Exact density functional and wave function embedding schemes based on orbital localization

Bence Hégely, Péter R. Nagy, György G. Ferenczy, and Mihály Kállay

Citation: The Journal of Chemical Physics **145**, 064107 (2016); doi: 10.1063/1.4960177
View online: http://dx.doi.org/10.1063/1.4960177
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/145/6?ver=pdfcov
Published by the AIP Publishing

Exact density functional and wave function embedding schemes based on orbital localization

Bence Hégely,[1] Péter R. Nagy,[1] György G. Ferenczy,[2,3] and Mihály Kállay[1,a)]
[1]*MTA-BME Lendület Quantum Chemistry Research Group, Department of Physical Chemistry and Materials Science, Budapest University of Technology and Economics, P.O. Box 91, H-1521 Budapest, Hungary*
[2]*Medicinal Chemistry Research Group, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Magyar tudósok körútja 2, H-1117 Budapest, Hungary*
[3]*Department of Biophysics and Radiation Biology, Semmelweis University, Tűzoltó u. 37-47, H-1094 Budapest, Hungary*

(Received 1 April 2016; accepted 20 July 2016; published online 11 August 2016)

Exact schemes for the embedding of density functional theory (DFT) and wave function theory (WFT) methods into lower-level DFT or WFT approaches are introduced utilizing orbital localization. First, a simple modification of the projector-based embedding scheme of Manby and co-workers [J. Chem. Phys. **140**, 18A507 (2014)] is proposed. We also use localized orbitals to partition the system, but instead of augmenting the Fock operator with a somewhat arbitrary level-shift projector we solve the Huzinaga-equation, which strictly enforces the Pauli exclusion principle. Second, the embedding of WFT methods in local correlation approaches is studied. Since the latter methods split up the system into local domains, very simple embedding theories can be defined if the domains of the active subsystem and the environment are treated at a different level. The considered embedding schemes are benchmarked for reaction energies and compared to quantum mechanics (QM)/molecular mechanics (MM) and vacuum embedding. We conclude that for DFT-in-DFT embedding, the Huzinaga-equation-based scheme is more efficient than the other approaches, but QM/MM or even simple vacuum embedding is still competitive in particular cases. Concerning the embedding of wave function methods, the clear winner is the embedding of WFT into low-level local correlation approaches, and WFT-in-DFT embedding can only be more advantageous if a non-hybrid density functional is employed. *Published by AIP Publishing.* [http://dx.doi.org/10.1063/1.4960177]

I. INTRODUCTION

Multilevel methods are extensively applied to describe local events in extended electronic systems. They are well suited to treat various phenomena including chemical reactions and electronic excitations in both gas and condensed phases. These methods combine an accurate, high-level quantum mechanical (QM) method for the description of changes in the electronic properties and take into account the effect of the environment by a computationally efficient lower-level method. The first multilevel methods[1,2] combined semi-empirical wave functions with molecular mechanics (MM). Recent approaches also apply density functional theory (DFT) or *ab initio* wave function theory (WFT) including accurate correlated wave functions for the higher-level methods, moreover, embedding into various level DFT and WFT methods has also been developed.[3–9]

The coupling of the different computational schemes is a critical point in multilevel methods. Standard wave function and Kohn–Sham (KS) DFT methods apply orthonormal orbitals, and this leads to significant simplifications both in the formalism and in the computation. However, when different methods are coupled, orthonormality may not be automatically guaranteed. Therefore, significant efforts have been devoted to derive multilevel methods where molecular orbitals (MOs) calculated with different QM methods are orthogonal. In the context of QM/MM methods with fixed orbitals at the subsystem boundary, the orthogonality between the fixed and optimized orbitals is achieved either with basis set orthogonalization[10–15] or with the application[16,17] of the Huzinaga equation.[18] The latter in its most general form can be written as

$$\left(\hat{F} - \hat{P}\hat{F} - \hat{F}\hat{P}\right)|\tilde{\phi}_i^{\mathrm{A}}\rangle = \tilde{\varepsilon}_i^{\mathrm{A}}|\tilde{\phi}_i^{\mathrm{A}}\rangle, \tag{1}$$

where $\hat{F}$ is the Fockian composed of the optimized MOs $\tilde{\phi}_i^{\mathrm{A}}$ of the embedded system A with $\tilde{\varepsilon}_i^{\mathrm{A}}$'s as the corresponding eigenvalues, and

$$\hat{P} = \sum_{i\in\mathrm{B}} \left|\phi_i^{\mathrm{B}}\right\rangle \left\langle\phi_i^{\mathrm{B}}\right| \tag{2}$$

projects onto space B spanned by frozen orbitals ϕ_i^{B}. A distinctive feature of the operator in Eq. (1) is that it is commutable with projector $\hat{P}$ irrespective of the choice of the latter. It means that solving Eq. (1) always guarantees that MOs orthogonal to those in set B can be obtained. It is worth mentioning here that, in addition to solving Eq. (1) for subsystem A with a fixed environment represented by subsystem B, Eq. (1) can also be solved self-consistently for the two subsystems. Moreover, it can also be extended for several frozen subsystems, and the subsystems can be treated at different levels of theory.

a)Electronic mail: kallay@mail.bme.hu

DFT-in-DFT and WFT-in-DFT methods face a similar problem in ensuring orthogonality between the orbitals representing the embedded and embedding electron density. Without orthogonality or in an orbital free representation of the embedding potential, the sum of the subsystem electron densities does not agree with the total density, and the kinetic energy contains a nonadditive term that can be calculated with approximate nonadditive functionals[19,20] or with optimized effective potential methods.[21–27] Alternative techniques proposed by Rajchel *et al.*[28] and by Manby and co-workers[29–32] represent the embedded electron density by MOs (approximately) orthogonal to those from which the embedding density is computed. The basic idea is the creation of a Hermitian operator whose eigenfunctions include the embedding and optimized orbitals with different eigenvalues. In Refs. 28 and 29 it was proposed to complement the Fockian of the embedded system by projector $\hat{P}$ and solving the

$$\left(\hat{F} + \mu\hat{P}\right)|\tilde{\phi}_i^{\mathrm{A}}\rangle = \tilde{\varepsilon}_i^{\mathrm{A}}|\tilde{\phi}_i^{\mathrm{A}}\rangle \tag{3}$$

eigenvalue equation, where μ is a large number. This type of eigenvalue equation has been extensively used in frozen core calculations[33] and is related to the Phillips–Kleinman pseudopotential[34] and also to the Huzinaga equation,[18] Eq. (1), but has less favorable properties than the latter. When the ϕ_i^{B} orbitals are eigenfunctions of $\hat{F}$, then $\hat{P}$ commutes with $\hat{F}$ and also with $\hat{F} + \mu\hat{P}$, the operator in Eq. (3). Consequently, there exists a set of orthogonal MOs that are eigenfunctions of both $\hat{F} + \mu\hat{P}$ and $\hat{P}$, and $\tilde{\phi}_i^{\mathrm{A}}$ orbitals orthogonal to ϕ_i^{B}'s can be obtained by solving Eq. (3). However, when $\hat{F}$ and $\hat{P}$ do not commute, orthogonal ϕ_i^{B} and $\tilde{\phi}_i^{\mathrm{A}}$ orbitals cannot be obtained with a finite μ.

A promising new approach to obtain orthogonal MOs is offered by the embedded mean-field theory (EMFT) developed by Miller III *et al.*[35] In EMFT the system is partitioned at the basis set level allowing for the combination of various mean field theories applied to atoms partitioned according to the basis set-atom assignment. The approach is parameter free and was shown to perform advantageously even for partitions across multiple or aromatic bonds.[35] Its major drawback is that it only enables DFT-in-DFT-type embedding, whereas it is not clear how to carry out WFT-in-DFT-type calculations since the orbitals cannot be unambiguously assigned to the subsystems.

WFT methods are most frequently embedded in DFT approaches, but the embedding into a less expensive WFT approach is also an alternative. A plausible choice for this purpose is the use of local correlation models, which are these days serious competitors to DFT methods concerning both accuracy and speed. Since the pioneering studies in the eighties,[36–38] a number of local correlation approaches have been developed (see, e.g., Refs. 39–47 for representative examples). The fragmentation-based local correlation models decompose the system into small fragments, which are treated by conventional correlation methods. Another essential part of the local correlation approaches describes the system without dividing it into fragments, but local domains of atoms and MOs or other functions are still used in the calculations. In both cases, trivial embedding theories can be defined if the fragments or domains are treated at different levels. This possibility has already been explored by several authors,[48–51] but the arising WFT-in-WFT embedding theories have not yet been compared to the corresponding WFT-in-DFT schemes.

In this paper, we introduce several exact embedding approaches which are applicable to DFT-in-DFT, WFT-in-DFT, and WFT-in-WFT embedding. In Sec. II A we provide a somewhat different derivation of the projector-based embedding approach of Manby and co-workers,[29,30] which will facilitate the further discussion, and we also present our proposal for the modifications of the theory. In Sec. II B, we discuss the possibilities for embedding wave function approaches into local correlation methods. The performance of the various schemes will be compared in Sec. III.

II. THEORY

A. Huzinaga-equation-based embedding

Let us divide the system into two subsystems defined by specifying the corresponding atoms: an embedded (active) subsystem hereafter denoted by A and its environment labeled by B. The subsystems are treated with a high- and low-level DFT method, respectively, or alternatively for subsystem A a correlated WFT method can also be used. The self-consistent field (SCF) energy, which can be either KS or Hartree–Fock (HF), for both subsystems can be written in the

$$E_S[\mathbf{D}] = \mathrm{Tr}(\mathbf{hD}) + G_S[\mathbf{D}] \qquad S = 1, 2 \tag{4}$$

form, where 1 and 2 refer to the SCF method used for subsystems A and B, respectively, $\mathbf{D}$ is a one-particle density matrix, $\mathbf{h}$ is the core Hamiltonian, and G_S includes the two-electron terms. The latter can be decomposed as

$$G_S[\mathbf{D}] = J[\mathbf{D}] + E_{\mathrm{xc},S}[\mathbf{D}], \tag{5}$$

where J is the Coulomb energy and $E_{\mathrm{xc},S}$ is the exchange-correlation functional including also the exact exchange term, if any. Following the work of Manby and co-workers,[29,30] we suppose that the SCF equations are solved for the entire system with the low-level DFT method defined by E_2, and the occupied MOs are localized using some localization criterion. For each localized MO (LMO), the atoms are identified on which the orbital is localized, e.g., an atom is assigned to an orbital if the corresponding Mulliken population of the MO is greater than a threshold. Then the LMOs can be classified into two groups: the MOs ϕ_i^{A}, which are localized on the atoms of the embedded subsystem and the remaining orbitals, ϕ_i^{B}, which will be kept frozen in the subsequent calculations. This assignment also fixes the number of electrons in the subsystems, which will be denoted by n_{A} and n_{B}, respectively. We can also compute the density matrices $\mathbf{D}^{\mathrm{A}}$ and $\mathbf{D}^{\mathrm{B}}$ of the A and B subsystems, respectively, from orbitals ϕ_i^{A} and ϕ_i^{B}, and obviously for the density matrix of the entire system, $\mathbf{D}^{\mathrm{AB}}$, the

$$\mathbf{D}^{\mathrm{AB}} = \mathbf{D}^{\mathrm{A}} + \mathbf{D}^{\mathrm{B}} \tag{6}$$

identity holds.

In the next step, the SCF equations are solved for the active subsystem of n_{A} electrons with an embedding potential describing the effect of the environment to calculate

the embedded energy of the entire system, E_{12}. An exact embedding can be defined if we consider the energy of the entire system calculated in the first SCF run and "replace" the contribution of the $\mathbf{D}^{\mathrm{A}}$ subsystem density by that of the corresponding self-consistent density $\tilde{\mathbf{D}}^{\mathrm{A}}$ computed with the higher-level DFT method in the second SCF run. In other words, E_{12} can be evaluated as

$$E_{12}[\tilde{\mathbf{D}}^{\mathrm{A}};\mathbf{D}^{\mathrm{A}},\mathbf{D}^{\mathrm{B}}] = E_2[\mathbf{D}^{\mathrm{AB}}] + E_1[\tilde{\mathbf{D}}^{\mathrm{A}}] - E_2[\mathbf{D}^{\mathrm{A}}]. \quad (7)$$

Note that the subscript of the energy refers to the method, and the superscript of the density refers to the subsystem; thus E_1 is the energy with the higher-level method, $\mathbf{D}^{\mathrm{A}}$ is the density of subsystem A with the lower-level method, and $\tilde{\mathbf{D}}^{\mathrm{A}}$ designates the density of subsystem A with the higher-level method. Provided that the reoptimized orbitals of subsystem A, $\tilde{\phi}_i^{\mathrm{A}}$, are orthogonal to ϕ_i^{B}, the Fock matrix that is compatible with Eq. (7) can be defined as

$$\mathbf{F} = \frac{\partial E_{12}}{\partial \tilde{\mathbf{D}}^{\mathrm{A}}} + \frac{\partial E_{12}}{\partial \mathbf{D}^{\mathrm{A}}}. \quad (8)$$

To ensure the orthogonality, Manby and co-workers proposed to solve the SCF equations with a modified Fock matrix $\mathbf{F} + \mu\mathbf{P}$, where μ is a level-shift parameter, $\mathbf{P} = \mathbf{SD}^{\mathrm{B}}\mathbf{S}$ is the matrix of the projector in Eq. (2) containing the localized occupied MOs of the environment, and $\mathbf{S}$ is the atomic orbital overlap matrix. If μ is infinity, the orbitals $\tilde{\phi}_i^{\mathrm{A}}$ are orthogonal to ϕ_i^{B}'s, and the embedding is exact; however, in practice, μ can only take a finite value. To avoid numerical instability, a value of 10^3 E_h was suggested for practical applications, which causes an error of a couple of tens of $\mu\mathrm{E}_h$ in the final energies.

The embedded energy can be further improved if E_{12} is augmented with a correction term, which is a first-order correction to the difference between $E_{12}[\tilde{\mathbf{D}}^{\mathrm{A}};\tilde{\mathbf{D}}^{\mathrm{A}},\mathbf{D}^{\mathrm{B}}]$ and $E_{12}[\tilde{\mathbf{D}}^{\mathrm{A}};\mathbf{D}^{\mathrm{A}},\mathbf{D}^{\mathrm{B}}]$. Thus, the final DFT-in-DFT energy expression reads as

$$E_{12}^{\mathrm{DFT}}[\tilde{\mathbf{D}}^{\mathrm{A}};\mathbf{D}^{\mathrm{A}},\mathbf{D}^{\mathrm{B}}] = E_{12}[\tilde{\mathbf{D}}^{\mathrm{A}};\mathbf{D}^{\mathrm{A}},\mathbf{D}^{\mathrm{B}}] + \mathrm{Tr}\left[\left(\tilde{\mathbf{D}}^{\mathrm{A}} - \mathbf{D}^{\mathrm{A}}\right)\frac{\partial E_{12}}{\partial \mathbf{D}^{\mathrm{A}}}\right], \quad (9)$$

where the last term obviously does not influence the exactness of the embedding. In the case of WFT-in-DFT embedding, a calculation is carried out with the WFT method using the above Fock matrix and keeping orbitals ϕ_i^{B} frozen, and the WFT-in-DFT energy is evaluated as

$$E_{12}^{\mathrm{WFT}}[\Psi^{\mathrm{A}};\tilde{\mathbf{D}}^{\mathrm{A}},\mathbf{D}^{\mathrm{A}},\mathbf{D}^{\mathrm{B}}] = E_2[\mathbf{D}^{\mathrm{AB}}] + E_1^{\mathrm{WFT}}[\Psi^{\mathrm{A}};\tilde{\mathbf{D}}^{\mathrm{A}},\mathbf{D}^{\mathrm{A}},\mathbf{D}^{\mathrm{B}}] - E_2[\mathbf{D}^{\mathrm{A}}] - \mathrm{Tr}\left[\mathbf{D}^{\mathrm{A}}\frac{\partial E_{12}}{\partial \mathbf{D}^{\mathrm{A}}}\right], \quad (10)$$

where E_1^{WFT} is the total energy evaluated with the WFT method and Fock matrix $\mathbf{F}$, and Ψ^{A} is the corresponding wave function.

In this work, most of the constituents of the projector-based embedding theory are retained, but we propose a theoretical modification, and some algorithmic changes are also put forth to improve the selection of the orbital spaces. The former modification affects the SCF procedure: instead of solving the KS or HF equations with the modified Fock matrix $\mathbf{F} + \mu\mathbf{P}$, we solve the Huzinaga equations, Eq. (1), whose matrix form reads as

$$\left(\mathbf{F} - \mathbf{SD}^{\mathrm{B}}\mathbf{F} - \mathbf{FD}^{\mathrm{B}}\mathbf{S}\right)\tilde{\mathbf{C}}^{\mathrm{A}} = \mathbf{S}\tilde{\mathbf{C}}^{\mathrm{A}}\tilde{\mathbf{E}}^{\mathrm{A}}, \quad (11)$$

where $\tilde{\mathbf{C}}^{\mathrm{A}}$ contains the MO coefficients of the $\tilde{\phi}_i^{\mathrm{A}}$ orbitals, and $\tilde{\mathbf{E}}^{\mathrm{A}}$ is a diagonal matrix with the corresponding $\tilde{\varepsilon}_i^{\mathrm{A}}$ orbital energies on its diagonal. As discussed above, the $\mathbf{F} - \mathbf{SD}^{\mathrm{B}}\mathbf{F} - \mathbf{FD}^{\mathrm{B}}\mathbf{S}$ Huzinaga matrix is Hermitian and commutes with the $\mathbf{SD}^{\mathrm{B}}\mathbf{S}$ projector. Consequently, the ϕ_i^{B} orbitals are eigenfunctions of the Huzinaga matrix, and the $\tilde{\phi}_i^{\mathrm{A}}$ orbitals will be orthogonal to them. It can be shown[18] that the $\tilde{\phi}_i^{\mathrm{A}}$ orbitals satisfying Eq. (11) make the energy in Eq. (9) stationary with the condition that they form an orthonormal set and they are orthogonal to the fixed ϕ_i^{B} orbitals.

The use of the Huzinaga equations eliminates the bias caused by the arbitrary level shift parameter μ and makes the embedding "strictly exact." Nevertheless, the difference of the numerical results obtained with the original and the modified approach is expected to be small. The computational complexity of the two schemes is very similar. The extra price we have to pay for the solution of the Huzinaga equations instead of the level-shifted KS or HF equations is just one matrix multiplication per iteration step in place of a matrix addition, which incurs a negligible overhead with respect to other operations in an SCF procedure (see the supplementary material[53] for timings).

Besides this theoretical improvement, we also tested new selection schemes for the active and frozen orbitals since our experience showed that for relatively complicated systems, such as conjugated and aromatic molecules, the simple Mulliken-population-based algorithm can select the wrong orbitals or, in the case of a reaction, the number of orbitals is not compatible for the educt and product. We have found that the Boughton–Pulay (BP) algorithm,[52] which is widely used in local correlation methods for the assignment of atoms to LMOs, with a completeness criterion of 0.985 helps to overcome these problems for conjugated systems. With these modifications, the embedding scheme outlined above can be used as a black-box method for most molecules except for extended aromatic systems (see Sec. III for examples). For the latter, we developed a two-level orbital selection strategy. First, the core, σ, and the π orbitals are distinguished: an LMO is regarded as a π orbital if more than 4 atoms are assigned to it by the BP scheme, otherwise it is supposed to be core or σ. The latter orbitals are added to the active subsystem based on the atom assignment of the BP algorithm. For the number of π orbitals, an upper limit is set to avoid the incorrect number of embedded orbitals for the reactants and products. If one of the atoms assigned to a π orbital by the BP algorithm is part of the embedded subsystem, the Mulliken populations of the orbital on the active atoms are summed up, and the given number of orbitals with the largest sums is selected. This selection scheme clearly requires the *a priori* knowledge of the maximum number of π orbitals in subsystem A and is thus not fully black-box. Our test calculations proved that the Mulliken population-based scheme performs well for aromatic systems with active subsystems of reasonable size. However, for extremely small embedded subsystems, which are normally considered only for testing purposes, the

σ orbitals must also be selected with a similar algorithm as the π orbitals.

An important feature of the above method, which is also peculiar to other embedding schemes, is that the density of the environment, $\mathbf{D}^{\mathrm{B}}$, is kept frozen. One would expect better results if $\mathbf{D}^{\mathrm{B}}$ was also relaxed during the SCF procedure, that is, the electrons of the environment also felt the potential of the active electrons treated at a higher level. The EMFT of Miller III *et al.*[35] is an elegant approach, which is free of this problem. In EMFT, the density matrix of the entire system, $\mathbf{D}^{\mathrm{AB}}$, is partitioned into subsystem densities $\mathbf{D}^{\mathrm{A}}$ and $\mathbf{D}^{\mathrm{B}}$ depending on if the corresponding basis functions reside on the atoms of subsystem A or B. The energy is written in the same form as Eq. (7),

$$E_{12}[\mathbf{D}^{\mathrm{AB}}] = E_2[\mathbf{D}^{\mathrm{AB}}] + E_1[\mathbf{D}^{\mathrm{A}}] - E_2[\mathbf{D}^{\mathrm{A}}], \tag{12}$$

which now depends on the total density $\mathbf{D}^{\mathrm{AB}} = \mathbf{D}^{\mathrm{A}} + \mathbf{D}^{\mathrm{B}}$, and the Fock matrix is defined by the derivative of the energy with respect to the latter,

$$\mathbf{F} = \frac{\partial E_{12}}{\partial \mathbf{D}^{\mathrm{AB}}}. \tag{13}$$

We also attempted to develop a self-consistent generalization of the Huzinaga-equation-based embedding scheme along this line. We again suppose that the system is partitioned in the same way as for the projector-based embedding, but now the densities of the two subsystems, $\mathbf{D}^{\mathrm{A}}$ and $\mathbf{D}^{\mathrm{B}}$, are optimized self-consistently in a single SCF run. To that end, in each iteration step of the SCF procedure the occupied orbitals are localized, the LMOs are assigned to the two subsystems, and the $\mathbf{D}^{\mathrm{A}}$ and $\mathbf{D}^{\mathrm{B}}$ density matrices are computed and used to evaluate the energy and the Fock matrix according to the same equations as for EMFT, Eqs. (12) and (13). The converged E_{12} energy is directly our DFT-in-DFT energy, and the definition of the WFT-in-DFT energy is straightforward. This self-consistent approach can also be regarded as an alternative to the EMFT scheme. Compared to EMFT, its major advantage is that it enables both DFT-in-DFT- and WFT-in-DFT-type embedding calculations. We also implemented and tested this approach, but we have found that it does not perform better than the parent Huzinaga-equation-based scheme. It can be explained by that the latter benefits from error cancellation, which has been also realized by Miller III and co-workers.[30] Since the results are not convincing, and, in addition, the self-consistent scheme is numerically less stable, we refrain from the detailed discussion of the algorithm and the results obtained with this approach; further details and selected examples are presented in the supplementary material.[53]

We also note that there is another simple possibility for the self-consistent generalization of the Huzinaga-equation-based embedding scheme. As mentioned in Sec. I, the Huzinaga-equations can be solved repeatedly, i.e., after converging the Huzinaga-equations for the active subsystem, we can change the role of the active and the embedding subsystem and solve the Huzinaga-equations for the environment keeping the active orbitals frozen. This procedure can be repeated until self-consistency. Concerning that this scheme is relatively costly, and taking into account the moderate success of the above self-consistent approach, we did not embark on implementing it.

B. Embedding into local correlation methods

In our fragmentation-based local correlation approach,[47,51,54] which is related to the cluster-in-molecule[55,56] and the incremental schemes,[42,57,58] the MOs are localized, and a domain of LMOs, $\mathcal{E}_i$, is constructed for each occupied LMO i. The approximate contribution of LMO i to the correlation energy, hereafter denoted by $\delta E_i(\mathcal{E}_i)$, is evaluated in this domain. Furthermore, to account for the long-range correlation (LRC) neglected by this approximation, for each i-j pair of occupied LMOs that are not included in any $\mathcal{E}_k$ domain, a pair domain, $\mathcal{P}_{ij}$, is assembled, and an approximate pair correlation energy, $\delta E_{ij}(\mathcal{P}_{ij})$, is evaluated within the pair domain. By summing up these increments, we arrive at our final correlation energy expression,

$$E^{\mathrm{c}} = \sum_i \delta E_i(\mathcal{E}_i) + {\sum_{ij}}' \delta E_{ij}(\mathcal{P}_{ij}), \tag{14}$$

where the prime indicates a restricted summation over the pairs which are not included in any $\mathcal{E}_k$ domain. The construction of the domains and other details can be found in our previous papers[47,54] and are not recapitulated here. In our implementation, the $\delta E_i(\mathcal{E}_i)$ contributions can be computed with the second-order Møller–Plesset (MP2) approach[59] and its spin-scaled variants,[60,61] the direct random-phase approximation (dRPA)[62] and its second-order screened exchange (SOSEX) extension,[63] and arbitrary iterative and perturbative coupled-cluster (CC) methods, such as CC with single and double excitations (CCSD) and CCSD with perturbative triples [CCSD(T)], CCSD with triple excitations (CCSDT), and so on (Ref. 64). The evaluation of the $\delta E_{ij}(\mathcal{P}_{ij})$ pair correlation energies is currently possible at the MP2, opposite-spin MP2, and dRPA levels.

The above partitioning of the correlation energy suggests a trivial embedding scheme: the LMOs can be classified as active or inactive using the algorithms presented in Sec. II A, and the $\delta E_i(\mathcal{E}_i)$ and $\delta E_{ij}(\mathcal{P}_{ij})$ contributions can be evaluated at different levels of theory depending on the type of the corresponding orbitals. An obvious possibility is to compute the $\delta E_i(\mathcal{E}_i)$ contributions at the CCSD(T) and MP2 levels for an active and an inactive i, respectively, and approximate the pair correlation energies also with MP2. The resulting approach, which will be tested here, can be denoted as LCCSD(T)-in-LMP2, where L indicates the local approximation. A further choice is to treat the environment simply at the HF level, that is, to neglect the $\delta E_i(\mathcal{E}_i)$ contributions for the environmental orbitals. Then we have the option to only evaluate the pair correlation energies within the active subsystem, but the resulting LCCSD(T)-in-HF approach does not offer any advantage over the LCCSD(T)-in-DFT-type embedding, which can be easily implemented with the infrastructure described in Sec. II A. On the other hand, we can take into account the LRC between the embedded subsystem and the environment by

also calculating the $\delta E_{ij}(\mathcal{P}_{ij})$ increments, where one of i and j is an inactive LMO. In this study, we also tested this approximation calculating the pair correlation energies at the MP2 level; the LCCSD(T)-in-HF+LRC abbreviation will stand for this approach.

Concerning the speed of the calculations, it is important to note that, for a medium-sized or large molecule, the evaluation of the local MP2 correlation energy is considerably cheaper than the preceding HF calculation even if reduced-scaling algorithms are used for the calculation of the exchange contribution. The calculation of the MP2 pair correlation energies takes just a fraction of the entire computation time. Consequently, the speed of the LCCSD(T)-in-LMP2 and LCCSD(T)-in-HF+LRC calculations is comparable to that for a LCCSD(T)-in-DFT run, where the density functional includes exact exchange.

III. BENCHMARK CALCULATIONS

A. Computational details

The localization-based embedding models proposed in this paper just as the original projector-based embedding scheme[29,30] have been implemented in the Mrcc suite of quantum chemical programs and will be available in the next release of the package.[65] For DFT-in-DFT embedding we evaluate the performance of the Huzinaga-equation-based scheme in comparison to QM/MM and simple vacuum embedding for energy differences. Concerning the embedding of WFT methods, in addition to the latter two schemes WFT-in-DFT embedding is also compared to the LCCSD(T)-in-LMP2 and LCCSD(T)-in-HF+LRC approaches.

The Huzinaga-equation-based scheme was also benchmarked against the parent projector-based approach, and, as expected, we have found that the two methods behave similarly. In accordance with Ref. 29, the difference between total energies obtained with the two approaches is lower than 50 μE_h (see the supplementary material[53] for examples), and the resulting difference in reaction energies is significantly smaller than the intrinsic errors caused by the embedding approximations. Consequently, in the following we only present the results of the Huzinaga-equation-based scheme.

In our calculations, Pople's 6-31G* and 6-311G**[66,67] and Dunning's (augmented) correlation-consistent polarized valence triple-ζ basis set [(aug-)cc-pVTZ] were employed.[68–70] If the aug-cc-pVTZ basis is used for the heavy atoms, but the diffuse functions are removed from the hydrogens, the resulting basis set will be termed aug'-cc-pVTZ. The density-fitting approximation was employed throughout both for the SCF and the subsequent WFT calculations. The auxiliary basis sets were those of Weigend and co-workers optimized at the HF[71] and MP2[72] levels.

The DFT calculations used various functionals: the local (spin) density approximation (LDA),[73,74] the Perdew–Burke–Ernzerhof (PBE)[75] functional, and B3LYP (Becke's three-parameter hybrid functional including the correlation functional of Lee, Yang, and Parr).[76,77] These functionals were chosen because they represent different rungs on the "Jacob's ladder" of the DFT functionals. LDA is the simplest existing functional depending only on the electron density itself. PBE is a prototype of a functional derived using the generalized gradient approximation. In addition to the density it is also a function of its gradient, which makes the approach more accurate but also results in somewhat increased computational expenses with respect to the simple LDA. The most complete functional considered here is B3LYP, which is a hybrid functional, that is, it also contains HF exchange. The inclusion of the latter further improves the performance of the method, but B3LYP is also considerably more costly than the two other functionals due to the expensive exact exchange term.

For the localization of the occupied orbitals the Pipek–Mezey[78] method was invoked. The MO spaces were partitioned by the algorithm discussed in Sec. II A. Though this scheme unambiguously selects the MOs, to avoid any confusion about the partitioning of the systems considered, for each of them we give the serial numbers of the atoms assigned to the localized orbitals and the serial numbers of the localized MOs included in the embedded subsystem in the supplementary material.[53]

For the QM/MM calculations, Mrcc was interfaced to the Amber molecular mechanics code.[79,80] For the evaluation of the MM energy functional, the general Amber force field (GAFF)[81] was used where the corresponding MM parameters were generated by the Antechamber program[82] using the AM1-BCC charge scheme.[83,84] The electrostatic and van der Waals terms were not cut off, and only those bonded terms were retained which contained at least one MM or MM host atom. Residual charges which come from zeroing the charges of the QM and MM host atoms were distributed equally among all MM atoms.

The benchmark test sets were adopted from Ref. 35: the substitution reaction 1-chlorodecane to yield 1-decanol, the deprotonation of decanoic acid, the Diels–Alder reaction between the conjugated octadecanonaene and 1,3-butadiene (see Fig. 2(d) of Ref. 35), and the hydrogenation of pentacene (see Fig. 3(d) of Ref. 35). The DFT-in-DFT, WFT-in-DFT, and WFT-in-WFT calculations used the geometries of Ref. 35 except for OH^- and H_2, whose bond lengths were reoptimized at the same level as in Ref. 35 and the resulting bond lengths of 0.978 Å and 0.744 Å, respectively, were used. For the QM/MM and vacuum embedding calculation the C—C bond that is cut when fragmenting the system was replaced by a C—H bond with a standardized bond length of 1.09 Å pointing in the original direction. In the latter calculations, only the single C—C bonds of octadecanonaene were broken to preserve the conjugated system, and similarly, for pentacene the aromatic rings were not split up, that is, benzene, naphthalene, anthracene, and tetracene were taken as model systems.

In Subsections III B and III C, the reference energies are substracted from reaction energies, which were calculated by the various embedding techniques, and the error is plotted as a function of embedded carbon atoms if not otherwise stated. Throughout the discussion of the results, the absolute errors will be simply referred to as errors. Reference energies are obtained by using the corresponding high-level methods and basis sets for the aforementioned reactions. We also adopt the convention of Ref. 35 that the results presented

for the case where zero carbon atoms are included in the embedded subsystem are the errors of the reaction energies obtained with the low-level method.

It is worth emphasizing that the Huzinaga embedding with the BP algorithm-based system partitioning required no user intervention, such as the manual selection of the orbitals included in the embedded subsystem or setting the number of those orbitals, for any system separation even when conjugated π-electron systems were considered. The only exception was the hydrogenation reaction of pentacene, where the number of π-orbitals must have been specified (*vide infra*). However, other than setting these parameters, no human intervention was required either for pentacene or for its hydrogenated derivative.

B. DFT-in-DFT embedding

1. The substitution of 1-chlorodecane

The first test for the various embedding techniques is the simplest case, where subsystems A and B are connected through a single σ-bond. In this reaction 1-chlorodecane reacts with a hydroxide anion to form 1-decanol and a chloride anion. The results are presented in Fig. 1.

The errors show similar tendencies for the hybrid and the non-hybrid functional, as well as for the smaller and the larger basis set. The error (relative error) of all the embedding techniques is less than 1 kcal/mol (1%) if the embedded subsystem includes at least two carbon atoms. Surprisingly, the cost–efficient vacuum embedding and QM/MM techniques converge rapidly and produce errors similar to those of the more advanced scheme when sufficiently large subsystem A is applied. The error of the Huzinaga embedding is still below 1 kcal/mol when the active subsystem includes only one carbon atom, while the error of the two simple approaches is somewhat larger.

2. The deprotonation of decanoic acid

For the deprotonation reaction of decanoic acid the subsystems are also separated through a single σ-bond, however, the total charge of the reactant (decanoic acid) and the product (decanoate) is different. This makes the test more sensitive because the effect of error cancellation is likely to be less significant. The results are shown in Fig. 2.

As expected, the embedding techniques produce larger errors, and the errors converge more slowly with the enlargement of the embedded subsystem compared with the previous reaction. The Huzinaga embedding provides satisfactory results even for small active subsystems, but its convergence is not monotonic, and for bigger embedded systems the simplistic methods outperform the Huzinaga scheme. The trends are similar for both basis sets, but with the larger basis set the errors of the Huzinaga embedding are slightly smaller. The PBE-in-LDA reaction energies calculated with the Huzinaga approach are accurate to 1 kcal/mol with active subsystems of any size, while the B3LYP-in-LDA embedding requires quite large embedded systems to reach this accuracy. In relative terms, the errors are well below 1% for all the combinations of functionals and basis sets. Nevertheless, if one chooses a sufficiently large embedded region, the vacuum embedding or the QM/MM method is the most cost efficient option.

3. The Diels–Alder reaction of octadecanonaene

The next test reaction is the Diels–Alder addition of 1,3-butadiene and octadecanonaene. The main difference of this example compared with the previous ones is that the system is separated not through simple single C—C bonds but bonds of order higher than one. The modeling of a delocalized, conjugated π system is a serious test for the embedding methods because these systems are sensitive to the polarizing effects of the environment, hence distortions or the truncation of the conjugated π orbitals can cause significant errors. In the case of vacuum embedding and QM/MM, the systems are only partitioned across single bonds to maintain the conjugation in subsystem A. At all the system separations 1,3-butadiene is part of the embedded subsystem. The results are depicted in Fig. 3, where we adopted the convention of Ref. 35: the horizontal axis indicates the number of carbon atoms of the octadecanonaene chain included in the embedded

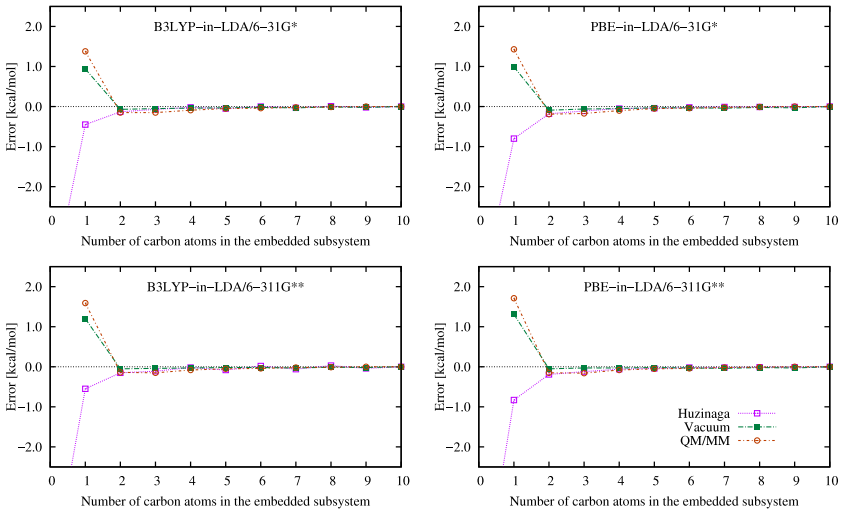

FIG. 1. Error of the reaction energy for the substitution reaction of 1-chlorodecane with various embedding schemes as a function of the number of carbon atoms included in the embedded subsystem using the 6-31G* and 6-311G** basis sets. The B3LYP/6-31G*, B3LYP/6-311G**, PBE/6-31G*, and PBE/6-311G** reference reaction energies are −89.0, −92.1, −88.8, and −91.2 kcal/mol, respectively. The reaction energies of the LDA/6-31G* and LDA/6-311G** low-level methods are −93.7 and −97.5 kcal/mol, respectively.

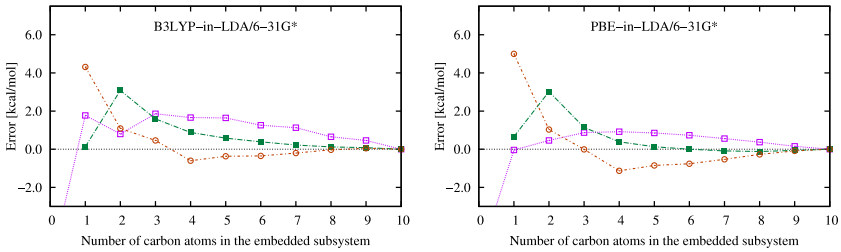

FIG. 2. Error of the reaction energy for the deprotonation of decanoic acid with various embedding schemes as a function of the number of carbon atoms included in the embedded subsystem using the 6-31G* and 6-311G** basis sets. The B3LYP/6-31G*, B3LYP/6-311G**, PBE/6-31G*, and PBE/6-311G** reference reaction energies are 363.3, 359.9, 363.0, and 359.3 kcal/mol, respectively. The reaction energies of the LDA/6-31G* and LDA/6-311G** low-level methods are 357.1 and 351.0 kcal/mol, respectively.

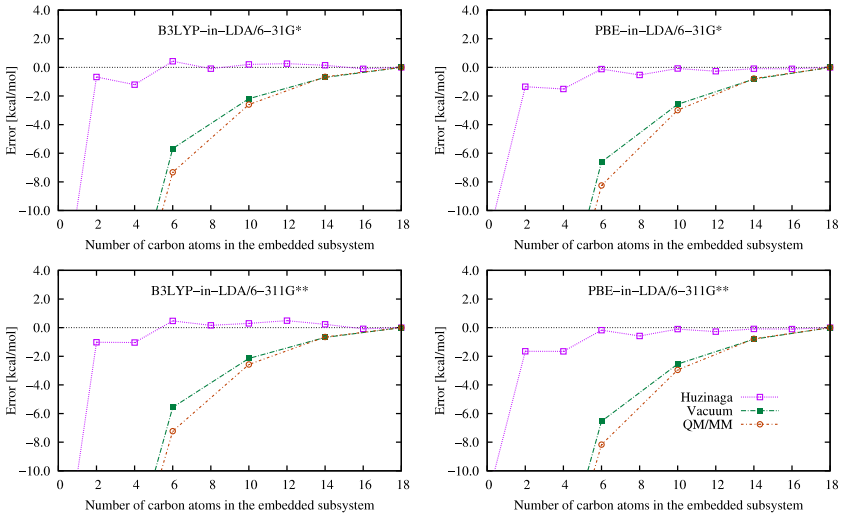

FIG. 3. Error of the reaction energy for the Diels–Alder reaction of 1,3-butadiene and octadecanonaene with various embedding schemes as a function of the number of carbon atoms included in the embedded subsystem using the 6-31G* and 6-311G** basis sets. The B3LYP/6-31G*, B3LYP/6-311G**, PBE/6-31G*, and PBE/6-311G** reference reaction energies are −21.5, −16.3, −27.9, and −23.6 kcal/mol, respectively. The reaction energies of the LDA/6-31G* and LDA/6-311G** low-level methods are −40.1 and −35.6 kcal/mol, respectively.

subsystem, and subsystem A enlarges symmetrically from the central C═C bond, where the reaction takes place.

The results show similar trends with both DFT functionals and basis sets. The Huzinaga embedding performs the best of all the embedding methods because its error (relative error) is well below 1 kcal/mol (3%) if at least 6 carbon atoms are included in the active subsystem, and it does not exceed 1.7 kcal/mol with any embedded system size. The convergence is not smooth with the Huzinaga method, and the errors usually increase when the system is partitioned across double bonds, which can be rationalized by the destruction of the conjugated π system. The vacuum embedding and QM/MM are not competitive since for quantitative accuracy they require that almost the entire molecule be included in the active subsystem.

We note again that the BP-algorithm-based partitioning scheme is completely black-box even for this complicated system as not even the number of orbitals of the embedded subsystem had to be specified for any active subsystem of octadecanonaene, except for the smallest embedded subsystem where we had to set the number of embedded orbitals. It is worth mentioning that the success of this scheme is probably a consequence that the algorithm also selects those π orbitals, which are close to the subsystem boundaries but localized mainly in subsystem B.

4. The hydrogenation of pentacene

The last test system for the embedding methods is the hydrogenation of pentacene. This is the most challenging reaction because the π orbitals of the aromatic system are even more sensitive to polarization compared with the conjugated π orbitals of the previous test. Moreover, the orbitals are also more delocalized, which increases the risk that the automated orbital selection process selects different number of orbitals for subsystem A in the educts and the product. We experienced that the standard BP algorithm greatly suffers from this problem, consequently the two-level orbital selection scheme is applied for these calculations. This orbital selection technique worked well for this test system as no user intervention was necessary other than setting the upper limit for the embedded π orbitals. These orbital limits were incremented by 2 at each new aromatic ring broken except for the smallest embedded subsystem, that is, for the product the limits were 0, 2, 2, 4, 4, 6, 6, 8, 8, 10, and 10 for the embedded subsystems of $2, 4, 6, \ldots, 22$ carbon atoms, respectively, while the above limits were increased by one for pentacene. The results are presented in Fig. 4, where the numbering of the horizontal axis starts from the end of the pentacene molecule where the hydrogen addition occurs.

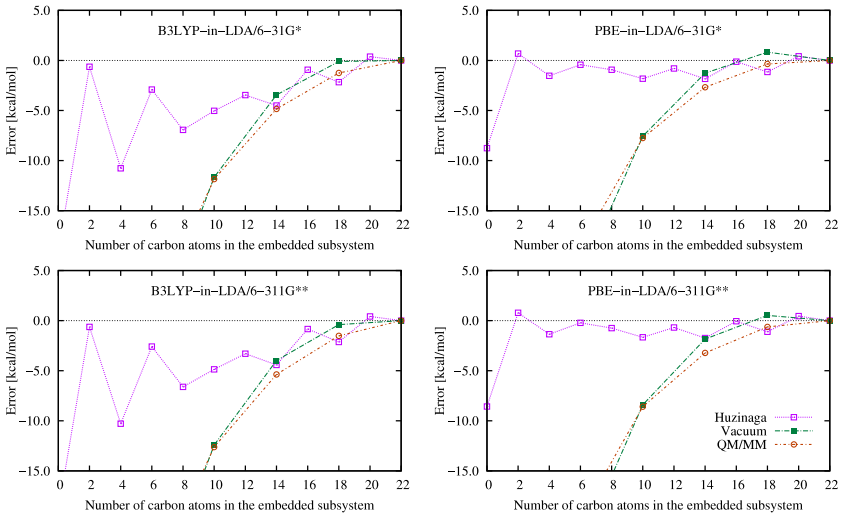

FIG. 4. Error of the reaction energy for hydrogenation of pentacene with the various embedding schemes as a function of the number of carbon atoms included in the embedded subsystem using the 6-31G* and 6-311G** basis sets. The B3LYP/6-31G*, B3LYP/6-311G**, PBE/6-31G*, and PBE/6-311G** reference reaction energies are 32.4, 34.7, 22.2, and 24.3 kcal/mol, respectively. The reaction energies of the LDA/6-31G* and LDA/6-311G** low-level methods are 13.4 and 15.8 kcal/mol, respectively.

In general the behavior of the Huzinaga embedding is similar to the B3LYP and PBE functionals, but the errors are higher and the convergence is slower for the B3LYP-in-LDA embedding. With the smallest embedded subsystem the error is moderate probably due to error compensation, and after a big jump at four carbon atoms it starts to decrease. Surprisingly, for bigger active subsystems the embedding is more accurate when full carbon rings are not part of subsystem A, which can also be explained by error cancellation effects. The error only drops below 1 kcal/mol, which still corresponds to a considerable relative error of 5% (PBE) or 3% (B3LYP) if 20 carbon atoms are part of the embedded subsystem. Nevertheless, the error does not depend on the chosen basis set. The vacuum embedding and QM/MM approaches are again not competitive for small embedded subspaces. However, if chemical accuracy is desired, we can also choose these methods since they reach this limit in the same region as the more advanced method.

All in all, the aromatic systems are separable with the embedding techniques considered if sufficiently large active subsystems are chosen. For qualitative accuracy, the Huzinaga approach is recommended in conjunction with embedded subsystems of medium size. For accurate reaction energies big active subsystems are required, but any of the embedding techniques can be employed.

C. WFT-in-DFT and WFT-in-WFT embedding

The above test set is also used for benchmarking the WFT-in-DFT and the WFT-in-WFT embedding techniques, consequently, the general remarks and technical considerations of Sec. III B also hold here.

1. The substitution of 1-chlorodecane

The results for the substitution reaction calculated by the various embedding techniques are shown in Fig. 5. The results reveal that the performance of all the embedding methods is good because the errors are below 1 kcal/mol, in addition, the relative error is not greater than 1.1%. The LCCSD(T)-in-HF+LRC approach provides the smallest maximum error, followed by the embedding into PBE, LMP2, vacuum, B3LYP, LDA, and MM. As to the convergence rate, the WFT-in-WFT methods perform the best (LMP2 and HF+LRC embedding), followed by the WFT-in-DFT methods (PBE, B3LYP, and LDA), and the cost-efficient techniques are the slowest to converge with respect to the size of subsystem A. However, in the case of the B3LYP-based embedding, the error fluctuates in the (−0.2, 0.2) kcal/mol interval when larger embedded subsystems are applied. All in all, the considered methods prove their capability of handling this simple test case, but the most accurate technique is the LCCSD(T)-in-LMP2 scheme. If one takes into account the computational cost as well, the vacuum embedding seems to be the best option here, followed by the QM/MM method and the PBE-based embedding.

2. The deprotonation of decanoic acid

Fig. 6 presents the errors in the deprotonation energies calculated by the embedding techniques investigated. These results show more pronounced differences among the embedding methods than those for the previous reaction;

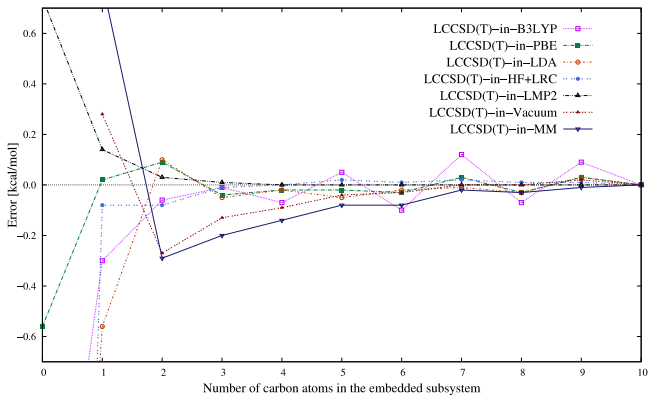

FIG. 5. Error of the reaction energy for the substitution reaction of 1-chlorodecane with various embedding schemes as a function of the number of carbon atoms included in the embedded subsystem using the cc-pVTZ basis set. The LCCSD(T) reference reaction energy is −71.3 kcal/mol. The reaction energies of the B3LYP, PBE, LDA, HF, and LMP2 low-level methods are −73.3, −71.8, −75.4, −78.7, and −70.5 kcal/mol, respectively.

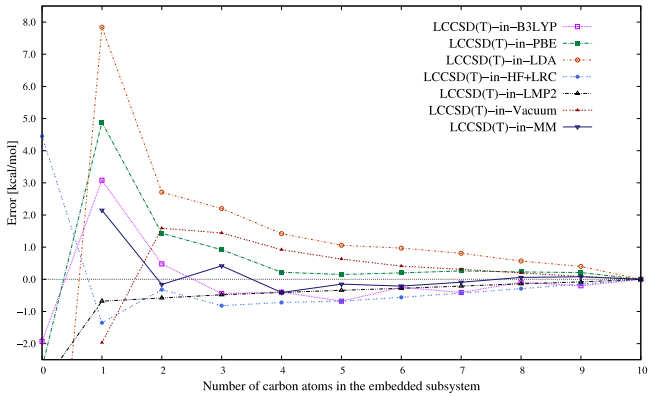

FIG. 6. Error of the reaction energy for the deprotonation reaction of decanoic acid with the various embedding schemes as a function of the number of carbon atoms included in the embedded subsystem using the aug'-cc-pVTZ basis set. The LCCSD(T) reference reaction energy is 353.1 kcal/mol. The reaction energies of the B3LYP, PBE, LDA, HF, and LMP2 low-level methods are 351.2, 350.4, 340.3, 357.6, and 349.9 kcal/mol, respectively.

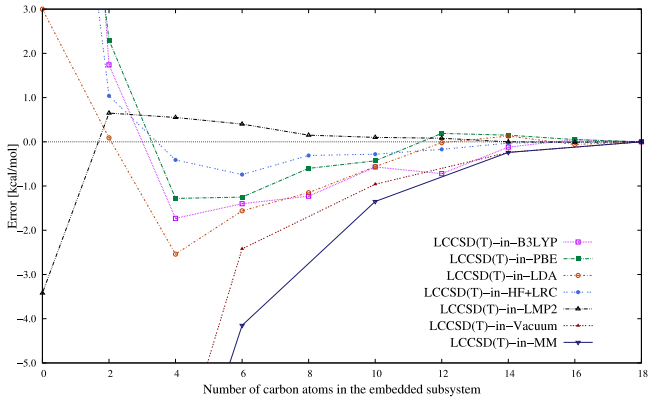

FIG. 7. Error of the reaction energy for the Diels–Alder reaction of 1,3-butadiene and octadecanonaene with various embedding schemes as a function of the number of carbon atoms included in the embedded subsystem using the cc-pVTZ basis set. The LCCSD(T) reference reaction energy is −35.4 kcal/mol. The reaction energies of the B3LYP, PBE, LDA, HF, and LMP2 low-level methods are −13.0, −20.5, −32.4, −21.6, and −38.8 kcal/mol, respectively.

however, if a sufficiently large embedded subsystem is taken, all the techniques are still capable of describing the deprotonation reaction. If we inspect the plot, we find that the LCCSD(T)-in-LMP2 method provides the smallest maximum error (0.7 kcal/mol), which is followed by the HF+LRC, the vacuum, the QM/MM, the B3LYP, the PBE, and the LDA embeddings, respectively. In relative terms, the error of the LCCSD(T)-in-LMP2 approach is less than 0.2%, while that for the other schemes still does not exceed 3%. Interestingly, the fastest method to converge with respect to the size of subsystem A is the QM/MM approximation: if subsystem A consists of at least two carbon atoms, the most accurate embedding scheme is QM/MM, which is also the second cheapest method. However, its convergence is not monotonic and is only marginally better than that for LCCSD(T)-in-LMP2, which is followed by the HF+LRC, the B3LYP, the PBE, the vacuum, and the LDA embeddings, respectively. That is, the WFT-in-WFT schemes are proven to be more accurate than the LCCSD(T)-in-B3LYP method, which is the most complete WFT-in-DFT embedding but only slightly less expensive than the WFT-in-WFT models considered here. As expected, the WFT-in-DFT methods become more accurate as the quality of the exchange-correlation functional increases, but the vacuum embedding is still more accurate and, of course, cheaper than the simplest WFT-in-DFT method, LCCSD(T)-in-LDA.

3. The Diels–Alder reaction of octadecanonaene

Fig. 7 presents the errors of the various embedding schemes calculated for the addition reaction of octadecanonaene and 1,3-butadiene. The results for this reaction show that again the LCCSD(T)-in-LMP2 method produces the smallest maximum error (0.6 kcal/mol = 1.7%), followed by the HF+LRC, PBE, B3LYP, LDA, vacuum, and the MM embedding, respectively. The latter methods are significantly less accurate than LCCSD(T)-in-LMP2, and their maximum errors are higher than 3%, even if most of them seem to benefit from error cancellation for the smallest embedded subsystems. For instance, the most accurate technique is LCCSD(T)-in-LDA when only two carbon atoms are included in the embedded subsystem, while it shows otherwise the poorest performance amongst the WFT-in-WFT and WFT-in-DFT methods. The convergence of the LCCSD(T)-in-LMP2 results is smooth, which is not true for the other WFT-in-WFT and WFT-in-DFT methods. The order of the latter as well as the vacuum and QM/MM schemes regarding the convergence with the size of subsystem A is the same as their order of accuracy. Just as for the DFT-in-DFT embedding, vacuum embedding and QM/MM cannot compete with the other methods for this reaction, though it is interesting to observe that vacuum embedding always outperforms QM/MM. Similar to the previous reactions, the WFT-in-WFT methods perform consistently better than LCCSD(T)-in-B3LYP, which has comparable computational costs. Moreover, LCCSD(T)-in-PBE also proves to be more accurate than the latter method, and LCCSD(T)-in-LDA has similar accuracy as the B3LYP-based embedding when a larger embedded subsystem is assembled.

4. The hydrogenation of pentacene

Fig. 8 depicts the errors obtained with the embedding techniques studied for the hydrogenation reaction of pentacene. The two-level selection scheme was used for the WFT-in-WFT and WFT-in-DFT methods using the same upper limits for the number of π orbitals as at the DFT-in-DFT benchmarks.

The results show that the most accurate technique is LCCSD(T)-in-LMP2 because this method gives by far the smallest errors among all the embedding schemes. The second best performance is observed for the LCCSD(T)-in-PBE and LCCSD(T)-in-B3LYP schemes, which is followed by the LDA, the HF+LRC, the QM/MM, and the vacuum embeddings, respectively. Even though LCCSD(T)-in-LMP2 is significantly more accurate than the others, its error (relative error) is still considerable, more than 5 kcal/mol (20%) if the carbon atoms of the first benzene ring are included in the embedded subsystem. The convergence of

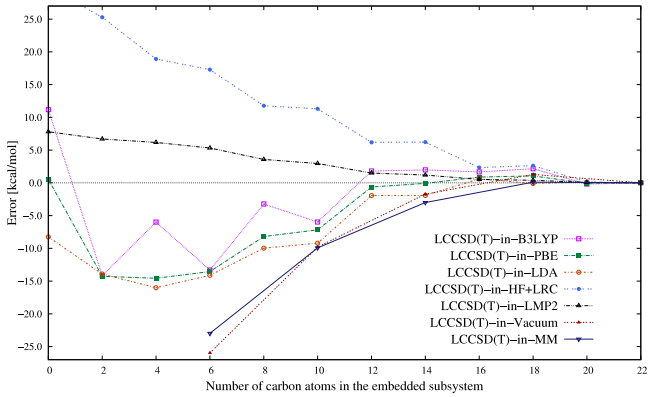

FIG. 8. Error of the reaction energy for the pentacene hydrogenation reaction with various embedding schemes as a function of the number of carbon atoms included in the embedded subsystem using the cc-pVTZ basis set. The LCCSD(T) reference reaction energy is 25.9 kcal/mol. The reaction energies of the B3LYP, PBE, LDA, HF, and LMP2 low-level methods are 37.1, 26.5, 17.7, 55.9, and 33.7 kcal/mol, respectively.

LCCSD(T)-in-LMP2 is monotonic, though it requires 14 carbons in the active subsystem to reduce the error to 1 kcal/mol, while the fluctuating LCCSD(T)-in-PBE crosses this border earlier. The LCCSD(T)-in-B3LYP errors oscillate and reach convergence more slowly than those of the aforementioned two approaches, which means that it does not offer any advantage even over the cheaper LCCSD(T)-in-PBE scheme. In general, this aromatic system is separable with the LCCSD(T)-in-LMP2 and LCCSD(T)-in-DFT methods, but they require at least three carbon rings in subsystem A for quantitative reaction energies. The vacuum, QM/MM, and the HF+LRC embeddings are only suitable for modeling aromatic systems when a very large embedded subsystem is chosen.

IV. CONCLUSIONS

Several schemes for the embedding of DFT and wave function methods into lower-level DFT or WFT approaches have been studied. The common feature of the embedding theories is that they utilize orbital localization and are exact, that is, the energy of the whole systems is recovered if the DFT or WFT method is embedded into itself. First, a simple theoretical modification of the projector-based embedding scheme of Manby and co-workers[29] is introduced. The proposed modification, the use of the Huzinaga-equation, eliminates the bias caused by the arbitrary level shift parameter of the original theory, though it has a very small effect on the numerical results. A modified algorithm is also suggested for the partitioning of the molecular orbital space, which increases the applicability of the theory. Second, straightforward approaches for the embedding into local correlation methods are proposed exploiting their intrinsic feature, that is, they split up the system by default into smaller species, which can be treated at different levels of theory.

Benchmark calculations have been carried out to assess the performance of the embedding schemes introduced in comparison to QM/MM and vacuum embedding. Our results suggest that, if reasonable active subsystems are chosen, the proposed models can be applied as black-box methods except for extended aromatic systems, where the maximum number of embedded π orbitals have to be specified. For DFT embedding, the performance of the Huzinaga-equation-based scheme seems to be superior to the other approaches, but QM/MM or even the simple vacuum embedding performs surprisingly well if the system is partitioned across single bonds. Concerning the embedding of wave function methods, the best choice is to employ a low-level local correlation method, such as local MP2 for the environment. Such embedding schemes are considerably more accurate than the corresponding WFT-in-DFT approaches, and considering that a local MP2 calculation is significantly less expensive than the preceding HF calculation, WFT-in-DFT embedding can only be competitive if a non-hybrid density functional is used.

ACKNOWLEDGMENTS

The computing time granted on the Hungarian HPC Infrastructure at NIIF Institute, Hungary, is gratefully acknowledged. The research work has been accomplished in the framework of the "BME R+D+I project," supported by Grant No. TÁMOP 4.2.1/B-09/1/KMR-2010-0002. The authors are grateful for the financial support from the Hungarian Scientific Research Fund (OTKA, Grant No. K111862).

[1]A. Warshel and M. Karplus, J. Am. Chem. Soc. **94**, 5612 (1972).
[2]A. Warshel and M. Levitt, J. Mol. Biol. **103**, 227 (1976).
[3]F. Maseras and K. Morokuma, J. Comput. Chem. **16**, 1170 (1995).
[4]H. M. Senn and W. Thiel, Angew. Chem., Int. Ed. **48**, 1198 (2009).
[5]A. S. P. Gomes and C. R. Jacob, Annu. Rep. Prog. Chem., Sect. C: Phys. Chem. **108**, 222 (2012).
[6]M. S. Gordon, D. G. Fedorov, S. R. Pruitt, and L. V. Slipchenko, Chem. Rev. **112**, 632 (2012).
[7]F. Libisch, C. Huang, and E. A. Carter, Acc. Chem. Res. **47**, 2768 (2014).
[8]T. A. Wesolowski, S. Shedge, and X. Zhou, Chem. Rev. **115**, 5891 (2015).
[9]G. Knizia and G. K.-L. Chan, J. Chem. Theory Comput. **9**, 1428 (2013).
[10]X. Assfeld and J. L. Rivail, Chem. Phys. Lett. **263**, 100 (1996).
[11]D. M. Philipp and R. A. Friesner, J. Comput. Chem. **20**, 1468 (1999).
[12]J. Pu, J. Gao, and D. G. Truhlar, J. Phys. Chem. A. **108**, 632 (2004).
[13]J. Jung, C. H. Choi, Y. Sugita, and S. Ten-no, J. Chem. Phys. **127**, 204102 (2007).
[14]J. Jung and S. Ten-no, Chem. Phys. Lett. **484**, 344 (2010).
[15]Q. Sun and G. K.-L. Chan, J. Chem. Theory Comput. **10**, 3784 (2014).
[16]G. G. Ferenczy, J. Comput. Chem. **34**, 854 (2013).
[17]B. Hégely, F. Bogár, G. G. Ferenczy, and M. Kállay, Theor. Chem. Acc. **134**, 132 (2015).
[18]S. Huzinaga and A. A. Cantu, J. Chem. Phys. **55**, 5543 (1971).
[19]P. Cortona, Phys. Rev. B **44**, 8454 (1991).
[20]T. A. Wesolowski and A. Warshel, J. Phys. Chem. **97**, 8050 (1993).
[21]J. D. Goodpaster, N. Ananth, F. R. Manby, and T. F. Miller III, J. Chem. Phys. **133**, 084103 (2010).
[22]P. Elliott, M. H. Cohen, A. Wasserman, and K. Burke, J. Chem. Theory Comput. **5**, 827 (2009).
[23]P. Elliott, K. Burke, M. H. Cohen, and A. Wasserman, Phys. Rev. A **82**, 024501 (2010).
[24]S. Fux, C. R. Jacob, J. Neugebauer, L. Visscher, and M. Reiher, J. Chem. Phys. **132**, 164101 (2010).
[25]C. Huang, M. Pavone, and E. A. Carter, J. Chem. Phys. **134**, 154110 (2011).
[26]J. D. Goodpaster, T. A. Barnes, and T. F. Miller III, J. Chem. Phys. **134**, 164108 (2011).
[27]J. Nafziger, Q. Wu, and A. Wasserman, J. Chem. Phys. **135**, 234101 (2011).
[28]L. Rajchel, P. S. Żuchowski, M. M. Szczęśniak, and G. Chałasiński, Chem. Phys. Lett. **486**, 160 (2010).
[29]F. R. Manby, M. Stella, J. D. Goodpaster, and T. F. Miller III, J. Chem. Theory Comput. **8**, 2564 (2012).

[30]J. D. Goodpaster, T. A. Barnes, F. R. Manby, and T. F. Miller III, J. Chem. Phys. **140**, 18A507 (2014).
[31]T. A. Barnes, J. D. Goodpaster, F. R. Manby, and T. F. Miller III, J. Chem. Phys. **139**, 024103 (2013).
[32]S. J. Bennie, M. Stella, T. F. Miller III, and F. R. Manby, J. Chem. Phys. **143**, 024105 (2015).
[33]L. Seijo and Z. Barandiaran, "The *ab initio* model potential method: A common strategy for effective core potential and embedded cluster calculations", in *Computational Chemistry: Reviews of Current Trends*, edited by J. Leszczynski (World Scientific, Singapore, 1999), Vol. 4.
[34]J. C. Phillips and L. Kleinman, Phys. Rev. **116**, 287 (1959).
[35]M. E. Fornace, J. Lee, K. Miyamoto, F. R. Manby, and T. F. Miller III, J. Chem. Theory Comput. **11**, 568 (2015).
[36]P. Pulay, Chem. Phys. Lett. **100**, 151 (1983).
[37]E. Kapuy, Z. Csépes, and C. Kozmutza, Int. J. Quantum Chem. **23**, 981 (1983).
[38]W. Förner, J. Ladik, P. Otto, and J. Čížek, Chem. Phys. **97**, 251 (1985).
[39]C. Hampel and H.-J. Werner, J. Chem. Phys. **104**, 6286 (1996).
[40]M. Schütz, J. Chem. Phys. **113**, 9986 (2000).
[41]C. Riplinger and F. Neese, J. Chem. Phys. **138**, 034106 (2013).
[42]H. Stoll, Phys. Rev. B. **46**, 6700 (1992).
[43]J. Friedrich and M. Dolg, J. Chem. Theory Comput. **5**, 287 (2009).
[44]S. Li, J. Ma, and Y. Jiang, J. Comput. Chem. **23**, 237 (2002).
[45]M. Ziółkowski, B. Jansík, T. Kjærgaard, and P. Jørgensen, J. Chem. Phys. **133**, 014107 (2010).
[46]N. Flocke and R. J. Bartlett, J. Chem. Phys. **121**, 10935 (2004).
[47]Z. Rolik, L. Szegedy, I. Ladjánszki, B. Ladóczki, and M. Kállay, J. Chem. Phys. **139**, 094105 (2013).
[48]S. Li, J. Shen, W. Li, and Y. Jiang, J. Chem. Phys. **125**, 074109 (2006).
[49]R. A. Mata, H.-J. Werner, and M. Schütz, J. Chem. Phys. **128**, 144106 (2008).
[50]W. Li and P. Piecuch, J. Phys. Chem. A. **114**, 6721 (2010).
[51]Z. Rolik and M. Kállay, J. Chem. Phys. **135**, 104111 (2011).
[52]J. W. Boughton and P. Pulay, J. Comput. Chem. **14**, 736 (1993).
[53]See supplementary material at http://dx.doi.org/10.1063/1.4960177 for test calculations with the self-consistent localization-based approach discussed in Sec. II A, for the comparison of the projector-based embedding scheme and its Huzinaga-equation-based variant proposed in this work, and for the specification of the orbital partitionings.
[54]M. Kállay, J. Chem. Phys. **142**, 204105 (2015).
[55]S. Li, W. Li, and J. Ma, Chin. J. Chem. **21**, 1422 (2003).
[56]W. Li, P. Piecuch, J. R. Gour, and S. Li, J. Chem. Phys. **131**, 114109 (2009).
[57]K. Rościszewski, K. Doll, B. Paulus, P. Fulde, and H. Stoll, Phys. Rev. B **57**, 14667 (1998).
[58]J. Friedrich and K. Walczak, J. Chem. Theory Comput. **9**, 408 (2013).
[59]C. Møller and M. S. Plesset, Phys. Rev. **46**, 618 (1934).
[60]S. Grimme, J. Chem. Phys. **118**, 9095 (2003).
[61]Y. Jung, R. C. Lochan, A. D. Dutoi, and M. Head-Gordon, J. Chem. Phys. **121**, 9793 (2004).
[62]H. Eshuis, J. E. Bates, and F. Furche, Theor. Chem. Acc. **131**, 1084 (2012).
[63]A. Grüneis, M. Marsman, J. Harl, L. Schimka, and G. Kresse, J. Chem. Phys. **131**, 154115 (2009).
[64]J. Gauss, in *Encyclopedia of Computational Chemistry*, edited by P. R. Schleyer, W. L. Jorgensen, H. F. Schaefer III, P. R. Schreiner, and W. Thiel (Wiley, New York, 1998), p. 615.
[65]Mrcc, a quantum chemical program suite written by M. Kállay, Z. Rolik, J. Csontos, I. Ladjánszki, L. Szegedy, B. Ladóczki, G. Samu, K. Petrov, M. Farkas, P. Nagy, D. Mester, and B. Hégely. See also Ref. 47 as well as http://www.mrcc.hu/.
[66]P. C. Hariharan and J. A. Pople, Theor. Chim. Acta **28**, 213 (1973).
[67]R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople, J. Chem. Phys. **72**, 650 (1980).
[68]T. H. Dunning, Jr., J. Chem. Phys. **90**, 1007 (1989).
[69]R. A. Kendall, T. H. Dunning, Jr., and R. J. Harrison, J. Chem. Phys. **96**, 6796 (1992).
[70]D. E. Woon and T. H. Dunning, Jr., J. Chem. Phys. **98**, 1358 (1993).
[71]F. Weigend, J. Comput. Chem. **29**, 167 (2008).
[72]F. Weigend, A. Köhn, and C. Hättig, J. Chem. Phys. **116**, 3175 (2002).
[73]P. A. M. Dirac, Proc. R. Soc. A **123**, 714 (1929).
[74]J. C. Slater, Phys. Rev. **81**, 385 (1951).
[75]J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[76]A. D. Becke, Phys. Rev. A. **38**, 3098 (1988).
[77]A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).
[78]J. Pipek and P. Mezey, J. Chem. Phys. **90**, 4916 (1989).
[79]D. A. Case, V. Babin, J. T. Berryman, R. M. Betz, Q. Cai, D. S. Cerutti, T. E. Cheatham III, T. A. Darden, R. E. Duke, H. Gohlke, A. W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T. S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K. M. Merz, F. Paesani, D. R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C. L. Simmerling, W. Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, and P. A. Kollman, AMBER 14, University of California, San Francisco, 2014.
[80]R. Salomon-Ferrer, D. A. Case, and R. C. Walker, WIREs: Comput. Mol. Sci. **3**, 198 (2013).
[81]J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, J. Comput. Chem. **25**, 1157 (2004).
[82]J. Wang, W. Wang, P. A. Kollman, and D. A. Case, J. Mol. Graphics Modell. **25**, 247 (2006).
[83]A. Jakalian, B. L. Bush, D. B. Jack, and C. I. Bayly, J. Comput. Chem. **21**, 132 (2000).
[84]A. Jakalian, D. B. Jack, and C. I. Bayly, J. Comput. Chem. **23**, 1623 (2002).