Huang, I. Shin, and E. A. Carter, J. Chem. Phys. **136**, 084102 (2012).
[25]Q. Wu and W. Yang, J. Chem. Phys. **118**, 2498 (2003).
[26]C. R. Jacob, J. Chem. Phys. **135**, 244102 (2011).
[27]A. D. Becke and E. R. Johnson, J. Chem. Phys. **124**, 221101 (2006).
[28]V. N. Staroverov, G. E. Scuseria, and E. R. Davidson, J. Chem. Phys. **124**, 141103 (2006).
[29]F. R. Manby, M. Stella, J. D. Goodpaster, and T. F. Miller, J. Chem. Theory Comput. **8**, 2564 (2012).
[30]H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, 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, Y. Liu, A. W. Lloyd, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, 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, and M. Wang, MOLPRO, version 2012.1, a package of *ab initio* programs, 2012, see http://www.molpro.net.
[31]H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, and M. Schütz, WIREs: Comput. Mol. Sci. **2**, 242 (2012).
[32]T. A. Barnes, J. D. Goodpaster, F. R. Manby, and T. F. Miller, J. Chem. Phys. **139**, 024103 (2013).
[33]I. Mayer, *Simple Theorems, Proofs, and Derivations in Quantum Chemistry* (Springer, 2003).
[34]R. McWeeny, Rev. Mod. Phys. **32**, 335 (1960).
[35]T. H. Dunning, J. Chem. Phys. **90**, 1007 (1989).
[36]J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[37]J. Pipek and P. G. Mezey, J. Chem. Phys. **90**, 4916 (1989).
[38]G. Knizia, J. Chem. Theor. Comput. **9**, 4834 (2013).
[39]I.-M. Høyvik, B. Jansik, and P. Jørgensen, J. Chem. Phys. **137**, 224114 (2012).
[40]D. N. Laikov, Int. J. Quantum Chem. **111**, 2851 (2011).
[41]W. J. Hehre, R. F. Stewart, and J. A. Pople, J. Chem. Phys. **51**, 2657 (1969).
[42]J. S. Binkley, J. A. Pople, and W. J. Hehre, J. Am. Chem. Soc. **102**, 939 (1980).
[43]F. Weigend and R. Ahlrichs, Phys. Chem. Chem. Phys. **7**, 3297 (2005).
[44]R. M. Richard, K. U. Lao, and J. M. Herbert, J. Chem. Phys. **141**, 014108 (2014).
[45]M. J. Gillan, D. Alfé, P. J. Bygrave, C. R. Taylor, and F. R. Manby, J. Chem. Phys. **139**, 114101 (2013).
[46]P. J. Bygrave, N. L. Allan, and F. R. Manby, J. Chem. Phys. **137** (2012).
[47]B. Santra, A. Michaelides, M. Fuchs, A. Tkatchenko, C. Filippi, and M. Scheffler, J. Chem. Phys. **129**, 194111 (2008).
[48]N. J. Russ and T. D. Crawford, J. Chem. Phys. **121**, 691 (2004).
[49]A.-R. Allouche, J. Comput. Chem. **32**, 174 (2011).
[50]See http://dx.doi.org/10.5523/bris.ugahdopfnzk51i6fkkxxuhu99 for all data pertaining to this work.