Development of analytic gradients for the Huzinaga quantum embedding method and its applications to large-scale hybrid and double hybrid DFT forces

József Csóka[a)], Bence Hégely[b)], Péter R. Nagy[c)], and Mihály Kállay[d)][1, 2, 3]

[1)] *Department of Physical Chemistry and Materials Science, Faculty of Chemical Technology and Biotechnology, Budapest University of Technology and Economics, Műegyetem rkp. 3., H-1111 Budapest, Hungary*

[2)] *HUN-REN–BME Quantum Chemistry Research Group, Műegyetem rkp. 3., H-1111 Budapest, Hungary*

[3)] *MTA–BME Lendület Quantum Chemistry Research Group, Műegyetem rkp. 3., H-1111 Budapest, Hungary*

[a)] Electronic mail: csoka.jozsef94@gmail.com

[b)] Electronic mail: hegely.bence@vbk.bme.hu

[c)] Electronic mail: nagy.peter@vbk.bme.hu

[d)] Electronic mail: kallay.mihaly@vbk.bme.hu

The theory of analytic gradients is presented for the projector-based density functional theory (DFT) embedding approach utilizing the Huzinaga-equation. The advantages of the Huzinaga-equation-based formulation are demonstrated. In particular, it is shown that the projector employed does not appear in the Lagrangian, and the potential risk of numerical problems is avoided at the evaluation of the gradients. The efficient implementation of the analytic gradient theory is presented for the approaches where hybrid DFT, second-order Møller–Plesset perturbation theory (MP2), or double hybrid DFT is embedded in lower-level DFT environments. To demonstrate the applicability of the method and to gain insight into its accuracy, it is applied to equilibrium geometry optimizations, transition state searches, and potential energy surface scans. Our results show that bond lengths and angles converge rapidly with the size of the embedded system. While providing structural parameters close to high-level quality for the embedded atoms, the embedding approach has the potential to relax the coordinates of the environment as well. Our demonstrations on a 171-atom zeolite and a 570-atom protein system show that the Huzinaga-equation-based embedding can accelerate (double) hybrid gradient computations by an order of magnitude with sufficient active regions and enables affordable force evaluations or geometry optimizations for molecules of hundreds of atoms.

I. INTRODUCTION

Modeling complex chemical systems poses a difficult challenge for electronic structure theory. Many of these systems contain hundreds if not thousands of atoms while exhibiting complicated interactions. The application of many correlation methods like coupled-cluster (CC) or second-order Møller–Plesset perturbation theory (MP2) is prohibited due to the excessive computer time that they would require. On the other side, density functional theory (DFT)[1] provides reasonable accuracy at a modest price. As a result, DFT has become one of the most prominent tools in computational chemistry. Despite its favorable properties, there are applications, for example, molecular dynamics simulations, where the cost of DFT calculations is still beyond reach, or its accuracy is not satisfactory.

A possible solution to these problems is offered by quantum embedding methods, where the system is divided into smaller subsystems, which can be treated at different levels of theory. Thus, for the chemically active subsystem, we can use a more accurate and more costly high-level method, while for the remaining environmental subsystem, a faster but probably less accurate low-level method can be applied. Accordingly, embedding methods suit the best for the situations where the chemically relevant region is spatially localized to a relatively small portion of the whole system. The usefulness of this approach has been demonstrated in many applications, the interested readers are referred to reviews on this topic.[2–7]

In quantum embedding methods, the active subsystem is typically described with a quantum mechanical (QM) method, for example, CC, MP2, or hybrid[8] or double hybrid (DH)[9] DFT. For the environment, the preferred methods are force-field molecular mechanics (MM), semi-empirical, or QM methods including mainly DFT without Hartree–Fock (HF) exchange.

Although the division of the modeled system is a common idea of all embedding methods, there are many different realizations. The most notable ones are QM/MM[4,10], density matrix embedding[11], embedded mean-field theory[12], frozen density embedding[13–15], ONIOM[16–18], subsystem DFT[19], fragmentation methods[20–25], projection based embedding (PbE)[6,26–30], and its Huzinaga-equation-based variant[31–34].

In PbE, the system is divided into two parts, an environment and an active subsystem. The environment is treated with a low-level DFT approach, while for the active part, a

high-level DFT or a wave function (WF) method is employed. The molecular orbitals (MOs) of the environment and the active part are kept orthogonal to each other. This way nonadditive interaction terms do not appear in the theory, which significantly simplifies embedding calculations. In the original formulation of PbE, the orthogonality is imposed approximately by a projector ($\mathbf{P}$) added to the Fock matrix ($\mathbf{F}$): $\mathbf{F} + \mu\mathbf{P}$, where μ is a large level shift parameter and $\mathbf{P}$ projects onto the subspace spanned by the environmental MOs. In the $\mu \to \infty$ limit, the PbE is exact, meaning that the orbitals are orthogonal to each other, and the projector's contribution to the embedding energy vanishes. However, in practice, only a finite μ can be used, which introduces numerical errors in both the energy and the orthogonality.

On the other side, Huzinaga embedding utilizes the Huzinaga equation[35] to maintain the orthogonality of the MOs. This approach is theoretically exact, that is, it does not introduce any numerical artifacts in either the energy values or the orthogonality constraints resulting in a more stable algorithm. We give a more detailed description of the method in Sect. II A.

Single point energy calculations are sufficient for particular applications. In many cases, however, the analytic gradients of the energy with respect to the nuclear coordinates are also required. Such applications are, for instance, equilibrium and transition state (TS) optimizations, reaction path searches, or molecular dynamics simulations. Analytic gradients were implemented for many of the aforementioned methods[18,36–49], and some of them (e.g., ONIOM, QM/MM, frozen density embedding, embedded mean field theory) have a relatively straightforward gradient theory. In the case of projection based embedding, the final energy is not variational with respect to all of its parameters. The general and well established way of dealing with this problem is the method of Lagrange multipliers, which was successfully used for the original formulation of PbE as well[36].

In this study, we present the theory and efficient implementation of the analytic gradients for the Huzinaga embedding method alongside illustrative applications. After describing the theoretical background, we benchmark our implementation on various test molecules and reactions. We demonstrate its efficiency and reliability with geometry optimizations, TS searches, and potential energy surface scans, and we also provide representative timing data.

II. THEORY

In this section, we present the theoretical background of our work. First, the Huzinaga embedding scheme is discussed briefly, then the gradient theory is described in more detail supposing closed-shell systems and restricted orbitals. For reference, we compiled the most important notations used in this article in Table I. The superscripts A, B, and AB refer to the active part, the environment, and the whole system, respectively. The high- and low-level self-consistent field (SCF) calculations are labeled with 1 and 2, respectively.

TABLE I: List of the most important notations used in this study

Notation	Description	Notation	Description
$\mu, \nu, \rho, \ldots$	Atomic orbitals	$C_{\mu p}$	Low-level MO coefficients
$i, j, k, \ldots$	Occupied orbitals	$C^B_{\mu p}$	Low-level environment MO coefficients
$a, b, c, \ldots$	Virtual orbitals	$C^A_{\mu p}$	Low-level active MO coefficients
$p, q, r, \ldots$	General MO	$\tilde{C}^A_{\mu p}$	High-level active MO coefficients
$D^{AB}_{\mu\nu}$	Low-level SCF density	$D^A_{\mu\nu}$	Low-level active SCF density
$D^B_{\mu\nu}$	Low-level environment SCF density	$\tilde{D}^A_{\mu\nu}$	High-level active SCF density

A. Huzinaga embedding

We assume that the low-level SCF equations have been solved for the entire system with $E_2[\mathbf{D}^{AB}]$ being the final SCF energy. In the next step, we want to replace the low-level active subsystem energy $E_2[\mathbf{D}^A]$ with the corresponding high-level subsystem energy $E_1[\tilde{\mathbf{D}}^A]$ resulting in the energy expression

$$E_{12} = E_2[\mathbf{D}^{AB}] - E_2[\mathbf{D}^A] + E_1[\tilde{\mathbf{D}}^A]. \tag{1}$$

The SCF energies E_2 and E_1 are defined as

$$E_i[\mathbf{D}] = \mathrm{Tr}\left\{\mathbf{h}_{\mathrm{core}}\mathbf{D}\right\} + \frac{1}{2}\mathrm{Tr}\left\{\mathbf{D}\mathbf{G}^i[\mathbf{D}]\right\} + E_i^{xc}, \quad (i = 1, 2) \tag{2}$$

where $\mathbf{h}_{\mathrm{core}}$ is the core Hamiltonian, E_i^{xc} is the exchange-correlation contribution in the case of DFT calculations, and $\mathbf{G}^i$ is the sum of the Coulomb and exchange matrices

$$G^i_{\mu\nu}[\mathbf{D}] = \sum_{\sigma\lambda} D_{\lambda\sigma}\left[(\mu\nu|\lambda\sigma) - \frac{1}{2}c_{\mathrm{HFx}}(\mu\lambda|\sigma\nu)\right]. \tag{3}$$

In this equation, $(\mu\nu|\lambda\sigma)$ is a four-center Coulomb integral, and c_{HFx} stands for the coefficient of the exact exchange ($c_{\mathrm{HFx}} = 1$ for HF). To improve the energy value in Eq. (1), a first-order correction is added

$$E_{\mathrm{tot}}^{\mathrm{SCF}} = E_2[\mathbf{D}^{AB}] - E_2[\mathbf{D}^{A}] + E_1[\tilde{\mathbf{D}}^{A}] + \mathrm{Tr}\left\{(\tilde{\mathbf{D}}^{A} - \mathbf{D}^{A})\frac{\partial E_{12}}{\partial \mathbf{D}^{A}}\right\}, \tag{4}$$

which is our final SCF energy. The last term describes the interaction between the subsystems, and it will be referred to as the embedding energy and denoted by E_{embed}.

To calculate the high-level MOs $\tilde{\mathbf{C}}^{A}$, and thus the $\tilde{\mathbf{D}}^{A}$ density, $E_{\mathrm{tot}}^{\mathrm{SCF}}$ is optimized with respect to the $\tilde{C}_{\mu i}^{A}$ coefficients assuming that the high-level MOs are orthogonal to each other and to the environmental MOs $\mathbf{C}^{B}$. To that end, we define the Lagrangian

$$\mathcal{L}_{\mathrm{Huzinaga}} = E_{\mathrm{tot}}^{\mathrm{SCF}} + \sum_{ij}\tilde{x}_{ij}\sum_{\mu\nu}\left[\tilde{C}_{\mu i}^{A}S_{\mu\nu}\tilde{C}_{\nu j}^{A} - \delta_{ij}\right] + \sum_{ij}y_{ij}\sum_{\mu\nu}\tilde{C}_{\mu i}^{A}S_{\mu\nu}C_{\nu j}^{B}, \tag{5}$$

where the second and third terms are the orthogonality conditions with $\mathbf{S}$ being the atomic orbital (AO) overlap matrix, and $\tilde{\mathbf{x}}$ and $\mathbf{y}$ are the Lagrange multipliers.

To determine the environmental orbitals and an initial set of high-level MOs, the low-level ones are localized. For each localized MO (LMO), we identify the atoms on which it is localized. The identification can be based on different measures. In this work, we used the Mulliken population meaning that an LMO is added to the set of high-level MOs if its Mulliken population on the active atoms exceeds a certain threshold[31]. The remaining orbitals are labeled as environmental orbitals.

The Lagrangian in Eq. (5) has to be stationary with respect to the optimal high-level MO coefficients leading to the equation

$$\frac{\partial \mathcal{L}_{\mathrm{Huzinaga}}}{\partial \tilde{C}_{\omega k}^{A}} = 4\sum_{\nu}\tilde{F}_{\omega\nu}\tilde{C}_{\nu k}^{A} + 2\sum_{i}\tilde{x}_{ik}\sum_{\nu}S_{\omega\nu}\tilde{C}_{\nu i}^{A} + \sum_{i}y_{ki}\sum_{\nu}S_{\omega\nu}C_{\nu i}^{B} = 0. \tag{6}$$

Here, we introduced the notation

$$\tilde{F}_{\mu\nu} = \frac{\partial E_{\mathrm{tot}}^{\mathrm{SCF}}}{\partial \tilde{D}_{\mu\nu}^{A}} = \frac{\partial E_1[\tilde{\mathbf{D}}^{A}]}{\partial \tilde{D}_{\mu\nu}^{A}} + \frac{\partial E_{12}[\mathbf{D}^{AB}]}{\partial D_{\mu\nu}^{A}} = F_{\mu\nu}^{1}[\tilde{\mathbf{D}}^{A}] + F_{\mu\nu}^{2}[\mathbf{D}^{AB}] - F_{\mu\nu}^{2}[\mathbf{D}^{A}], \tag{7}$$

where $\mathbf{F}^2$ denotes the low-level Kohn–Sham (KS) matrix calculated with either the $\mathbf{D}^{AB}$ or the $\mathbf{D}^{A}$ densities, while $\mathbf{F}^1[\tilde{\mathbf{D}}^{A}]$ is the KS matrix of the high-level SCF method. The last two terms in Eq. (7) define the embedding potential. Eq. (6) is called the Huzinaga equation[35], which can be solved by a regular SCF procedure with an effective Fock matrix defined by

$$\mathbf{H} = \tilde{\mathbf{F}} - \mathbf{S}\mathbf{D}^{B}\tilde{\mathbf{F}} - \tilde{\mathbf{F}}\mathbf{D}^{B}\mathbf{S}, \tag{8}$$

to which we will refer as the Huzinaga matrix in the following.

If we want to use a WF method for the active subsystem, we still solve Eq. (6) to calculate the high-level orbitals. Then, the final energy is defined as the sum of the WF correlation energy E_1^{WF} and the total SCF energy as

$$E_{\mathrm{tot}} = E_{\mathrm{tot}}^{\mathrm{SCF}} + E_1^{\mathrm{WF}}. \tag{9}$$

B. Gradient theory

In this section, we discuss the theory of the analytic gradients. We divide our analysis into two parts. First, the gradients for the DFT-in-DFT embedding are derived, while in the second part, we present the WF-in-DFT gradient theory. Although there are numerous WF methods, in this study, we opt to implement the gradients for MP2 only due to its applicability for larger active spaces, and this way we can use DH functionals as well.

1. DFT-in-DFT gradient theory

The Huzinaga embedding method is not fully variational, which can be handled with the method of Lagrange multipliers[50,51]. In the case of DFT-in-DFT embedding, we start from the Lagrangian of Eq. (5). As described in Sect. II A, this functional is stationary with respect to the high-level subsystem orbitals. However, we still have to add further terms to make it variational with respect to the low-level orbitals as well.

To enforce that the low-level MOs satisfy the low-level SCF equations, we have to add the Brillouin and the orthogonality constraints,

$$\sum_{ai} z_{ai} \sum_{\mu\nu} C_{\mu a} C_{\nu i} F^2_{\mu\nu}[\mathbf{D}^{AB}] + \sum_{pq} x_{pq} \sum_{\mu\nu} \left[C_{\mu p} S_{\mu\nu} C_{\nu q} - \delta_{pq} \right], \tag{10}$$

to our Lagrangian, where $\mathbf{z}$ and $\mathbf{x}$ are Lagrange multipliers.

For the selection of the active subspace orbitals, the low-level occupied MOs have to be localized. To incorporate this step into our Lagrangian, we have to include a localization constraint. In principle, any localization method could be employed, however, we only implemented this constraint for the Boys–Foster localization[52], which reads as

$$\sum_{i>j} z_{ij}^{\mathrm{loc}} r_{ij}, \quad r_{ij} = \sum_{\alpha=x,y,z} \left[\sum_{\mu\nu} \left(C_{\mu i} C_{\nu i} - C_{\mu j} C_{\nu j} \right) J^{\alpha}_{\mu\nu} \sum_{\mu\nu} C_{\mu i} C_{\nu i} J^{\alpha}_{\mu\nu} \right] \tag{11}$$

with $\mathbf{J}^\alpha$ being the AO dipole moment integral matrices, that is,

$$J^\alpha_{\mu\nu} = \langle \mu | \, \alpha \, | \nu \rangle \quad (\alpha = x, y, z), \tag{12}$$

and z^{loc}_{ij} is the corresponding Lagrange multiplier. Note that in the case of core-valence separation, we have to add a separate localization condition for both the core and the valence orbitals:

$$\sum_{\substack{i,j\in\mathrm{core}\\ i>j}} z^{\mathrm{loc,c}}_{ij} r_{ij} + \sum_{\substack{i,j\in\mathrm{valence}\\ i>j}} z^{\mathrm{loc,v}}_{ij} r_{ij}. \tag{13}$$

Moreover, a core-valence separation condition has to be explicitly included that reads as

$$\sum_{\substack{i\in\mathrm{core}\\ j\in\mathrm{valence}}} z^{cv}_{ij} \sum_{\mu\nu} C_{\mu i} C_{\nu j} F^2_{\mu\nu}[\mathbf{D}^{AB}]. \tag{14}$$

With all this in mind, we can conclude our final Lagrangian for DFT-in-DFT embedding:

$$\begin{aligned}\mathcal{L}_{\text{DFT-in-DFT}} = \mathcal{L}_{\text{Huzinaga}} + \sum_{ai} z_{ai} \sum_{\mu\nu} C_{\mu a} C_{\nu i} F^2_{\mu\nu}[\mathbf{D}^{AB}] + \sum_{pq} x_{pq} \sum_{\mu\nu} [C_{\mu p} S_{\mu\nu} C_{\nu q} - \delta_{pq}] + \\ \sum_{\substack{i,j\in\mathrm{core}\\ i>j}} z^{\mathrm{loc}}_{ij} r_{ij} + \sum_{\substack{i,j\in\mathrm{valence}\\ i>j}} z^{\mathrm{loc,v}}_{ij} r_{ij} + \sum_{\substack{i\in\mathrm{core}\\ j\in\mathrm{valence}}} z^{cv}_{ij} \sum_{\mu\nu} C_{\mu i} C_{\nu j} F^2_{\mu\nu}[\mathbf{D}^{AB}].\end{aligned} \tag{15}$$

The next step is to determine the Lagrange multipliers in Eq. (15). They can be calculated by using the fact that the Lagrangian has to be stationary with respect to the high- and low-level orbitals. One can easily see that by taking the derivative of the Lagrangian with respect to the high-level occupied orbitals, we arrive at the Huzinaga equation [Eq. (6)], and thus the $\tilde{\mathbf{x}}$ and $\mathbf{y}$ multipliers can be computed as

$$\tilde{x}_{kl} = -2 \sum_{\mu\nu} \tilde{C}^A_{\nu k} \tilde{C}^A_{\mu l} \tilde{F}_{\mu\nu}, \qquad y_{kl} = -4 \sum_{\mu\nu} C^B_{\mu l} \tilde{C}^A_{\nu k} \tilde{F}_{\mu\nu}. \tag{16}$$

The rest of the multipliers are determined from the condition

$$\sum_{\omega} C_{\omega p} \frac{\partial \mathcal{L}}{\partial C_{\omega q}} = 0. \tag{17}$$

Let us introduce the following notations:

$$A_{pq} = \sum_{\omega} C_{\omega p} \frac{\partial}{\partial C_{\omega q}} \left[E^{\mathrm{SCF}}_{\mathrm{tot}} + \sum_{ij} y_{ij} \sum_{\mu\nu} \tilde{C}^A_{\mu i} S_{\mu\nu} C^B_{\nu j} \right], \tag{18}$$

$$\tilde{A}_{pq}[\mathbf{z}] = \sum_{ai} z_{ai} \sum_{\omega} C_{\omega p} \frac{\partial}{\partial C_{\omega q}} \left[\sum_{\mu\nu} C_{\mu a} C_{\nu i} F^2_{\mu\nu}[\mathbf{D}^{AB}] \right], \tag{19}$$

$$\tilde{A}^{cv}_{pq}[\mathbf{z}^{cv}] = \sum_{\substack{i\in\text{core}\\ j\in\text{valence}}} z^{cv}_{ij} \sum_{\omega} C_{\omega p} \frac{\partial}{\partial C_{\omega q}} \left[\sum_{\mu\nu} C_{\mu i} C_{\nu j} F^2_{\mu\nu}[\mathbf{D}^{AB}] \right], \tag{20}$$

$$B_{pq}[\mathbf{z}^{\text{loc}}] = \sum_{ij} z^{\text{loc}}_{ij} \sum_{\omega} C_{\omega p} \frac{\partial r_{ij}}{\partial C_{\omega q}} \qquad (i > j). \tag{21}$$

To simplify the derivation, we introduce the antisymmetric part of Eq. (17):

$$\sum_{\omega} C_{\omega p} \frac{\partial \mathcal{L}}{\partial C_{\omega q}} - \sum_{\omega} C_{\omega q} \frac{\partial \mathcal{L}}{\partial C_{\omega p}} = 0. \tag{22}$$

Assuming that $\mathbf{x}$ is symmetric, it drops out of this equation. The multipliers $\mathbf{z}^{cv}$, $\mathbf{z}^{\text{loc,c}}$, and $\mathbf{z}^{\text{loc,v}}$ are calculated from the occupied-occupied part of Eq. (22). For $\mathbf{z}^{\text{loc,c}}$ and $\mathbf{z}^{\text{loc,v}}$, this equation reads as

$$B_{ij}[\mathbf{z}^{\text{loc}}] - B_{ji}[\mathbf{z}^{\text{loc}}] + A_{ij} - A_{ji} = 0 \quad (i > j), \tag{23}$$

where i and j label core orbitals for $\mathbf{z}^{\text{loc,c}}$ and valence orbitals for $\mathbf{z}^{\text{loc,v}}$. For $\mathbf{z}^{\text{cv}}$, we get a similar equation:

$$\tilde{A}^{cv}_{ij}[\mathbf{z}^{cv}] - \tilde{A}^{cv}_{ji}[\mathbf{z}^{cv}] + B_{ij}[\mathbf{z}^{\text{loc,v}}] - B_{ji}[\mathbf{z}^{\text{loc,c}}] + A_{ij} - A_{ji} = 0, \tag{24}$$

where i and j label core and valence orbitals, respectively. The multiplier of the Brillouin condition is calculated by solving the occupied-virtual part of Eq. (22):

$$\tilde{A}_{ai}[\mathbf{z}] - \tilde{A}_{ia}[\mathbf{z}] + \tilde{A}^{cv}_{ai}[\mathbf{z}^{cv}] + B_{ai}[\mathbf{z}^{\text{loc,c}}] + B_{ai}[\mathbf{z}^{\text{loc,v}}] + A_{ai} - A_{ia} = 0, \tag{25}$$

which is called the coupled-perturbed KS equation (CPKS) or coupled-perturbed HF equation (CPHF) if the low-level method is HF. The last remaining multiplier, $\mathbf{x}$, is determined from Eq. (17) by substituting all the other known multipliers.

Although all the above equations are linear, their matrices are generally too large to be stored in the memory. To circumvent this issue, we use a subspace method to solve these equations. Our approach was detailed in Ref. 53.

C. MP2-in-DFT and DH-in-DFT gradient theory

The derivatives of MP2-in-DFT and DH-in-DFT are similar, therefore, we discuss them together. To calculate the total energy, we have to add the (possibly scaled) MP2 energy to the total SCF energy, that is,

$$E_{\text{tot}} = E^{\text{SCF}}_{\text{tot}} + a_{\text{MP2}} E_{\text{MP2}}[\tilde{\mathbf{C}}^A], \tag{26}$$

where $E_{\mathrm{MP2}}[\tilde{\mathbf{C}}^A]$ is the Hylleraas functional

$$E_{\mathrm{MP2}}[\tilde{\mathbf{C}}^A] = \sum_{abij} \tilde{T}_{ab}^{ij}(ai|bj) + \sum_{pq} D_{pq}^{(2)} \tilde{F}_{pq}[\tilde{\mathbf{D}}^A]. \tag{27}$$

In this equation, $(ai|bj)$ are the four-center two-electron integrals in the MO basis, $\tilde{T}_{ab}^{ij}$ stands for the contravariant MP2 amplitudes

$$\tilde{T}_{ab}^{ij} = 2T_{ab}^{ij} - T_{ab}^{ji}, \tag{28}$$

where T_{ab}^{ij} denotes the MP2 amplitudes. The MP2 density $\mathbf{D}^{(2)}$ is defined as

$$D_{ij}^{(2)} = -2\sum_k \mathrm{Tr}\Big\{\mathbf{T}^{ik}\tilde{\mathbf{T}}^{kj}\Big\}, \quad D_{ab}^{(2)} = 2\sum_{ij}[\mathbf{T}^{ij}\tilde{\mathbf{T}}^{ji}]_{ab}, \quad D_{ia}^{(2)} = D_{ai}^{(2)} = 0. \tag{29}$$

For MP2-in-DFT, $a_{\mathrm{MP2}} = 1$, and the high-level SCF method is HF. In the case of DH-in-DFT, a_{MP2} equals the weight of the MP2 term in the DH-DFT energy expression, and the high-level SCF is a general KS calculation.

Unfortunately, the Hylleraas functional[54,55] is not variational with respect to the MOs, thus further constraints are required. Due to the fact that the MP2 energy depends on the high-level virtual orbitals, we have to include them in the orthogonality conditions of Eq. (5):

$$\sum_{pq} \tilde{x}_{pq} \sum_{\mu\nu} \Big[\tilde{C}_{\mu p}^A S_{\mu\nu} \tilde{C}_{\nu q}^A - \delta_{pq}\Big] + \sum_{pj} y_{pj} \sum_{\mu\nu} \tilde{C}_{\mu p}^A S_{\mu\nu} C_{\nu j}^B. \tag{30}$$

A Brillouin condition has to be added for the high-level orbitals as well:

$$\sum_{ai} \tilde{z}_{ai} \sum_{\mu\nu} \tilde{C}_{\mu a}^A \tilde{C}_{\nu i}^A \tilde{F}_{\mu\nu}[\tilde{\mathbf{D}}^A] \tag{31}$$

with $\tilde{z}_{ai}$ being the corresponding Lagrange multiplier. Please note that we use the $\tilde{\mathbf{F}}$ KS matrix defined by Eq. (7) instead of the Huzinaga matrix of Eq. (8). This is due to the fact that the extra terms in the Huzinaga matrix are required to enforce the orthogonality conditions, which are already included in our Lagrangian.

Finally, if the frozen core approximation is also applied in the MP2 calculation, then a new constraint has to be introduced to the Lagrangian to ensure the core-valence separation:

$$\sum_{\substack{i\in\mathrm{core}\\ j\in\mathrm{valence}}} \tilde{z}_{ij}^{cv} \sum_{\mu\nu} \tilde{C}_{\mu i}^A \tilde{C}_{\nu j}^A \tilde{F}_{\mu\nu}[\tilde{\mathbf{D}}^A]. \tag{32}$$

Despite of the formal similarity, Eqs. (32) and (14) are different constraints, and in a frozen core calculation, both have to be included in the Lagrangian. It is also important to note that the summation in Eq. (32) runs over the core and valence orbitals that are included in the active subsystem. With these modifications, the final MP2-in-DFT Lagrangian reads as

$$\mathcal{L} = E_{\mathrm{tot}} + \sum_{pq} \tilde{x}_{pq} \sum_{\mu\nu} \left[\tilde{C}^A_{\mu p} S_{\mu\nu} \tilde{C}^A_{\nu q} - \delta_{pq} \right] + \sum_{pj} y_{pj} \sum_{\mu\nu} \tilde{C}^A_{\mu p} S_{\mu\nu} C^B_{\nu j} +$$
$$\sum_{\substack{i\in\mathrm{core}\\ j\in\mathrm{valence}}} \tilde{z}^{cv}_{ij} \sum_{\mu\nu} \tilde{C}^A_{\mu i} \tilde{C}^A_{\nu j} \tilde{F}_{\mu\nu}[\tilde{\mathbf{D}}^A] + \sum_{ai} \tilde{z}_{ai} \sum_{\mu\nu} \tilde{C}^A_{\mu a} \tilde{C}^A_{\nu i} \tilde{F}_{\mu\nu}[\tilde{\mathbf{D}}^A] + \sum_{pq} x_{pq} \left[\sum_{\mu\nu} C_{\mu p} S_{\mu\nu} C_{\nu q} - \delta_{pq} \right] +$$
$$\sum_{\substack{i,j\in\mathrm{core}\\ i>j}} z^{\mathrm{loc,c}}_{ij} r_{ij} + \sum_{\substack{i,j\in\mathrm{valence}\\ i>j}} z^{\mathrm{loc,v}}_{ij} r_{ij} + \sum_{ai} z_{ai} \sum_{\mu\nu} C_{\mu a} C_{\nu i} F^2_{\mu\nu}[\mathbf{D}^{AB}] +$$
$$\sum_{\substack{i\in\mathrm{core}\\ j\in\mathrm{valence}}} z^{cv}_{ij} \sum_{\mu\nu} C_{\mu i} C_{\nu j} F^2_{\mu\nu}[\mathbf{D}^{AB}], \quad (33)$$

where E_{tot} is defined by Eq. (26).

The Lagrange multipliers are determined in two main steps. In the first step, we calculate the multipliers associated with the high-level SCF and MP2 methods, namely $\tilde{\mathbf{z}}^{cv}$, $\mathbf{y}$, $\tilde{\mathbf{x}}$, and $\tilde{\mathbf{z}}$. Next, the remaining multipliers are determined corresponding to the low-level MOs. Both steps are carried out in a similar way as described for the DFT-in-DFT embedding. For more details, see Appendix A.

Although the theory of the Huzinaga embedding gradient seems to be similar to that of the original PbE[36], there are particular differences. In PbE, the orthogonality constraints are enforced by a projector, which enters the Lagrangian through the total energy and the Brillouin condition for the high-level SCF. In our case, the projector appears in the Huzinaga matrix, which is not present in the Lagrange function. Instead, our Lagrangian contains the exact orthogonality conditions, which results in slightly more multipliers to be determined. However, they can be calculated at practically no cost [see Eq. (16)]. Meanwhile, our approach is theoretically exact and avoids the potential risk of numerical problems in the evaluation of both the energy and the gradient due to the level shift of the projector.

The handling of the core orbitals is also different. In the implementation of PbE gradient[36], the core and valence orbitals are separated neither during the localization nor in the Lagrangian despite using the frozen-core approximation for the correlation calculations. Localizing the core and valence orbitals together can introduce an unwanted mixing

of the two, while the Lagrange function loses its exactness due to omitting the constrain in Eq. (32). In our code, the core and valence orbitals are localized separately, which is included in the Lagrangian through the constraints in Eqs. (13) and (14), moreover the frozen-core approximation is also handled exactly with Eq. (32).

D. Gradient calculation

To calculate the final gradient, the Lagrange function has to be differentiated with respect to the nuclear coordinates. In order to save computer time, we have to collect similar terms and calculate their derivatives together. We discuss the gradient evaluation for the DH-in-DFT embedding scheme, the other two, DFT-in-DFT and MP2-in-DFT, are special cases of it.

By collecting the various orthogonality conditions, it is easy to see that the derivative of the AO overlap matrix $\mathbf{S}^{(q)}$ is contracted with the effective density:

$$P^S_{\mu\nu} = \sum_{pq} \tilde{x}_{pq} \tilde{C}^A_{\mu p} \tilde{C}^A_{\nu q} + \sum_{pj} y_{pj} \tilde{C}^A_{\mu p} C^B_{\nu j} + \sum_{pq} x_{pq} C_{\mu p} C_{\nu q}. \tag{34}$$

The derivatives of the localization conditions can be determined by differentiating Eq. (11). Without the core-valence separation, the $(\mathbf{J}^\alpha)^{(q)}$ dipole integral derivatives are multiplied with

$$(P^{\mathrm{loc}}_\alpha)_{\mu\nu} = \sum_{i>j} \left[z^{\mathrm{loc}}_{ij} J^\alpha_{ii} (C_{\mu i} C_{\nu i} - C_{\mu j} C_{\nu j}) + z^{\mathrm{loc}}_{ij} (J^\alpha_{ii} - J^\alpha_{jj}) C_{\mu i} C_{\nu j} \right]. \tag{35}$$

If the core and valence orbitals are localized separately, the effective density is the sum of the core and valence contributions:

$$\mathbf{P}^{\mathrm{loc}}_\alpha = \mathbf{P}^{\mathrm{loc,c}}_\alpha + \mathbf{P}^{\mathrm{loc,v}}_\alpha, \tag{36}$$

where $\mathbf{P}^{\mathrm{loc,c}}_\alpha$ and $\mathbf{P}^{\mathrm{loc,v}}_\alpha$ are formally defined as in Eq. (35), but the summation is restricted to the core or valence orbitals, respectively.

The remaining terms contain the core Hamiltonian $\mathbf{h}_{\mathrm{core}}$, four-center AO integrals, Coulomb and exchange matrices, and exchange-correlation contributions. For brevity, let us introduce the auxiliary quantities

$$\tilde{d}_{\mu\nu} = \sum_{\substack{i\in\mathrm{core} \\ j\in\mathrm{valence}}} \tilde{z}^{cv}_{ij} \tilde{C}^A_{\mu i} \tilde{C}^A_{\nu j} + \sum_{ai} \tilde{z}_{ai} \tilde{C}^A_{\mu a} \tilde{C}^A_{\nu i} + a_{\mathrm{MP2}} \sum_{pq} D^{(2)}_{pq} \tilde{C}^A_{\mu p} \tilde{C}^A_{\nu q}, \tag{37}$$

$$d_{\mu\nu} = \sum_{\substack{i\in\text{core}\\ j\in\text{valence}}} z_{ij}^{cv} C_{\mu i} C_{\nu j} + \sum_{ai} z_{ai} C_{\mu a} C_{\nu i}, \tag{38}$$

which are the effective AO matrices multiplying $\tilde{\mathbf{F}}$ and $\mathbf{F}^2$ in Eq. (33). The core Hamiltonian is present in the total SCF energy and the KS matrices $\tilde{\mathbf{F}}$ and $\mathbf{F}^2$. Therefore, the effective density contracting $\mathbf{h}_{\text{core}}^{(q)}$ is

$$\mathbf{P}^h = \mathbf{D}^B + \tilde{\mathbf{D}}^A + \tilde{\mathbf{d}} + \mathbf{d}. \tag{39}$$

The Coulomb and exchange matrices $\mathbf{G}^i[\mathbf{D}]$ defined by Eq. (3) appear in both the total SCF energy and the KS matrices. Starting with $\mathbf{G}^2[\mathbf{D}^{AB}]$, it contributes to the energies $E_2[\mathbf{D}^{AB}]$, E_{embed} [see Eqs. (2) and (4)], and the KS matrices $\tilde{\mathbf{F}}$ and $\mathbf{F}^2$. Consequently, the corresponding effective density is

$$\mathbf{P}_G^{AB} = \frac{1}{2}\mathbf{D}^{AB} + \mathbf{P}^{AB}, \quad \mathbf{P}^{AB} = (\tilde{\mathbf{D}}^A - \mathbf{D}^A) + \tilde{\mathbf{d}} + \mathbf{d}, \tag{40}$$

where $\frac{1}{2}\mathbf{D}^{AB}$ and $\tilde{\mathbf{D}}^A - \mathbf{D}^A$ come from $E_2[\mathbf{D}^{AB}]$ and E_{embed}, respectively. In a similar fashion, we can derive the effective densities for $\mathbf{G}^2[\mathbf{D}^A]$ and $\tilde{\mathbf{G}}^1[\mathbf{D}^A]$

$$\mathbf{P}_G^{A} = \frac{1}{2}\mathbf{D}^{A} + \mathbf{P}^{A}, \quad \mathbf{P}^{A} = (\tilde{\mathbf{D}}^A - \mathbf{D}^A) + \tilde{\mathbf{d}} \tag{41}$$

$$\tilde{\mathbf{P}}_G^{A} = \frac{1}{2}\tilde{\mathbf{D}}^{A} + \tilde{\mathbf{P}}^{A}, \quad \tilde{\mathbf{P}}^{A} = \tilde{\mathbf{d}}. \tag{42}$$

Let us denote the exchange-correlation matrices by $\mathbf{F}_{\text{xc}}^2[\mathbf{D}^{AB}]$, $\mathbf{F}_{\text{xc}}^2[\mathbf{D}^A]$, and $\mathbf{F}_{\text{xc}}^1[\tilde{\mathbf{D}}^A]$, where the naming convention follows that of the $\mathbf{G}^i$ matrices. They appear in the KS matrices alongside the $\mathbf{G}^i$ matrices. Therefore, their effective densities are $\mathbf{P}^{AB}$, $\mathbf{P}^A$ and $\tilde{\mathbf{P}}^{AB}$, respectively. The final terms are the derivatives of the exchange-correlation energies and the amplitude-dependent parts of the Hylleraas functional, which are evaluated as for regular DH-DFT calculations.

To sum up our results, we present the full energy gradient:

$$\begin{aligned}
\frac{\mathrm{d}E_{\text{tot}}}{\mathrm{d}q} = \mathcal{L}^{(q)} = \operatorname{Tr}\left\{\mathbf{h}_{\text{core}}^{(q)}\mathbf{P}^h\right\} + \operatorname{Tr}\left\{\mathbf{S}^{(q)}\mathbf{P}^S\right\} + \sum_{\alpha=x,y,z} \operatorname{Tr}\left\{(\mathbf{J}^\alpha)^{(q)}\mathbf{P}_\alpha^{\text{loc}}\right\} + \\
E_2^{\text{xc}}[\mathbf{D}^{AB}]^{(q)} - E_2^{\text{xc}}[\mathbf{D}^A]^{(q)} + E_1^{\text{xc}}[\tilde{\mathbf{D}}^A]^{(q)} + \\
\operatorname{Tr}\left\{\mathbf{F}_{\text{xc}}^2[\mathbf{D}^{AB}]^{(q)}\mathbf{P}^{AB}\right\} - \operatorname{Tr}\left\{\mathbf{F}_{\text{xc}}^2[\mathbf{D}^A]^{(q)}\mathbf{P}^A\right\} + \operatorname{Tr}\left\{\mathbf{F}_{\text{xc}}^1[\tilde{\mathbf{D}}^A]^{(q)}\tilde{\mathbf{P}}^A\right\} + \\
\operatorname{Tr}\left\{\mathbf{G}^2[\mathbf{D}^{AB}]^{(q)}\left(\frac{1}{2}\mathbf{D}^{AB} + \mathbf{P}^{AB}\right)\right\} - \operatorname{Tr}\left\{\mathbf{G}^2[\mathbf{D}^A]^{(q)}\left(\frac{1}{2}\mathbf{D}^{A} + \mathbf{P}^{A}\right)\right\} + \\
\operatorname{Tr}\left\{\mathbf{G}^1[\tilde{\mathbf{D}}^A]^{(q)}\left(\frac{1}{2}\tilde{\mathbf{D}}^{A} + \tilde{\mathbf{P}}^{A}\right)\right\} + 2\sum_{ijab}\tilde{T}_{ab}^{ij}(ai|bj)^{(q)}.
\end{aligned} \tag{43}$$

The most time consuming step of the evaluation of Eq. (43) is the calculation and contraction of the integral derivatives. To save computer time, in both the energy and the gradient evaluation, we used the density fitting approximation (DF)[56–60] for both SCF and MP2[61]. The three-center integral derivatives are computed four times, three of them required for the derivative of the three different $\mathbf{G}^i$ matrices and one for the last term in Eq. (43).

At this state, our program supports only MP2 as high-level wave function method. It should be straightforward to generalize the theory for other WF methods (for example CC) as well. If analytical gradients are already implemented for these WF methods, then it can be adapted for embedding calculations with minor modifications.

III. COMPUTATIONAL DETAILS

Our method has been implemented in the MRCC quantum chemistry program package[62,63] and it will be available in the next release. For details concerning the testing of the correctness of our implementation, see Appendix B. We also demonstrate the utility of our gradient implementation with geometry optimizations, TS searches, and potential energy surface (PES) scans.

In our first test set, we took the molecules from three reactions[12,31]: the substitution reaction of 1-chlorodecane with a hydroxide anion to form 1-decanol and a chloride anion, the Diels–Alder reaction of octadecanonaene and 1,3-butadiene, and the deprotonation of decanoic acid (see Fig. 1). Geometry optimizations were performed on both the reactants and the products. Using the Slater–Dirac (LDA)[64,65]; the Perdew, Burke, and Ernzerhof (PBE)[66]; the hybrid PBE (PBE0)[67]; Becke's three-parameter hybrid (B3LYP)[8,68]; and Grimme's two-parameter DH (B2PLYP)[9] functionals, three embedding schemes were tested for each molecule: PBE0-in-LDA, B2PLYP-in-PBE, and MP2-in-B3LYP. To avoid any ambiguity stemming from the integration grid, a 770-point Lebedev grid[69] with the Log3 radial quadrature[70] was employed without any pruning. Dunning's correlation consistent double-ζ (cc-pVDZ)[71,72] basis set was applied with the density fitting approximation for the SCF and the MP2 calculations as well. The auxiliary basis sets were cc-pVDZ-RI-JK[73] and cc-pVDZ-RI[74,75], respectively.

To test our method for TS optimization, we considered a Michael-addition reaction[76] where propanal and β-nitrostyrene react catalyzed by morpholine (Fig. 5). Key interme-

diates with a cyclobutane (CB) and a dihydrooxazine *N*-oxide (OO) motifs are studied alongside the TS between them, as the relative stability of the intermediates and the barrier height was found highly sensitive to the quality of the employed DFT method.[76] The structure of this TS was optimized with a variety of methods: PBE, PBE0, PBE0-in-PBE, PBE0-2-in-PBE and PBE0-DH-in-PBE, where PBE0-2[77] and PBE0-DH[78] are PBE0-based DH functionals, and the resulting geometries were compared. Weigend and Ahlrichs' triple-ζ valence polarized (def2-TZVPP)[79] AO basis set and the D3 dispersion correction[80] with Becke–Johnson damping[81] were used in all calculations. The D3 coefficients for PBE0-2 and PBE0-DH were taken from Ref. 82. The fitting basis set was def2-QZVPP-RI-JK[73] (def2-QZVPP[79] for hydrogen atoms) for all SCF and def2-TZVPP-RI[74] for MP2 calculations. Here, we used the default 590-point adaptive Lebedev grid with the Log3 quadrature.

Next, we performed a potential energy surface scan for a zeolite catalyzed methylation of propylene. Following Ref. 83, we employ an extensive cluster model of 171 atoms (Fig. 7). In the reactant state, a methyl group is bound to the acidic oxygen atom of the zeolite. Then, the methyl cation is transferred to the propylene, and finally, a hydrogen of the propylene bonds to an oxygen of the zeolite. The initial geometry of the model system is displayed in Fig. 7, while the reaction is shown in Fig. 8. For the active space, we selected the atoms of the methyl, propylene, the oxygen which the methyl is bonded to, and its first and second neighbors. In all calculations the def2-SVP[79] basis set was used with the def2-QZVPP-RI-JK[73] fitting basis set for the SCF calculations (def2-QZVPP[79] for the H atom) and def2-SVP-RI for MP2. The DFT grid was the Treutler–Ahlrichs number 3 grid[84] using the Log3 radial quadrature.

Finally, we present timing data as well. For this purpose, the zeolite model and the catechol-O-methyltransferase (COMT) enzyme[32] was selected. All timing measurements were performed on an AMD EPYC 7763 CPU using eight cores.

IV. RESULTS

A. Equilibrium geometry optimization

Figs. 2, 3, and 4 show the results for the three investigated reactions (substitution of 1-chlorodecane, Diels–Alder addition, and deprotonation of decanoic acid). The mean absolute

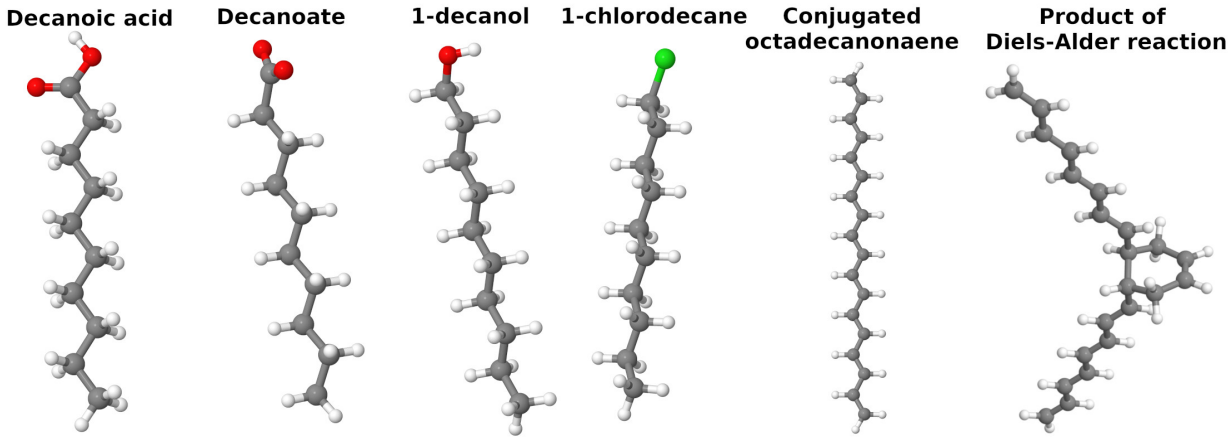

FIG. 1: Molecules used for geometry optimization.

errors (MAEs) of the bond lengths, bond angles, and dihedral angles are plotted with respect to the size of the active system on the left panel of the figures. The errors were calculated with respect to a reference geometry determined using the high-level method exclusively. A particular bond length was chosen for each system between the heavy atoms of the smallest active space, and its error was also plotted on the top right panels of the figures. Additionally, the reaction energies were also determined and plotted on the bottom right panels. The first data point on each subplot (i.e., zero carbon atom) shows the results calculated with the low-level method only.

Starting with the geometry errors, it is apparent that they decrease with increasing active subsystem size in almost all cases, which is the expected behavior. The mean errors in the bond lengths are below or around 0.04 bohr for all molecules and subsystem sizes, which is an acceptable difference. The low-level methods give larger errors than the corresponding embedding schemes with small active spaces in almost all cases. The only exception is the octadecanonaene with MP2-in-B3LYP.

The errors in the bond angles are typically below 0.3° and show a decaying trend. The exceptions are the octadecanonaene and the decanoate molecules with MP2-in-B3LYP. The errors for octadecanonaene start slightly higher, around 0.4°, and decrease rather slowly with increasing subsystem size. The error gets smaller at every second configuration, where the border of the active region cuts a single C–C bond. On the other hand, the accuracy does not increase when a double C–C bond is cut.

In the case of decanoate (Fig. 3), a jump in the error of the MP2-in-B3LYP bond and

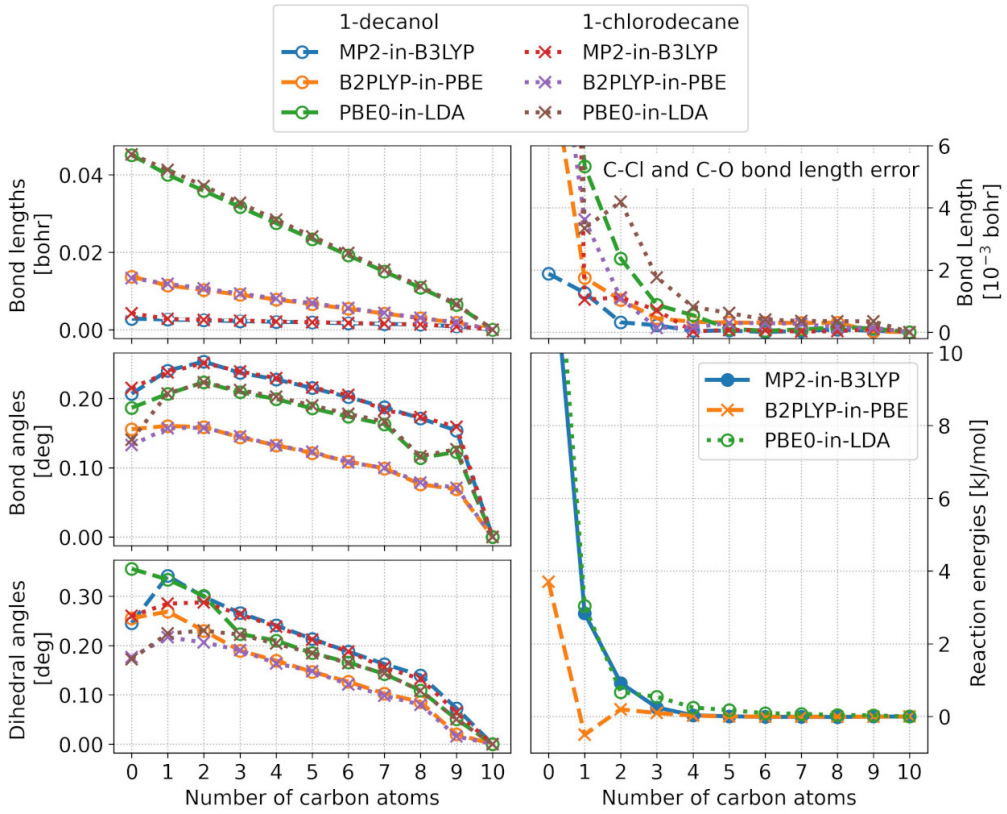

FIG. 2: MAEs for the optimized geometry parameters of 1-chlorodecane and 1-decanol and the reaction energies of the substitution reaction. The size of the embedded subsystem was systematically increased by one carbon atom. The embedded subsystem of one carbon atom contains also the hydroxyl group or the chlorine atom.

dihedral angles can be observed with two carbon atoms in the active subspace. This calculation predicts a symmetric structure where the carboxylate group is within the plane of the carbon chain instead of the correct, tilted geometry. Although it could be an artifact of the optimizer due to the flat potential energy surface, it turns out that the symmetric configuration is the energetically optimal one with this subspace selection. The optimizer predicts the same structure even if the maximal step size is set to a small value (0.05 atomic units). This indicates that small subsystems (i.e., 1–2 bonds away from the chemical change) cannot always describe the important part of a molecule, and thus, their general application is not recommended without thorough benchmarking.

For bond angles, the low-level method performs similarly as the embedding calculations with small active spaces, except for 1-decanol and 1-chlorodecane. In these cases, we need medium-sized active regions to get errors below those of the low-level methods. This indi-

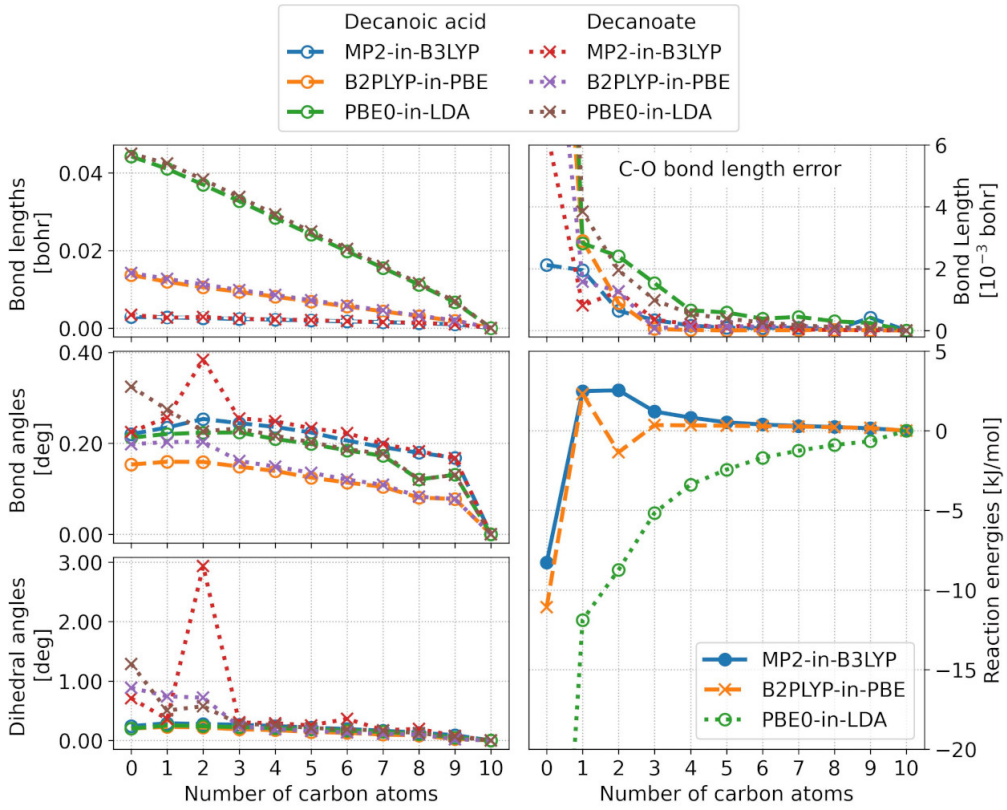

FIG. 3: MAEs for the optimized geometry parameters of decanoic acid and decanoate and the reaction energies of the deprotonation reaction. The size of the embedded subsystem was systematically increased by one carbon atom. The embedded subsystem of one carbon atom contains also the two oxygen atoms or the oxygen atom plus the hydroxyl group.

cates that there are relatively strong interactions between the subsystems, thus more active atoms are required to be able to relax the geometry of the environment as well.

In spite of these exceptions, our general observation is that the errors of the bond angles decay quickly and converge to the reference values systematically. In most cases, the errors remain controlled even for small active spaces, however, some systems with too compact active regions may exhibit larger errors or unusual behavior.

Regarding the dihedral angles, we get a relatively small mean error of less than 0.3° in most cases, and they converge to zero systematically with increasing subsystem size. Larger errors can be seen for the product of the Diels–Alder reaction (Fig. 4) for small active spaces, especially with the MP2-in-B3LYP embedding scheme, where the MAE exceeds 1.2°. This error drops below reasonable values quickly as we expand the subsystem size, and a fast convergence to the reference structure can be observed. As in the case of bond angles, a

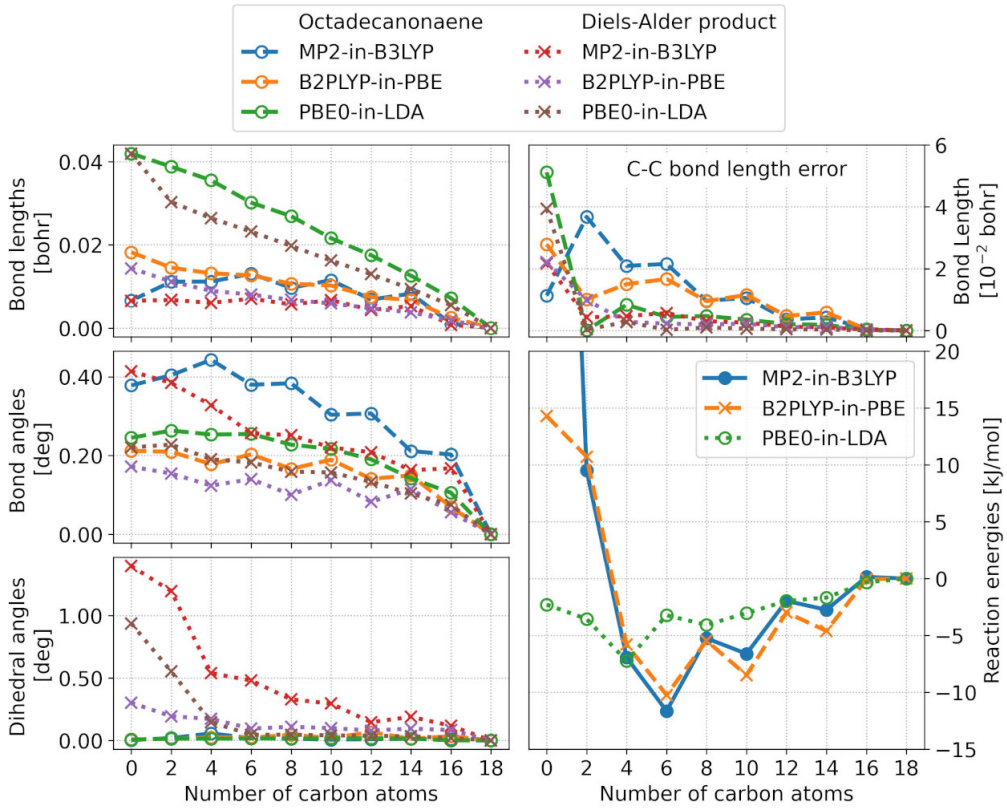

FIG. 4: MAEs for the optimized geometries of the reactant and product of the Diels–Alder reaction and the corresponding reaction energies. The embedded subsystem of two carbon atoms includes the two carbons in the middle of the molecules and it is extended symmetrically in both directions.

huge jump can be seen for decanoate with MP2-in-B3LYP which can be explained as before.

The embedding methods perform better than the low-level one in the case of dihedral angles. The only exceptions are the 1-decanol and 1-chlorodecane molecules due to similar reasons as for the bond angles.

To sum up, the errors in the geometry are moderate in most cases even for small active spaces, and there is a decaying trend in the mean errors in all three types of geometrical parameters. This means that the high-level method applied to a small active subspace has the potential to relax the coordinates of the environment as well resulting in a more accurate structure for the whole molecule and not just for the active part.

Looking at the errors of the bond lengths in the active space (top right panel in Figs. 2–4), similar observations can be made as before. For most subsystem sizes, the errors are small, typically below $5 \cdot 10^{-3}$ bohr. The only outlier is octadecanonaene, where errors up

to $2 \cdot 10^{-2}$ bohr are measured with the smaller active regions. Nonetheless, a relatively fast convergence can be seen for all cases and accurate results can be achieved with small to medium-sized active regions. Importantly, the bond lengths in the embedded region are about an order of magnitude more accurate than for the entire molecule (cf. the top left and right panels of Figs. 2–4). Moreover, the embedding calculations give smaller errors than the low-level method in almost all cases. There is only one exception, octadecanonaene with MP2-in-B3LYP, which is expected based on the top left plot of Fig. 4.

The reaction energies show similar convergence properties as the geometric coordinates. In most cases, a fast and systematic convergence is measured, accurate result are achieved with medium-sized active spaces and even small active regions can outperform the low-level methods. Moreover, when σ-bonds are cut, 2–3 active carbon atoms are sufficient for sub-kJ/mol accuracy. The only out of ordinary case is the Diels–Alder addition with the methods MP2-in-B3LYP and B2PLYP-in-PBE. Although convergence is achieved, it is slow and less systematic than in other cases. Interestingly, PBE0-in-LDA exhibits much faster and smoother convergence properties. These results indicate that modeling conjugated systems with embedding methods requires extra attention, not just in the selection of the active space but also at pairing of the high- and low-level methods. The other exception is the deprotonation of the decanoic acid, where PBE0-in-LDA gives relatively large errors at small active regions. This can be explained with the large error of the LDA functional, which is corrected by PBE0, but several active carbon atoms are required to fully mitigate this effect. Nonetheless, this error gets smaller with expanding subspace, and chemical accuracy can be achieved with medium-sized active regions including 3–4 bonds away from the deprotonated oxygen.

B. Transition state optimization

Next, we tested our embedding gradient program by optimizing the TS for an organic reaction containing 40-atom species. The reaction and the chosen active space are depicted in Fig. 5. The PBE, PBE0, PBE0-in-PBE, PBE0-2-in-PBE and PBE0-DH-in-PBE models were employed, and the resulting geometries were compared with the ones optimized with PBE0-DH and PBE0-2.

The deviations of the geometrical parameters (bond lengths, bond angles, dihedral angles)

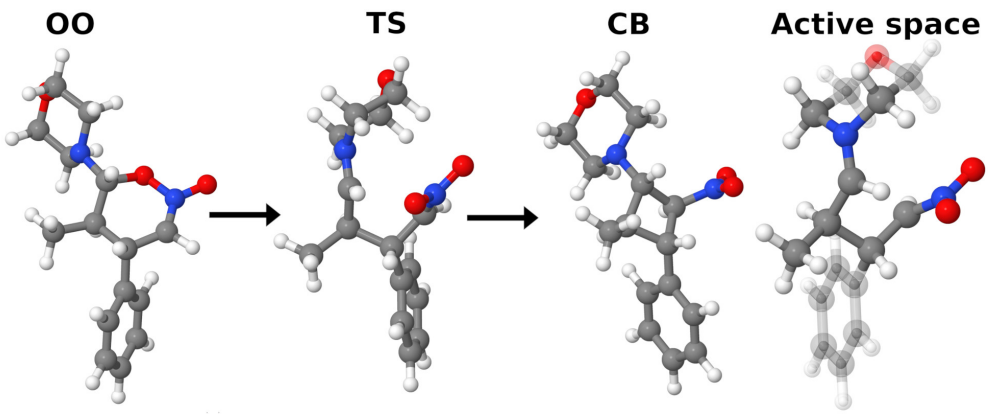

FIG. 5: The reaction studied for the TS optimization. The rightmost picture shows the active space: the atoms with a solid color are the active ones, while the transparent atoms belong to the environment.

were calculated and plotted in histograms (Fig. 6). The latter are made from 20 bins between 0 and the maximal error of the given coordinate type among all functionals. On the y axis, we plotted the normalized frequency of errors,

$$y_i = \frac{n_i}{\sum_i n_i},$$

where i runs through the bins, and n_i denotes the number of points in the ith bin.

It is easy to see from the data that the maximum error of the bond lengths is below 0.05 bohr with all functionals. This is acceptable from a practical point of view, especially since the active region covers only the atoms within 2 bonds of the breaking C–O bond. The differences are much more pronounced for the bond and dihedral angles. With PBE, we get relatively large errors with maxima around 2° and 8° in the bond and dihedral angles, respectively. The spread of the errors is slightly larger in the case of the PBE0-2 reference. With PBE0, the errors drop significantly, the largest differences for bond and dihedral angles are below 1° and 3°, respectively. As for PBE, we get less accurate agreement with the PBE0-2 reference structure.

Using embedding methods, we can clearly improve the results calculated with PBE. For PBE0-in-PBE the spread of the errors is just slightly larger than for PBE0 and it is more visible for the bond lengths and bond angles. Regarding the DH functionals, we get similar results as for PBE0-in-PBE. In the case of the PBE0-DH reference, the best accuracy is achieved by PBE0-DH-in-PBE, although its advantage is marginal over the other methods.

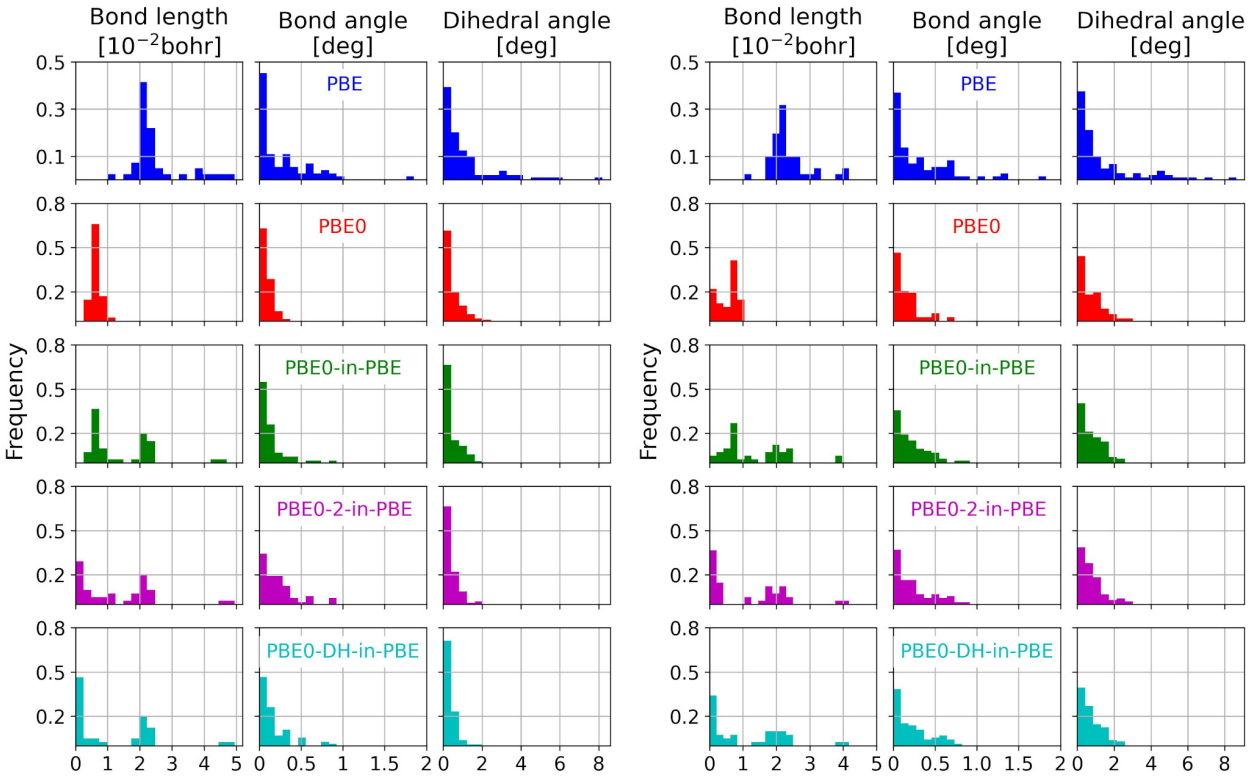

FIG. 6: TS structure errors compared to PBE0-DH (left) and PBE0-2 (right) references. Each row shows the error of the three different geometrical parameters for a given computational method.

On the other side, PBE0-2-in-PBE predicts less accurate dihedral angles than the other embedding schemes, however, the differences are small.

Nonetheless, these observations show that treating just a small part of the molecule at a higher level of theory can lead to substantial improvements in molecular structures. This agrees with our conclusions drawn for equilibrium geometries that the high-level method applied to the active space can relax the environmental coordinates as well. Most probably this effect gets weaker with increasing distance from the active region, however, we could demonstrate it for smaller molecules containing a couple of dozens of atoms.

The geometry errors were calculated for the active subspace as well (Tables II and III). To that end, all internal coordinates within the active region were collected, and the maximum and mean absolute differences were determined compared to the PBE0-DH and PBE0-2 reference structures. The largest errors are produced by PBE as expected based on Fig. 6. The rest of the methods performed similarly, larger differences are observed only for the

TABLE II: Geometry errors of the TS for the active region compared to the PBE0-DH reference. The error of the bond length is given in bohr, the errors of the bond and dihedral angles are measured in degrees.

	Bond Length		Bond angle		Dihedral angle	
Method	Max	Mean	Max	Mean	Max	Mean
PBE	0.050	0.025	1.85	0.35	8.19	1.96
PBE0	0.009	0.006	0.34	0.10	2.17	0.59
PBE0-in-PBE	0.010	0.006	0.30	0.09	1.74	0.52
PBE0-2-in-PBE	0.010	0.003	0.51	0.17	1.40	0.42
PBE0-DH-in-PBE	0.003	0.001	0.29	0.09	0.84	0.31

dihedral angles. PBE0-in-PBE yields fairly good results seemingly outperforming PBE0 in most cases. However, it is true only for the active part of the molecule, the overall performance of PBE0 is better as it was discussed above.

The DH functionals give similar results, and we get higher accuracy for both the PBE0-2 and the PBE0-DH references with the corresponding DH-in-PBE embedding scheme. Generally speaking, the lowest errors are achieved with these embedding methods, but PBE0-in-PBE yields similar or better errors for the bond angles against the PBE0-DH reference and has the lowest maximal error in the dihedral angles for the PBE0-2 reference. Comparing the bond length errors only for the active atoms to those for the entire molecule (Fig. 6 vs. Tables II and III), these structural parameters are again an order of magnitude more accurate for the active region.

The barrier heights were also calculated for this reaction and they are shown in Table IV. The various levels of DFT predict significantly different barrier heights, the difference is over 20 kJ/mol between PBE and the DH functionals. The underestimated GGA barrier and the improved hybrid and DH results are consistent with the trends obtained with the local natural orbital CCSD(T)[85,86] reference in Ref. 76. Embedding methods can improve upon the pure DFT barrier, and the resulting barriers are close to that of the high-level methods. For the DH functionals, the values predicted with the embedding schemes are within 1 kJ/mol with respect to the parent DH method, while PBE0-in-PBE is even more accurate giving an error less than 0.1 kJ/mol compared to PBE0. These errors support our

TABLE III: Geometry errors of the TS in the active space compared to PBE0-2 reference. The error of the bond lengths is given in bohr, the errors of the bond and dihedral angles are measured in degrees.

	Bond length		Bond angle		Dihedral angle	
Method	Max	Mean	Max	Mean	Max	Mean
PBE	0.042	0.024	1.85	0.43	8.64	2.47
PBE0	0.009	0.006	0.70	0.22	2.70	1.22
PBE0-in-PBE	0.009	0.006	0.77	0.20	1.50	1.17
PBE0-2-in-PBE	0.004	0.002	0.52	0.15	2.83	0.79
PBE0-DH-in-PBE	0.017	0.003	0.64	0.16	2.26	0.95

previous findings that embedding calculations supply results similar to those of the high-level methods provided that the active region was selected carefully. Importantly, here, an active region of only 11 heavy atoms could provide sub-kJ/mol accuracy.

TABLE IV: Barrier heights (in kJ/mol) for the Michael-addition reaction of Fig. 5.

Method	Barrier height	Method	Barrier height
PBE	99.3		
PBE0	116.0	PBE0-in-PBE	116.1
PBE0-2	125.7	PBE0-2-in-PBE	126.6
PBE0-DH	124.5	PBE0-DH-in-PBE	125.2

C. Potential energy surface scan

To further demonstrate the possible applications of our embedding gradient implementation, we studied a zeolite catalyzed methylation reaction of propylene. The structure of the zeolite model and the reaction are shown in Figs. 7 and 8. It is known in the literature that various DFT functionals predict significantly different barrier heights[83]. Our goal is to show that embedding methods can provide accurate barriers without the large cost of performing (double) hybrid structure optimizations for the entire cluster. To that end, we recorded a

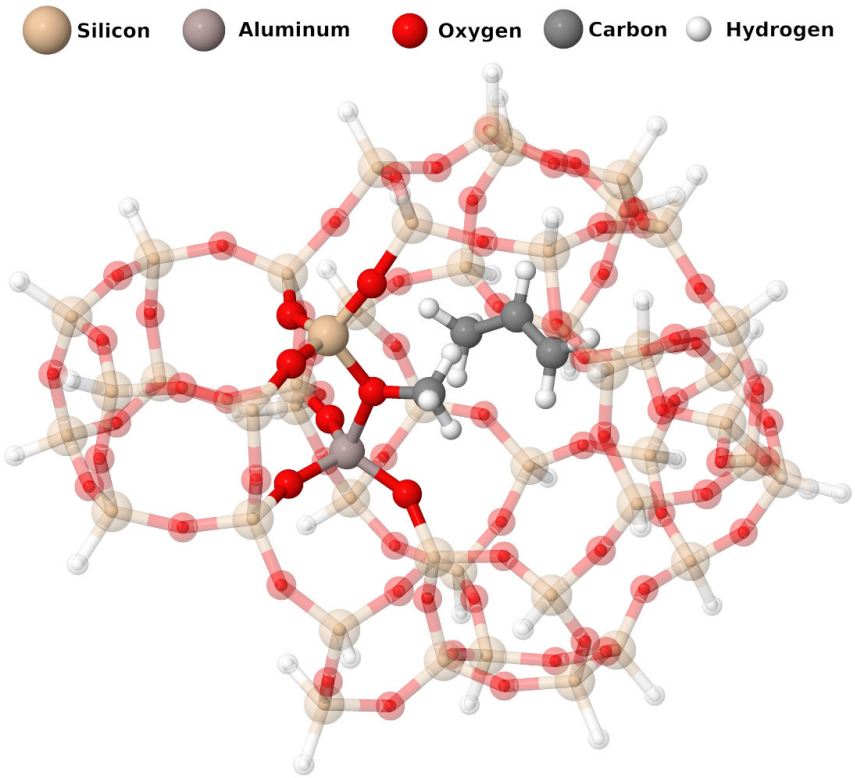

FIG. 7: The structure of the zeolite model. The transparent atoms constitute the environment, and the atoms with a solid color are added to the active space.

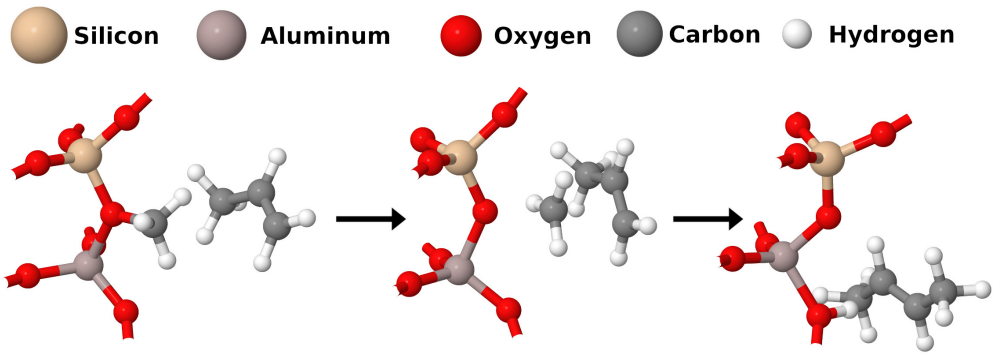

FIG. 8: The methylation of propylene inside the zeolite model. The pictures from left to right show the reactant, the TS, and the product configurations, respectively. Only the atoms of the active space are plotted.

potential energy surface around the transition state. The TS structure from Ref. 83 was used as a starting point, and the PES was scanned towards both the reactant and the product.

The geometries were relaxed only with PBE and PBE0-in-PBE, and then the PBE0-2-in-PBE and PBE0-DH-in-PBE energies were evaluated at the geometries optimized with

PBE0-in-PBE. The active space contained the atoms of the propylene, the methyl group of the zeolite, the oxygen to which the methyl is bonded, and the first and second neighbors of this oxygen. The resulting energy surfaces are plotted in Fig. 9.

The PBE and PBE0-in-PBE surfaces start to separate from each other already close to the reactant state, and the gap widens until the TS is reached. The estimated difference in the barrier heights from the surface scan is slightly above 20 kJ/mol, which is expected based on the previous results[83]. Passing the TS, the energy surfaces come closer to each other, and the differences become more modest.

For the DH functionals, we did not relax the geometries but used the ones calculated with PBE0-in-PBE. The PES curve of PBE0-DH-in-PBE is closer to PBE0-in-PBE with an estimated barrier height difference of 10 kJ/mol. The PBE0-2-in-PBE result deviates a lot more, the energy barrier is almost 20 kJ/mol larger than for PBE0-in-PBE. Relaxing the geometries would result in smaller estimated barrier energies. Based on the observations of Sect. IV B, this effect would be more pronounced for PBE0-2-in-PBE. As we have seen, the structure errors of PBE0-2-in-PBE were larger compared to PBE0-DH-in-PBE or even PBE0-in-PBE indicating that the geometries predicted with PBE0-2-in-PBE might differ more significantly from the ones calculated with PBE0-in-PBE.

D. Timings

We measured timing data for the zeolite and the COMT systems. Calculations were carried out with the def2-SVP and def2-TZVPP basis sets for the zeolite and with def2-SVP for the COMT system. def2-SVP was paired with the cc-pVDZ-RI-JK fitting basis set for SCF and def2-SVP-RI for MP2. In the case of def2-TZVPP, the def2-QZVPP-RI-JK and def2-TZVPP-RI fitting basis sets were used for the SCF and MP2 calculations, respectively. The timings are presented in Tables V, VI, and VII.

Starting with the zeolite system, it contains 171 atoms with 1292 electrons in total, of which 612 are core electrons. The number of AO basis functions for def2-SVP is 2137, whereas the SCF and correlation fitting basis sets contain 11599 and 7186 functions, respectively. With def2-TZVPP there are 4974 AOs, the number of auxiliary functions is 12614 for the SCF and 12904 for MP2.

Looking at the data in Table V, it is obvious that, as expected, PBE is significantly

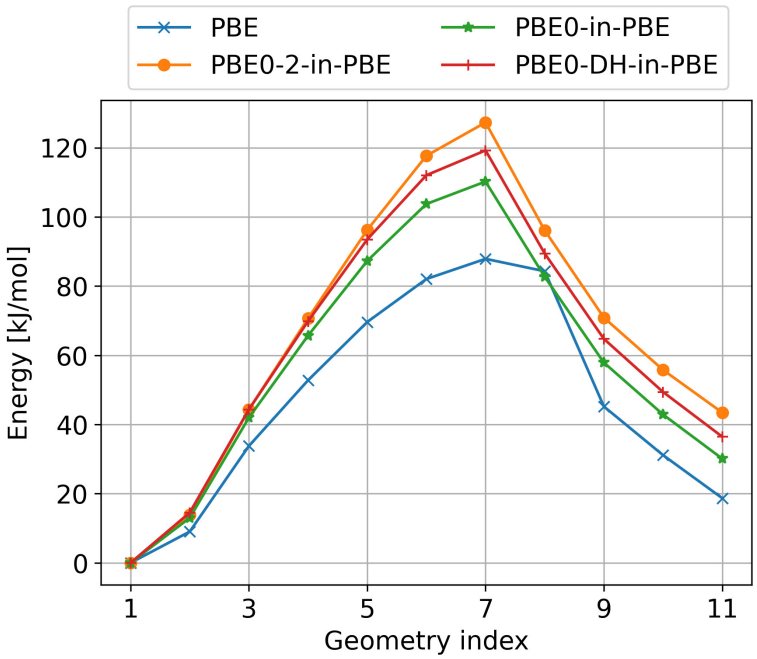

FIG. 9: The PES of the methylation reaction. The energy of the reactant was chosen to be zero for all methods. The DH-in-PBE surfaces were calculated using the geometries optimized with PBE0-in-PBE.

TABLE V: Timing data for the zeolite system using the def2-SVP basis. All timings are measured in minutes. The numbers in parenthesis show how much the given step contributes to the total runtime in percent.

Low-level method	-	-	-	PBE	PBE	PBE
High-level method	PBE	PBE0	PBE0-2	PBE0	PBE0-2	PBE0-DH
Low-level SCF	-	-	-	16.1 (16.7)	21.1 (6.9)	20.6 (7.5)
High-level SCF	18.1 (88.4)	733.3 (90.5)	806.0 (8.0)	55.2 (57.5)	81.5 (26.5)	63.4 (23.2)
MP2	-	-	7331.8 (72.8)	-	35.6 (11.6)	36.5 (13.3)
Total energy	18.1 (88.4)	733.3 (90.5)	8137.9 (80.8)	71.3 (74.2)	138.2 (44.9)	120.5 (44.0)
Multipliers	-	-	1412.9 (14.0)	11.8 (12.3)	101.6 (33.0)	88.0 (32.2)
Gradient	2.4 (11.6)	77.1 (9.5)	520.7 (5.2)	13.0 (13.5)	68.0 (22.1)	65.3 (23.8)
Total gradient	2.4 (11.6)	77.1 (9.5)	1933.6 (19.2)	24.8 (25.8)	169.6 (55.1)	153.4 (56.0)
Energy+gradient	20.5	810.3	10071.5	96.0	307.9	273.9

faster than any other method. It takes 50 times more time to calculate a PBE0 energy and gradient. It is also apparent that the gradient calculation itself takes only a small fraction of the total runtime for both PBE and PBE0.

With the embedding schemes, we can achieve convincing improvements in computer timings. It takes eight times less to calculate the energy and gradient for PBE0-in-PBE than for PBE0. We could measure around 10- and 3-times speedups for the energy and gradient evaluations, respectively.

The difference is more pronounced for the DH functionals. A PBE0-2 gradient calculation was also carried out for this system. The SCF iteration took around 13 hours, while the MP2 energy evaluation lasted 5 days and required 1 TB of disk space for the MP2 amplitudes. After this, the evaluation of the gradient was around 32 hours, thus, the whole calculation took just a few hours less than one week. Compared to this, the energy evaluation for the embedding methods PBE0-2-in-PBE and PBE0-DH-in-PBE required around 2 hours, which corresponds to a more than 60-fold speedup. The gradient calculation took approximately 2.5-3 hours, which is an order of magnitude faster than for PBE0-2. Although the total energy and gradient calculation still takes around 5 hours, these methods are already feasible options for geometry optimization.

For PBE and PBE0, the large portion of the total runtime (85-90%) is spent on the energy calculation. In the case of PBE0-2 the energy evaluation took 80% of the total runtime due to the large amount of data moved between the main memory and the disk. The relative runtime of the energy calculation drops to around 70% for PBE0-in-PBE due to the additional time needed for the determination of the multipliers. This gap shrinks further for the embedding schemes using DH functionals, where the gradient and energy calculations require around the same time, however, the gradient evaluation becomes the slightly more dominant step. This can be explained with the second, high-level CPKS equation which has to be solved. Moreover, the gradient of the last term in Eq. (43), the contribution of the amplitudes, also requires a significant amount of time.

By careful inspection of Table V, one can notice that there are different timing data for the low-level PBE calculations. This is the result of the different iteration count due to the tighter SCF convergence thresholds used if post-SCF correlation calculations are carried out.

The timings calculated with the def2-TZVPP basis set are shown in Table VI. For PBE,

TABLE VI: Timing data for the zeolite system using the def2-TZVPP basis. For details, see the caption of Table V.

Low-level method	-	-	PBE	PBE	PBE
High-level method	PBE	PBE0	PBE0	PBE02	PBE0-DH
Low-level SCF	-	-	52.4 (14.3)	59.7 (3.2)	50.63 (2.3)
High-level SCF	49.4 (85.3)	2072.2 (90.0)	228.1 (62.2)	259.8 (14.0)	281.9 (12.6)
MP2	-	-	-	818.6 (44.1)	1043.8 (46.7)
Total energy	49.4 (85.3)	2072.2 (90.0)	280.6 (76.5)	1138.2 (61.3)	1376.4 (61.6)
Multipliers	-	-	37.0 (10.1)	334.6 (18.0)	402.1 (18.0)
Gradient	8.5 (14.7)	229.1 (10.0)	49.2 (13.4)	385.0 (20.7)	456.2 (20.4)
Total gradient	8.5 (14.7)	229.1 (10.0)	86.163 (23.5)	719.6 (38.7)	858.3 (38.4)
Total Energy+Gradient	57.9	2301.4	366.729	1857.8	2234.7

PBE0, and PBE0-in-PBE, the particular steps contribute to the total runtime in similar proportions as for the def2-SVP basis set. The total runtime of PBE0-in-PBE is around 6 hours, which is a 6-fold speedup compared to PBE0. In the case of the DH functionals, the MP2 energy evaluation becomes the most time consuming step due to the increased memory requirements resulting in expensive data movements between the memory and the disk. These calculations took over 30 hours, which can be limiting in applications like geometry optimizations or TS searches. We note that typical organic molecules of the same size contain fewer electrons and thus fewer AOs. As a result, similar calculations would require less computer time.

We measured timing data for the COMT system as well. The whole protein contains 3420 atoms, of which 571 QM atoms were selected and surrounded by an MM environment. For the embedding calculations, two different active spaces were used, one with 50 atoms and another one with 64 atoms, denoted by COMT-1 and COMT-2, respectively. The different regions are shown in Fig. 10. We only tested the PBE and PBE0-in-PBE methods in this case, the measured data are collected in Table VII.

The majority of the computer time (over 80%) is spent on the solution of the high-level SCF equations. Compared to this, the low-level SCF requires only a small fraction (7–9%) of the computation time, and even the gradients can be evaluated relatively quickly (10%)

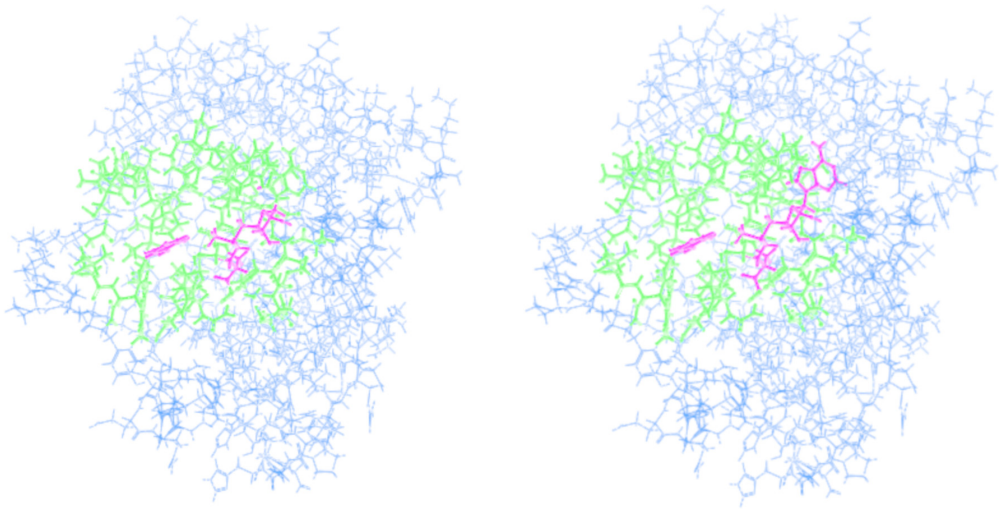

FIG. 10: The COMT-1 (left) and COMT-2 (right) test systems. The outer blue region was described with MM, the green colored part is the QM environment, while the QM active atoms are shown in magenta.

as well. The calculation of the Lagrange multipliers takes around a quarter of the gradient evaluation and just a few percent of the total runtime.

Although we did not perform PBE0 calculations for the full QM region, we can give an approximate timing. There are in total 1106 MOs in the QM region, and only 105 and 139 MOs were selected as high-level ones for the COMT-1 and COMT-2 systems, respectively. The computer time is dominated by the energy calculation, more precisely, by the calculation of the exchange matrix, which scales linearly with the number of MOs. Thus, the estimated runtimes are around 11 and 13 days for the COMT-1 and COMT-2 systems, respectively, which translate to speedups of around 7-10 times.

The entire calculation for these systems took 1-2 days, which is relatively lengthy for geometry optimization but suitable for the calculation of first-order properties. We note that our SCF implementation can handle multiple Message Passing Interface (MPI) threads. Running the SCF program in MPI parallel mode could improve the total runtime of our calculations. On the other side, our gradient code is not MPI parallel yet, therefore, the proportion of the gradient evaluation would most probably flip over in an MPI parallel run and it would require an overwhelmingly large part of the total runtime. This test shows clearly that increasing the scalability of our gradient code with MPI could extend the applicability of this implementation foreshadowing possible further developments.

TABLE VII: Timing data for the COMT systems using the PBE0-in-PBE approach. For details see the caption of Table V.

System	COMT-1	COMT-2
Low-level SCF	175.7 (8.9)	196.9 (6.7)
High-level SCF	1474.9 (74.8)	2298.4 (78.7)
Total energy	1644.8 (83.7)	2495.3 (85.5)
Multipliers	97.1 (4.9)	110.2 (3.8)
Gradient	224.2 (11.4)	313.3 (10.7)
Total gradient	321.2 (16.3)	423.5 (14.5)
Energy+gradient	1971.8	2918.7

V. CONCLUSIONS AND OUTLOOK

In this article, we have presented the theory, efficient implementation, and demonstrative applications of analytic gradients for the Huzinaga-equation-based embedding method. Several test calculations have been performed on various molecular systems, and the method has been applied to equilibrium geometry optimizations, TS searches, and potential energy surface scans.

We have demonstrated that in geometry and TS optimizations, the present embedding method can relax not just the structure of the active subspace but that of the environment as well, thus, increasing the accuracy of the low-level method for the environment. The calculated structures converge to the ones determined with the high-level methods by systematically expanding the active subsystem. With PES scans for the zeolite-catalyzed methylation reaction, we could demonstrate that treating just a small fraction (ca. 10%) of the whole system on a higher level of theory can drastically change the calculated properties, in this case, the reaction barrier height.

We have also presented timing data. It has been shown that with small active subsystems, significant, up to an order of magnitude, speedups can be achieved compared to the high-level methods. In the case of DFT-in-DFT embedding, the energy evaluation requires the majority of the computer time, while for DH-in-DFT, the gradient calculation is the more time consuming step. The presented developments accelerate (double) hybrid DFT gradient

computations for systems of up to a few hundred QM and ca. 20–50 embedded atoms to a widely accessible time already with 8 cores of a single processor.

At this point, our gradient code is shared-memory parallel only, and it can use around a dozen of processor cores efficiently. Further developments should focus on increasing scalability by implementing a hybrid distributed- and shared-memory parallel version. Other possible ways to broaden the applicability of our approach is the implementation of AO basis set truncation[29], the use of local fitting SCF methods[53,87], and the implementation of WF-in-DFT gradients for other WF methods as well.

ACKNOWLEDGMENTS

The authors are grateful for the financial support from the National Research, Development, and Innovation Office (NKFIH, Grant No. KKP126451 and FK142489), the ERC Starting Grant No. 101076972, "aCCuracy", the János Bolyai Research Scholarship of the Hungarian Academy of Sciences, the ÚNKP-23-5-BME-408 New National Excellence Program of the Ministry for Culture and Innovation sourced from the NKFIH fund, and the computing time granted by the Hungarian Governmental Information-Technology Development Agency on the Komondor HPC. The research reported in this paper is part of project BME-EGA-02, implemented with the support provided by the Ministry of Innovation and Technology of Hungary from the National Research, Development and Innovation Fund, financed under the TKP2021 funding scheme.

AUTHOR DECLARATIONS

Author declarations

The authors have no conflicts to disclose.

Author contributions

József Csóka: Methodology (equal); Software (lead); Data curation (equal); Writing – original draft (lead); Writing – review & editing (equal). Bence Hégely: Data curation (equal); Writing – review & editing (equal). Péter R. Nagy: Funding acquisition (equal);

Methodology (equal); Writing – review & editing (equal). Mihály Kállay: Conceptualization (lead); Funding acquisition (equal); Methodology (equal); Supervision (lead); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available within the article.

Appendix A: Calculation of Lagrange multipliers for DH-in-DFT and MP2-in-DFT embedding

Introducing the variables

$$A^A_{pq} = a_{\mathrm{MP2}} \sum_{\omega} \tilde{C}^A_{\omega p} \frac{\partial E_{\mathrm{tot}}}{\partial \tilde{C}^A_{\omega q}} \tag{A1}$$

$$\tilde{A}^{cv,A}_{pq}[\tilde{\mathbf{z}}^{cv}] = \sum_{\substack{i\in\mathrm{core}\\ j\in\mathrm{valence}}} \tilde{z}^{cv}_{ij} \sum_{\omega} \tilde{C}^A_{\omega p} \frac{\partial}{\partial \tilde{C}^A_{\omega q}} \left[\sum_{\mu\nu} \tilde{C}^A_{\mu i} \tilde{C}^A_{\nu j} \tilde{F}_{\mu\nu}[\tilde{\mathbf{D}}^A] \right] \tag{A2}$$

$$\tilde{A}^{A}_{pq}[\tilde{\mathbf{z}}] = \sum_{ai} \tilde{z}_{ai} \sum_{\omega} \tilde{C}^A_{\omega p} \frac{\partial}{\partial \tilde{C}^A_{\omega q}} \left[\sum_{\mu\nu} \tilde{C}^A_{\mu a} \tilde{C}^A_{\nu i} \tilde{F}_{\mu\nu}[\tilde{\mathbf{D}}^A] \right], \tag{A3}$$

we first solve the equation

$$\sum_{\omega} \tilde{C}^A_{\omega i} \frac{\partial \mathcal{L}}{\partial \tilde{C}^A_{\omega j}} - \sum_{\omega} \tilde{C}^A_{\omega j} \frac{\partial \mathcal{L}}{\partial \tilde{C}^A_{\omega i}} = \tilde{A}^{cv,A}_{ij}[\tilde{\mathbf{z}}^{cv}] - \tilde{A}^{cv,A}_{ji}[\tilde{\mathbf{z}}^{cv}] + A^A_{ij} - A^A_{ji} = 0 \tag{A4}$$

for $\tilde{\mathbf{z}}^{cv}$, where i and j label core and valence orbitals, respectively. Next, we determine the $\tilde{\mathbf{z}}$ multiplier from

$$\sum_{\omega} \tilde{C}^A_{\omega a} \frac{\partial \mathcal{L}}{\partial \tilde{C}^A_{\omega i}} - \sum_{\omega} \tilde{C}^A_{\omega i} \frac{\partial \mathcal{L}}{\partial \tilde{C}^A_{\omega a}} = \tilde{A}^{A}_{ai}[\tilde{\mathbf{z}}] - \tilde{A}^{A}_{ia}[\tilde{\mathbf{z}}] + \tilde{A}^{cv,A}_{ai}[\tilde{\mathbf{z}}^{cv}] + A^A_{ai} - A^A_{ia} = 0, \tag{A5}$$

where i denotes any occupied orbital. By substituting $\tilde{\mathbf{z}}$ and $\tilde{\mathbf{z}}^{cv}$ into the equation

$$\sum_{\omega} \tilde{C}^A_{\omega p} \frac{\partial \mathcal{L}}{\partial \tilde{C}^A_{\omega q}} = A^A_{pq} + \tilde{A}^{A}_{pq}[\tilde{\mathbf{z}}] + \tilde{A}^{cv,A}_{pq}[\tilde{\mathbf{z}}^{cv}] + x_{pq} = 0, \tag{A6}$$

we get the $\tilde{\mathbf{x}}$ multiplier. Finally, $\mathbf{y}$ is determined from the equation

$$\sum_{\omega} C^B_{\omega i} \frac{\partial \mathcal{L}}{\partial \tilde{C}^A_{\omega p}} = 0. \tag{A7}$$

The rest of the multipliers can be obtained similarly as in the case of the DFT-in-DFT embedding. The only difference is in the definition of the A_{pq} matrix of Eq. (18) due to the high-level Brillouin condition:

$$A_{pq} = \sum_{\omega} C_{\omega p} \frac{\partial}{\partial C_{\omega q}} \left[E_{\text{tot}} + \sum_{pj} y_{pj} \sum_{\mu\nu} \tilde{C}^A_{\mu p} S_{\mu\nu} C^B_{\nu j} + \sum_{ai} \tilde{z}_{ai} \sum_{\mu\nu} \tilde{C}^A_{\mu a} \tilde{C}^A_{\nu i} \tilde{F}_{\mu\nu}[\tilde{\mathbf{D}}^A] \right]. \quad \text{(A8)}$$

Appendix B: Comparison with numerical gradient

To verify the correctness of our implementation, the analytic gradients were compared to the numerical ones. To that end, we used a distorted ethanol molecule[36]. The analytic gradients were calculated with the default thresholds, whereas the numerical ones were determined using a four-point finite difference scheme with a step size of 0.005 bohr. We employed Dunning's cc-pVDZ basis set and a 770-point Lebedev grid in all calculations. The active space contains the OH group. The data are shown in Table VIII. The errors are well below our threshold for geometry convergence (10^{-4} a.u.). The differences are little bit larger for DFT calculations, which is a result of neglecting the geometry dependence of the integration grid.

TABLE VIII: Root-mean-square (RMS) error (in a.u.) of analytic gradients compared to numerical ones for a few methods.

Method	RMS error
PBE	$4.96 \cdot 10^{-7}$
PBE0-in-PBE	$1.62 \cdot 10^{-6}$
MP2-in-HF	$1.94 \cdot 10^{-8}$

We also have to note that the exact orthogonality condition is only fulfilled if the Huzinaga equations are solved exactly, which might bring in some error in the gradient as well. To check the accuracy of the orthogonality, the maximal element of the matrix $(\tilde{\mathbf{C}}^A)^{\mathrm{T}}\mathbf{S}\mathbf{C}^B$ was monitored for different convergence thresholds. For a looser threshold of 10^{-6} E_h, the maximal matrix element is in the order of 10^{-11}, while for the default threshold (10^{-8} E_h), it drops down to 10^{-13}. As a result, the orthogonality condition is satisfied at a very high accuracy even with loose convergence criteria, thus it should not introduce any measurable

error into the gradient.

REFERENCES

[1]W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[2]L. O. Jones, M. A. Mosquera, G. C. Schatz and M. A. Ratner, J. Am. Chem. Soc. **142**, 3281 (2020).

[3]L. W. Chung, W. M. C. Sameera, R. Ramozzi, A. J. Page, M. Hatanaka, G. P. Petrova, T. V. Harris, X. Li, Z. Ke, F. Liu, H.-B. Li, L. Ding, and K. Morokuma, Chem. Rev. **115**, 5678 (2015).

[4]C. E. Tzeliou, M. A. Mermigki and D. Tzeli, Molecules. **27**(9), 2660 (2022).

[5]F. Duarte, B. A. Amrein, D. Blaha-Nelson and S. C. Kamerlin, Biochim. Biophys. Acta. **1850**, 954 (2015).

[6]S. J. R. Lee, M. Welborn, F. R. Manby and T. F. Miller III, Acc. Chem. Res. **52**, 1359 (2019).

[7]T. Vreven and K. Morokuma, Chapter 3 hybrid methods: ONIOM(QM:MM) and QM/MM, in *Volume 2*, edited by D. C. Spellmeyer, volume 2 of *Annu. Rep. Comput. Chem.*, pp. 35–51. Elsevier, 2006.

[8]A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).

[9]S. Grimme, J. Chem. Phys. **124**, 034108 (2006).

[10]A. Warshel and M. Levitt, J. Mol. Biol. **103**, 227 (1976).

[11]G. Knizia and G. K.-L. Chan, J. Chem. Theory Comput. **9**, 1428 (2013).

[12]M. E. Fornace, J. Lee, K. Miyamoto, F. R. Manby and T. F. Miller III, J. Chem. Theory Comput. **11**, 568 (2015).

[13]T. A. Wesolowski and A. Warshel, J. Phys. Chem. **97**, 8050 (1993).

[14]T. A. Wesolowski, S. Shedge and X. Zhou, Chem. Rev. **115**, 5891 (2015).

[15]M. De Santis, D. Sorbelli, V. Vallet, A. S. P. Gomes, L. Storchi and L. Belpassi, J. Chem. Theory Comput. **18**, 5992 (2022).

[16]F. Maseras and K. Morokuma, J. Comput. Chem. **16**, 1170 (1995).

[17]T. Vreven, K. S. Byun, I. Komáromi, S. Dapprich, J. A. J. Montgomery, K. Morokuma and M. J. Frisch, J. Chem. Theory Comput. **2**, 815 (2006).

[18]M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber and K. Morokuma, J.

Chem. Phys. **100**, 19357 (1996).

[19]P. Cortona, Phys. Rev. B. **44**, 8454 (1991).

[20]K. Kitaura, E. Ikeo, T. Asada, T. Nakano and M. Uebayasi, Chem. Phys. Lett. **313**, 701 (1999).

[21]H.-A. Le, H.-J. Tan, J. F. Ouyang and R. P. A. Bettens, J. Chem. Theory Comput. **8**, 469 (2012).

[22]S. R. Gadre, R. N. Shirsat and A. C. Limaye, J. Phys. Chem. **98**, 9165 (1994).

[23]A. P. Rahalkar, M. Katouda, S. R. Gadre and S. Nagase, J. Comput. Chem. **31**, 2405 (2010).

[24]D. W. Zhang and J. Z. H. Zhang, J. Chem. Phys. **119**, 3599 (2003).

[25]O. R. Meitei and A. Heßelmann, J. Chem. Phys. **144**, 084109 (2016).

[26]F. R. Manby, M. Stella, J. D. Goodpaster and T. F. Miller III, J. Chem. Theory Comput. **8**, 2564 (2012).

[27]J. D. Goodpaster, T. A. Barnes, F. R. Manby and T. F. Miller III, J. Chem. Phys. **140**, 18A507 (2014).

[28]S. J. Bennie, M. W. van der Kamp, R. C. R. Pennifold, M. Stella, F. R. Manby and A. J. Mulholland, J. Chem. Theory Comput. **12**, 2689 (2016).

[29]S. J. Bennie, M. Stella, T. F. Miller III and F. R. Manby, J. Chem. Phys. **143**, 024105 (2015).

[30]M. Bensberg and J. Neugebauer, J. Chem. Phys. **150**, 184104 (2019).

[31]B. Hégely, P. R. Nagy, G. G. Ferenczy and M. Kállay, J. Chem. Phys. **145**, 064107 (2016).

[32]B. Hégely, P. R. Nagy and M. Kállay, J. Chem. Theory Comput. **14**, 4600 (2018).

[33]D. V. Chulhai and J. D. Goodpaster, J. Chem. Theory Comput. **13**, 1503 (2017).

[34]D. S. Graham, X. Wen, D. V. Chulhai and J. D. Goodpaster, J. Chem. Theory Comput. **16**, 2284 (2020).

[35]S. Huzinaga and A. A. Cantu, J. Chem. Phys. **55**, 5543 (1971).

[36]S. J. R. Lee, F. Ding, F. R. Manby and T. F. Miller, J. Chem. Phys. **151**, 064112 (2019).

[37]A. Kovyrshin and J. Neugebauer, Phys. Chem. Chem. Phys. **18**, 20955 (2016).

[38]D. Schlüns, M. Franchini, A. W. Götz, J. Neugebauer, C. R. Jacob and L. Visscher, J. Comput. Chem. **38**, 238 (2017).

[39]K. Klahr, D. Schlüns and J. Neugebauer, J. Chem. Theory Comput. **14**, 5631 (2018).

[40]S. Dapprich, I. Komáromi, K. S. Byun, K. Morokuma and M. J. Frisch, J. Mol. Struct.

(Theochem). **461-462**, 1 (1999).

[41]H. P. Hratchian, P. V. Parandekar, K. Raghavachari, M. J. Frisch and T. Vreven, J. Chem. Phys. **128**, 034107 (2008).

[42]N. J. Mayhall, K. Raghavachari and H. P. Hratchian, J. Chem. Phys. **132**, 114107 (2010).

[43]H. P. Hratchian, A. V. Krukau, P. V. Parandekar, M. J. Frisch and K. Raghavachari, J. Chem. Phys. **135**, 014105 (2011).

[44]O. R. Meitei and A. Heßelmann, J. Chem. Theory Comput. **17**, 1850037 (2018).

[45]M. Dułak, J. W. Kamiński and T. A. Wesołowski, J. Chem. Theory Comput. **3**, 735 (2007).

[46]J. Heuser and S. Höfener, J. Comput. Chem. **37**, 1092 (2016).

[47]J. Heuser and S. Höfener, J. Comput. Chem. **38**, 2316 (2017).

[48]J. Heuser and S. Höfener, J. Chem. Theory Comput. **14**, 4616 (2018).

[49]C. Huang and E. A. Carter, J. Chem. Phys. **135**, 194104 (2011).

[50]P. Jørgensen and T. Helgaker, J. Chem. Phys. **89**, 1560 (1988).

[51]T. Helgaker, P. Jørgensen and N. C. Handy, Theor. Chem. Acc. **76**(4), 227 (1989).

[52]J. M. Foster and S. F. Boys, Rev. Mod. Phys. **32**, 300 (1960).

[53]J. Csóka and M. Kállay, J. Chem. Phys. **158**, 024110 (2023).

[54]P. Pulay and S. Saebø, Theor. Chem. Acc. **69**, 357 (1986).

[55]E. A. Hylleraas, Z. Phys. **65**(3), 209 (1930).

[56]B. I. Dunlap, J. W. D. Connolly and J. R. Sabin, J. Chem. Phys. **71**, 3396 (1979).

[57]B. I. Dunlap, Phys. Chem. Chem. Phys. **2**, 2113 (2000).

[58]B. I. Dunlap, J. Mol. Struct. (Theochem). **529**, 37 (2000).

[59]K. Eichkorn, O. Treutler, H. Öhm, M. Häser and R. Ahlrichs, Chem. Phys. Lett. **240**, 283 (1995).

[60]F. Weigend, Phys. Chem. Chem. Phys. **4**, 4285 (2002).

[61]M. Feyereisen, G. Fitzgerald and A. Komornicki, Chem. Phys. Lett. **208**, 359 (1993).

[62]M. Kállay, P. R. Nagy, D. Mester, Z. Rolik, G. Samu, J. Csontos, J. Csóka, P. B. Szabó, L. Gyevi-Nagy, B. Hégely, I. Ladjánszki, L. Szegedy, B. Ladóczki, K. Petrov, M. Farkas, P. D. Mezei, and Á. Ganyecz, J. Chem. Phys. **152**, 074107 (2020).

[63]M. Kállay, P. R. Nagy, D. Mester, L. Gyevi-Nagy, J. Csóka, P. B. Szabó, Z. Rolik, G. Samu, J. Csontos, B. Hégely, Á. Ganyecz, I. Ladjánszki, L. Szegedy, B. Ladóczki, K. Petrov, M. Farkas, P. D. Mezei, and R. A. Horváth, Mrcc, a quantum chemical program suite. See https://www.mrcc.hu/. (Accessed December 1, 2023).

[64]P. A. M. Dirac, Proc. R. Soc. London A. **123**, 714 (1929).

[65]J. C. Slater, Phys. Rev. **81**, 385 (1951).

[66]J. P. Perdew, K. Burke and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[67]J. P. Perdew, M. Ernzerhof and K. Burke, J. Chem. Phys. **105**, 9982 (1996).

[68]C. Lee, W. Yang and R. G. Parr, Phys. Rev. B. **37**, 785 (1988).

[69]A. D. Becke, J. Chem. Phys. **88**, 2547 (1988).

[70]M. E. Mura and P. J. Knowles, J. Chem. Phys. **104**, 9848 (1996).

[71]T. H. Dunning Jr., J. Chem. Phys. **90**, 1007 (1989).

[72]D. E. Woon and T. H. Dunning Jr., J. Chem. Phys. **98**, 1358 (1993).

[73]F. Weigend, J. Comput. Chem. **29**, 167 (2008).

[74]F. Weigend, M. Häser, H. Patzelt and R. Ahlrichs, Chem. Phys. Lett. **294**, 143 (1998).

[75]F. Weigend, A. Köhn and C. Hättig, J. Chem. Phys. **116**, 3175 (2002).

[76]T. Földes, Á. Madarász, Á. Révész, Z. Dobi, S. Varga, A. Hamza, P. R. Nagy, P. M. Pihko and I. Pápai, J. Am. Chem. Soc. **139**, 17052 (2017).

[77]J.-D. Chai and S.-P. Mao, Chem. Phys. Lett. **538**, 121 (2012).

[78]É. Brémond and C. Adamo, J. Chem. Phys. **135**, 024106 (2011).

[79]F. Weigend and R. Ahlrichs, Phys. Chem. Chem. Phys. **7**, 3297 (2005).

[80]S. Grimme, J. Antony, S. Ehrlich and H. Krieg, J. Chem. Phys. **132**, 154104 (2010).

[81]S. Grimme, S. Ehrlich and L. Goerigk, J. Comput. Chem. **32**, 1456 (2011).

[82]L. Goerigk and S. Grimme, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **4**, 576 (2014).

[83]T. J. Goncalves, P. N. Plessow and F. Studt, ChemCatChem. **11**, 4368 (2019).

[84]O. Treutler and R. Ahlrichs, J. Chem. Phys. **102**, 346 (1995).

[85]P. R. Nagy, G. Samu and M. Kállay, J. Chem. Theory Comput. **14**, 4193 (2018).

[86]P. R. Nagy and M. Kállay, J. Chem. Theory Comput. **15**, 5275 (2019).

[87]R. Polly, H.-J. Werner, F. R. Manby and P. J. Knowles, Mol. Phys. **102**, 2311 (2004).