RUB

Implementation of the Frozen Density Embedding method in combination with Pair Natural Orbitals

Robert Sascha Treß

PhD Thesis – November, 2022.

Chair of Theoretical Chemistry

Supervisors: Prof. Dr. Christof Hättig and PD. Dr. Sebastian Höfener

Acknowledgment

First I want to to thank my supervisors Prof. Dr. Christof Hättig and PD Dr. Sebastian Höfener for all their time, help and advice they gave me during my work on this project and for introducing me to such an interesting project. Furthermore I thank Prof. Dr. Rochus Schmid for being a member of my PhD examination board. I also want to thank all members of the Quantum Chemistry group for the pleasant working atmosphere and the lots of fun I had during my time in this group thanks to them. Finally I want to thank all my friends and family for their emotional support.

Contents

1 Introduction **1**

1.1 Motivation and the state of research 1

1.2 Aim of the project 10

2 Theoretical Background and Layout of the Implementation **13**

2.1 Pair Natural Orbitals 13

2.2 Reaction Field Schemes 14

2.3 Implementation of the FDE method 18

3 Tackling the electron spill-out problem **25**

4 Combining PNOs with the FDE method **39**

5 Conclusions and Outlook **73**

6 Bibliography **77**

A Supporting Informations **87**

A.1 Tackling the electron spill-out problem 87

A.2 Combining PNOs with the FDE method 99

B FDE-Documentation **113**

B.1 FDE related keywords 113

B.2 Multilevel ansatz within FDE 115

B.3 The `FDE` script 116

C Structures **119**

Chapter 1

Introduction

1.1 Motivation and the state of research

The computation of properties of large systems such as excitation energies is a common problem in quantum chemistry. While accurate wavefunction methods are to expensive because of their steep scaling with the system size, methods like density functional theory (DFT) or molecular mechanics (MM) might make larger systems feasible but might not be able to describe the desired property with sufficient accuracy. In case of local excitations one additionally faces the problem to isolate them from the many excited states of a large system.

The simplest way to reduce the computation time to study local properties in a large system is a truncation of the total system to a smaller system. Such truncations are limited and usually still need most parts of the system in order to provide accurate results. Another way is provided by embedding methods, where the total system is divided into a subsystem of interest and its environment. Here the subsystem of interest is described with a method of choice, while the environment is described through a computationally less demanding method, and its influence on the active subsystem is included via an embedding potential $\nu_{emb}(r)$. Such methods can converge much faster to supermolecular results with the size of the active subsystem than simple truncation models. Many different embedding methods have been developed[1–5] so far and combined with various methods in order to describe different kinds of properties.

The simplest kind of embedding methods, beside mechanical and electrostatic coupling, are the continuum solvation models (CSM). Here the specific molecular structure of the environment is neglected and replaced by a continuum, which is characterized by its dielectric constant ε^{env}. The active subsystem is placed within a cavity inside this continuum and the electrostatic potential of the induced surface charges is included in the quantum chemical treatment of the active subsystem as the embedding potential. Commonly the cavity is discretized and the point charges q_s at the positions r_s on the cavity are used to approximate the surface charges leading to the following embedding potential:

$$\nu_{emb}^{CSM}[\rho_I](\mathbf{r}) = \sum_s \frac{q_s[\varepsilon^{env}, \rho_I]}{|\mathbf{r} - \mathbf{r}_s|} \tag{1.1}$$

Many varieties of these kind of models have been developed and mainly differ in their definition of the cavity and in the equations of the surface charges[2]. The most popular CSMs are the polarizable continuum model[6] (PCM), the conductor-like screening model[7] (COSMO) and the solvation model based on density[8] (SMD). CSMs are well suited to model solvent effects[9,10] since they implicitly include an average over the different solvent configurations by neglecting the specific molecular structure of the environment and using macroscopic solvent properties like the dielectric constant or the refractive index to describe the interaction with the solute. But this is also their biggest disadvantage, since the molecular structure of the environment can have important influence on the interaction between embedded species and its environment, in particular hydrogen-bonds where the solvent acts as hydrogen-bond donor can significantly shift excitation energies for transitions that involve a hydrogen-bond accepting lone-pair of the chromophore.

A simple way to introduce environment effects while keeping the information about the specific molecular structure is through classical MM as proposed by Warshel and Levitt[11]. Such methods where the embedded species are described through a quantum mechanical method (QM) and the effects of the environment through classical MM are called QM/MM methods. The environment is represented here through point charges q_B, whose positions are determined through MM calculations. This leads to an interaction energy that is given through the electrostatic interaction of these point charges with the nuclear charges Q_A and

the electron density $\rho_I(\mathbf{r})$ of the active subsystem:

$$E_{int}^{QM/MM} = \sum_{A}^{N_{at}} \sum_{B}^{N_{pc}} \frac{Q_A q_B}{R_{AB}} + \sum_{B}^{N_{pc}} \int \frac{\rho_I(\mathbf{r}) q_B}{|\mathbf{r} - \mathbf{R}_B|} d\mathbf{r} \ , \tag{1.2}$$

where R_{AB} is the distance between the nuclei A and the point charge B, N_{at} the number of atoms in the embedded system, and N_{pc} the number of point charges. The embedding potential simply given by:

$$\nu_{emb}^{QM/MM}(\mathbf{r}) = \sum_{B}^{N_{pc}} \frac{q_B}{|\mathbf{r} - \mathbf{R}_B|} \ . \tag{1.3}$$

The description of the environment can be further improved by taking the polarization of the environment into account. This was already proposed by Warshel and Levitt in their original work [11] and many ways to include polarization have been developed since then [12–16]. One way to include polarization is given by the direct reaction field (DRF) model. Here the electrostatic potential is generated through induced dipoles μ_B^{ind} at the positions $\mathbf{R}_B$:

$$\nu_{emb}^{DRF}[\rho_I](\mathbf{r}) = \sum_{B}^{N_{pc}} \frac{q_B}{|\mathbf{r} - \mathbf{R}_B|} + \sum_{B}^{N_{pc}} \frac{\mu_B^{ind}[\rho_I] \cdot (\mathbf{r} - \mathbf{R}_B)}{|\mathbf{r} - \mathbf{R}_B|^3} \tag{1.4}$$

Since the embedding potential depends on $\rho_I(r)$ and the induced dipoles depend on all other induced dipoles, the embedding potential needs now to be determined self consistently. This model was slightly improved by Kongsted and coworkers in their polarizable embedding (PE) method [17–19] by using not only partial charges but also multipole moments to describe the static part of the electrostatic potential. This kind of methods are widely used to study local properties in biological systems [3]. Nevertheless, in some cases the missing non-electrostatic contributions, such as the Pauli repulsion, can cause severe problems, e.g., electron spill-out, where the electron density is unphysically strong attracted by the positive point charges in the environment.

In purely quantum mechanical embedding methods, such as the frozen density embedding (FDE) method, also non-electrostatic contributions are included within the description of the interaction with the environment. The FDE method was developed by Wesołowski and Warshel [20] in 1993 within the framework of DFT and provides a formally exact description of the total system. Here the electron density of the total system (ρ_{tot}) is defined as the sum of the

electron density of the embedded species (ρ_I) and its environment (ρ_{II}):

$$\rho_{tot}(\mathbf{r}) = \rho_I(\mathbf{r}) + \rho_{II}(\mathbf{r}) \ . \tag{1.5}$$

With this partition of the electron density the energy of the total system (E_{tot}) can be expressed through the following equation:

$$E_{tot}[\rho_I(\mathbf{r}), \rho_{II}(\mathbf{r})] = E_I[\rho_I(\mathbf{r})] + E_{int}[\rho_I(\mathbf{r}), \rho_{II}(\mathbf{r})] + E_{II}[\rho_{II}(\mathbf{r})] \ , \tag{1.6}$$

where E_I represents the energy of the embedded species, E_{II} the energy of the environment, and the interaction energy E_{int} is defined as

$$\begin{aligned} E_{int}^{FDE}[\rho_I(\mathbf{r}), \rho_{II}(\mathbf{r})] = &\int\int \frac{\rho_I(\mathbf{r})\rho_{II}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} dr dr - \int \rho_I(\mathbf{r})\nu_{II}^N(\mathbf{r}) dr \\ &- \int \rho_{II}(\mathbf{r})\nu_I^N(\mathbf{r}) dr' + E_{kxc}^{nadd}[\rho_I(\mathbf{r}), \rho_{II}(\mathbf{r})] + E_{NN} \ . \end{aligned} \tag{1.7}$$

$\nu_X^N(\mathbf{r})$ is the nuclear potential of subsystem X and E_{kxc}^{nadd} contains kinetic and exchange-correlation contributions to $E_{kxc}[\rho_{tot}(\mathbf{r})]$, which are not covered by $E_I[\rho_I(\mathbf{r})]$ and $E_{II}[\rho_{II}(\mathbf{r})]$. The kinetic contributions to E_{kxc}^{nadd} can not be computed in the conventional way, with molecular orbitals as proposed by Kohn and Sham[21], since the wavefunction of the total system is not known. So similar to the exchange-correlation energy also the kinetic energy has to be computed using a density functional.

If the electron density of the environment is "frozen" and thus kept constant, the electron density of the active subsystem minimizes the energy of the total system. The functional derivative of the total energy with respect to the electron density of the active subsystem gives in addition to the effective potential of the isolated active subsystem the embedding potential for this subsystem:

$$\nu_{emb}^{FDE}[\rho_I(\mathbf{r}), \rho_{II}(\mathbf{r})] = \nu_{II}^N(\mathbf{r}) + \int \frac{\rho_{II}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}' + \frac{\partial E_{kxc}^{nadd}[\rho_I(\mathbf{r}), \rho_{II}(\mathbf{r})]}{\partial \rho_I(\mathbf{r})} \ . \tag{1.8}$$

Such a minimization leads to the exact total electron density if the electron density of the environment ρ_{II} and the density functionals are exact.

In practice neither the exact electron density of the environment nor the exact density functionals are available. Therefore several strategies have been developed to approximate both. The simplest approach to approximate the environment density is to use a computational inexpensive method to generate the electron density of the environment and use this density for the FDE method.

This approach is suited especially for environments which are not significantly polarizable[22], but depending on the size of the environment this can be computational demanding. Another approach is to generate the environment density based on a superposition of densities of fragments, which are commonly the different molecules within the environment. In this way the computational effort to generate ρ_{II} increases only linear with the number of fragments. This strategy was used in various studies[23–25] and offers a good approximation, if the environment comprises weakly interacting molecules. The most common approach simultaneously optimizes the electron density of the embedded species and the environment with the freeze-and-thaw (FaT) procedure[26]. Here, the role of the frozen density and the embedded species is interchanged in an iterative fashion between the two subsystems until both electron densities are converged. This then also includes the polarization of the environment, which leads to a qualitative improvement in particular for systems with charged species[27]. Other approaches include the dominant part of the external electric field generated by the embedded species during the generation of ρ_{II} to get a polarized electron density[27,28]. A continuum like representation of the environment within FDE[29,30] was achieved by using a statistically averaged electron density of the environment.

Beside approximations of the environment density, the exact density functionals are usually not known, so that approximation must be made for these as well. While for the subsystems themselves the common approaches of Kohn-Sham DFT can be used, other approaches are necessary to compute the nonadditive terms. A common approach is to use existing density functionals to compute the kinetic and exchange-correlation energies of the total electron density and the individual subsystem densities and then compute the nonadditive terms as the difference between them. In the same way, the derivatives necessary for the embedding potential are calculated. Such approximations for the nonadditive energy terms are denoted as decomposable approximations. While for the exchange-correlation energy various functionals have been developed for conventional DFT calculations, the kinetic energy is usually calculated using orbitals and therefore only a limited amount of kinetic energy functionals were developed. In regular density-gradient expansions (DGEs) the exact kinetic density

functionals for a uniform electron gas, the Thomas-Fermi functional[31,32], and for one electron or spin compensated two electron systems, the von Weizsäcker functional[33], were used to construct kinetic energy functionals, which proofed to be not very successful in modeling the kinetic energy or the nonadditive kinetic energy[34] if used self-consistently. Many kinetic energy functionals using the generalized gradient approximation (GGA) have been developed following the conjointness hypothesis[35,36]. According to this hypothesis the same enhancement factor used for the exchange functionals can be used for kinetic functionals as well. The widely used PW91k or LC94 kinetic energy functional[37] was developed following this strategy. Another type of kinetic energy functionals are based on the linear response of a homogeneous noninteracting electron gas given by the Lindhard response function[38–40]. The resulting non-local functionals are usually much more demanding than (semi)local density functionals, but can rather efficiently calculated in periodic systems.

Beside the decomposable approximations also so-called nondecomposable approximations exist, which directly calculate the nonadditive kinetic energy potential. An advantage of these kind of functionals is that the functionals can be constructed in such a way that quantities which can be directly extracted from the potential, i.e. multipole moments or excitation energies, are well reproduced instead of an approximate description of the kinetic energy[41,42]. With such functionals, however, the energy of the system can not be obtained. One example of such nondecomposable kinetic energy functionals is the "nondecomposable approximant using first and second derivatives" (NDSD). This functional enforces the correct behavior of the nonadditive kinetic potential at the frozen subsystem, such that for long distances the embedding potential vanishes at the frozen subsystem.[42]

If a KS-DFT calculation of the total system is possible the nonadditive kinetic energy potential can also be obtained through numerical inversion procedures. Here the nonadditive potential is obtained as the difference between the potentials of the total density and the embedded subsystem density. The potentials are obtained from an iterative scheme, where in the simplest case the potential is updated on a real space grid. In each iteration the current potential is used to compute an electron density, which is then compared with the

target density at each grid point. Based on the difference between the two densities the potential will either be made more repulsive or more attractive. Over the time, various formulas have been proposed for updating the potential[43–46]. Beside procedures on the real space grid various other procedures have been proposed[47–51] to increase efficiency and robustness of the potential reconstruction. Since supermolecular DFT calculations are necessary for these procedures the computational advantage gets lost, however these procedures can be used to study and develop new approximate kinetic energy functionals.

That the embedding potential depends only on the electron densities of the subsystems makes it possible to extend the FDE method to wave function theory (WFT) methods as well. Such WFT-in-DFT approaches have been developed for Hartree-Fock[52](HF), Møller Plesset perturbation theory[53] (MPn), Configuration interaction[54] (CI), CASSCF and coupled cluster (CC) methods.[55,56] A strict formulation of such WFT-in-DFT approaches was given by Wesołowski[57].

The FDE method has been used to study various properties like NMR shieldings[58,59] and ESR[22,60] properties, but one of its most important fields are excited states and electronic spectra of chromophores in various environments. The first excitation energies determined by the FDE method were extracted from orbital energy differences in metal complexes, for example, lanthanide-halide clusters in elpasolite crystals[27,28] or MnF_6 complexes in fluoroperovskites[61]. A first linear response formulation within FDE was presented by Casida and Wesołowski[62] in 2004 for time dependent DFT (TDDFT) and was later on implemented as a simplified version[63], where the response contributions of the environment are assumed to be zero. This approach was later denoted as "uncoupled FDE"[64] (FDEu) or as "neglect of dynamic response of environment"[65] (NDRE). In 2007 an extension to an arbitrary number of subsystems was published[64], which includes effects resulting from changes in the electron densities of all subsystems. This approach allows to calculate excitonically coupled monomer excitations and is therefore denoted as coupled FDE (FDEc). Later on this approach was rederived to introduce the Tamm-Dancoff approximation[66] and to distinguish coupled and uncoupled response functions with Dyson-type equations[67]. In various studies it was shown that the FDEu approach leads to reliable results if the excitation energies of the embedded subsystem and the

environment are far apart[68–71], but in case of several interacting chromophores where the excitation can no longer be described in terms of local transitions the FDEc approach has to be employed[72–74]. For WFT-based methods a linear response formulation was given by Höfener et. al.[55,56] who demonstrated their approach for excitations in water molecules within a water cluster using time dependent HF (TDHF). Since DFT provides in most cases already an accurate ground state density and WFT is usually only necessary to overcome the difficulties of TDDFT, a scheme was proposed which uses conventional DFT-in-DFT calculations to determine the embedding potential and use this fixed embedding potential in a following WFT-in-DFT calculation for local excited states[75]. Beside linear response formulations also a state-specific approach was formulated by Khait and Hoffman[76,77] for WFT-in-DFT approaches.

A common problem in QM/MM calculations is the so-called electron spill out, where the electrons of the embedded subsystem are unphysically strong attracted by regions in the environment with a positive electrostatic potential, in particular regions close to the nuclei of the environment, because the Pauli repulsion is not included within the purely electrostatic interaction. This problem especially occurs if diffuse basis functions are used since they reach further into the region of the environment. The FDE method includes besides electrostatic contributions also the nonadditive contributions of the exchange-correlation and kinetic energy within the embedding potential. The nonadditive kinetic energy contributions provide here an additional repulsive contributions to hinder electrons to move to the nuclei of the environment. Therefore it is expected that the FDE method suffers less from the electron spill-out problem than MM embedding methods that use a purely classical description of the environment. Dułak and Wesołowski[78] investigated this problem for the two systems $F^- \cdot H_2O$ and $Li^+ \cdot H_2O$ and could show based on the interaction energies that this problem occurs only in case of pure electrostatic embedding potentials but not if the full FDE embedding potential is used. In another study Fradelos and Wesołowski[79] studied excitation energies of hydrogen bonded complexes for different basis set sizes and embedding potentials. They showed that the excitation energies collapse in case of large basis sets with diffuse basis functions and a purely electrostatic embedding potential and that the numerical stability can be restored with

a full FDE embedding potential. Nevertheless improvements of the embedding potential have been developed to overcome problems caused by the approximate nonadditive kinetic energy contributions. Jacob et al[42] developed a long distance correction that assures the correct description of the embedding potential at the frozen environment in the long distance limit. An improved description of the Pauli repulsion is given in projection based embedding models[80–86], where the orthogonality of the molecular orbitals between the subsystems is assured. However, the orthogonalization has a cubic scaling of its computational demand with the size of the basis of the full system and is therefore less suited for large systems.

Another important route to make wavefunction-based methods applicable to larger molecular system, which several research groups have been working on in the last decades, are local correlation methods. In particular the Pair Natural Orbital methods, which truncate the virtual space based on the correlation contributions of individual pairs of orbitals, have proven to be very efficient. Here the spatial characteristics of the two electron integrals of the type $(ia|bj)$, which are related to the computational most demanding steps in second-order correlation methods, are exploited. These integrals are non-negligible only if the virtual molecular orbitals (MO) a and b are in the spatial vicinity of the occupied MOs i and j. Moreover, the magnitude of these integrals decreases with the power of three with the distance between the centers of i and j. Based on these relations PNOs with only small contributions can be discarded, which leads to a reduction of the computational demand for the following calculation. Beside PNOs such a truncation of the basis size is also performed in the projected atomic orbitals[87,88] (PAO) approximation and in the concept of orbital specific virtuals[89,90] (OSV), but compared to these two approaches, PNOs have the advantage that they require less number of pairs to achieve the same accuracy and that their truncation depends only on one parameter.

Already in the 1970s the PNO approximation was formulated and became popular[91–103] because due to their enormous reduction of the virtual space, they made wave function methods, such as coupled electron pair approximation (CEPA) and CI methods, feasible for molecules for which these calculations had at that time not been possible with the available computer hardware with

conventional implementations. The high computational demand for the transformation of the two electron integrals into the PNO basis, however, led to a loss of interest in PNOs. It was not until 2009 that the PNO approach was revived by Neese et. al.[104,105], who realized that this problem could be avoided through the meanwhile developed approximations for the decomposition of the two electron integrals[106,107]. The achieved efficiency and accuracy also motivated other groups to use the PNO approximation beyond ground-state correlation energies. With that the PNO approach has been applied for F12-methods[108–110], excitation energies[111–113] with CIS(D), ADC(2) and CC2, for MP2 gradients[114] and for CCSD[105] and CC3[115]. Further improvements in the PNO generation could be achieved by using an intermediate OSV basis[116,117].

1.2 Aim of the project

One of the largest problems within FDE results from the necessary approximation of the nonadditive kinetic energy contributions. In practice decomposable approximations are most commonly used, but here the problem arises that the errors in the kinetic energy for the electron densities of the total system and the active and environment subsystems do not compensate each other when functionals originating from the regular gradient expansion[5] for the kinetic energy are used. As shown by numerical results there is no correlation between the errors of the nonadditive kinetic energy contributions and the errors of the corresponding kinetic energy contributions for the individual electron densities[118]. The same applies for their derivatives[23,41,118]. Several studies on approximate nonadditive kinetic energy functionals indicate that simple GGA type approximations are only applicable when the electron densities are weakly overlapping.[41,119–121] In case of systems with covalent bonds between the embedded species and the environment applying the approximate nonadditive kinetic energy functionals usually leads to wrong results.[122–124] This leaves several systems with only one option, which is to increase the size of the embedded subsystem until there are no more covalent bonds with the environment. The resulting subsystems can be large, which may limit the number of methods available to model that subsystem due to the computational demands. Particularly accu-

rate second-order correlation methods with their steep scaling with the number of basis functions may no longer be practical. However, such second-order correlation methods are popular to compute excitation energies because of their reliable accuracy.

An improved description of the embedding potential, in particular the non-additive kinetic energy potential, with the above mentioned numerical inversion procedures, allows to describe covalent bonds between subsystems [123] and hence keep subsystems within sizes feasible for second-order correlation methods. However, the necessary supermolecular DFT calculations still make this method computationally demanding. Another approach to handle covalent bound environments is given by the so called "3-FDE" method [125]. Here, the subsystems are separated at the covalent bond and capping groups are added to each subsystem to ensure the structure of the covalent bond in each subsystem. A third subsystem is then defined as a combination of the capping groups at each subsystem, and the density of this "capping subsystem" is subtracted from the combined densities of the other two subsystems to form the total electron density. In FaT iterations, the densities of the two subsystems are then optimized while the density of the "capping subsystem" is kept fixed. The efficiency of this approach was demonstrated for proteins, where the electron densities of supermolecular DFT calculations could be reproduced with good accuracy. However, due to an additional condition that ensures the positivity of the electron density, the electron density around the covalent bond is fixed and a polarization of the bond cannot be described.

This work follows a different strategy to overcome the problem of large subsystems within FDE embedding calculations. Instead of developing ways to decrease the size of subsystems even further, the goal is to combine the FDE method with pair natural orbitals (PNOs) to make accurate second-order wave function methods within FDE feasible for even larger subsystems. While this combination offers the possibility to investigate large subsystems within complex environments it also introduces an additional approximation into the FDE calculation. Therefore, an investigation of the introduced PNO approximation error for environment-induced properties is required to ensure the accuracy of this combination. A very important quantity in theoretical spectroscopy is the

environment-induced shift of excitation energies. These shifts are themselves rather small, so that an additional approximation might render a qualitative description of these shifts no longer possible. Also, the convergence of properties of the embedded subsystem with the size of the environment is an important feature that must be ensured even after the introduction of the PNO approximation. Therefore an implementation of the combination of PNOs with FDE is presented within the TURBOMOLE program package and its efficiency and accuracy tested for the calculation of excitation energies.

Beside the problem of large subsystems a second problem caused by the approximate nonadditive kinetic contributions might arise in excitation energy calculations: the electron spill-out problem. This problem is well known in QM/MM embedding and arises due to the lack of Pauli repulsion between the subsystems. In FDE the nonadditive kinetic energy offers an additional repulsive contribution to the embedding potential and is able to prevent electron spill-out in systems, where pure electrostatic embedding methods failed[78,79]. Nevertheless the robustness depends on the quality of the approximation of the kinetic density functional and further investigations in case of excitation energy calculations is important to ensure the accuracy of these. Therefore also the electron spill-out problem is investigated and a possible improvement of the description of the Pauli repulsion is proposed.

In the following chapter (chapter 2), the theoretical background of PNOs and reaction field schemes is explained and the structure of the FDE implementation is presented. Then, in Chapter 3, the electron spill-out problem for FDE calculations is investigated and a solution is proposed. In Chapter 4, the proposed combination of PNO and FDE is tested for efficiency and accuracy. In conclusion, Chapter 5 provides a summary of the project and an outlook for further improvements of the implementation.

Chapter 2

Theoretical Background and Layout of the Implementation

2.1 Pair Natural Orbitals

The electron correlation is a short range effect that decays with the distance as R^{-6} and in order to exploit this short-ranged nature PNOs are defined for each occupied pair of MOs i and j as a linear combination of the virtual orbitals a:

$$|\bar{a}\rangle = \sum_a |a\rangle d^{ij}_{a\bar{a}} \ . \tag{2.1}$$

The coefficients $d^{ij}_{a\bar{a}}$ are chosen such that they diagonalize the pair contribution D^{ij} of pair ij to the one electron matrix[104]:

$$d^{ij}_{a\bar{a}} D^{ij}_{ab} d^{ij}_{b\bar{b}} = n^{ij}_{\bar{a}} \delta_{\bar{a}\bar{b}} \ . \tag{2.2}$$

The eigenvalues $n^{ij}_{\bar{a}}$ are pair-specific natural occupation numbers and show a fast convergence towards zero. Most contributions to the correlation energy are already obtained through the MP2 method with a wavefunction of first order. Therefore a evaluation of the significance of the PNOs can already be carried out at the MP2 level. The pair contributions to the density matrix are then

given for the closed-shell case as:

$$D_{ab}^{ij} = \frac{2}{N_{ij}} \left(u_{ab}^{ij} t_{bc}^{ij} + u_{ca}^{ij} t_{cb}^{ij} \right), \text{ with} \tag{2.3}$$

$$N_{ij} = 1 + \frac{2}{1+\delta_{ij}} \sum_{ab} \left(u_{ab}^{ij} t_{ab}^{ij} \right) \text{ and} \tag{2.4}$$

$$u_{ab}^{ij} = 2t_{ab}^{ij} - t_{ba}^{ij} \tag{2.5}$$

and the MP1 doubles amplitudes as

$$t_{ab}^{ij} = \frac{(ai|bj)}{\varepsilon_a + \varepsilon_b - \varepsilon_i - \varepsilon_j} \ . \tag{2.6}$$

The number of PNOs is truncated based on their natural occupation number $n_{\bar{a}}^{ij}$ and all PNOs below a defined threshold value T_{PNO} are neglected. This convergence can be further improved by using localized molecular orbitals for occupied orbitals since the PNOs with the highest occupation numbers are spatially close to pairs of occupied orbitals that are themselves spatially close to each other. Doubles amplitudes and two electron integrals are then transformed into the resulting truncated PNO basis. For excitation energies the constructed ground-state PNOs perform poorly since they do not account for the different characters of excited states[111]. To deal with this problem, state specific PNOs are constructed in a similar fashion, where the ground-state doubles amplitudes are replaced with the first-order excited state doubles amplitudes[111]:

$$\bar{D}_{ab}^{ij} = 2 \left(\bar{u}_{ac}^{ij} E_{bc}^{ji} + \bar{u}_{ca}^{ij} E_{cb}^{ij} \right) \ . \tag{2.7}$$

2.2 Reaction Field Schemes

Reaction field theory provides a theoretical framework how the Schrödinger equation for a system that consists of an "active" part of interest and a (polarizable) environment can be reduced to an effective Schrödinger equation for the wavefunction of the active part only:

$$\hat{H}^{tot} \Psi^{tot} = E^{tot} \Psi^{tot} \longrightarrow \ \hat{H}_{eff}^{act} \Psi^{act} = E_{eff}^{act} \Psi^{act} \ . \tag{2.8}$$

With that reaction field theory provides a common theoretical framework for approaches that embed a quantum system in a polarizable environment as it is f.x. the case for FDE. To achieve such a description the total wavefunction is

approximated by an antisymmetrized product of a wavefunction for the active subsystem $|\Psi^{act}\rangle$ and a wavefunction for the environment $|\Psi^{env}\rangle$:

$$|\Psi^{tot}\rangle \approx |\Psi^{act}\Psi^{env}\rangle = \hat{\mathcal{A}}|\Psi^{act}\rangle|\Psi^{env}\rangle \tag{2.9}$$

with $\hat{\mathcal{A}}$ as the antisymmetrizing operator. The Hamiltonian is partitioned into $\hat{H}^{act}$, which describes the isolated active subsystem, $\hat{H}^{env}$, which describes the isolated environment and $\hat{H}^{int}$ describing the interaction between the two subsystems.

$$\hat{H}^{tot} = \hat{H}^{act} + \hat{H}^{env} + \hat{H}^{int} \tag{2.10}$$

Inserting this into the formula for the total energy gives:

$$\begin{aligned} E^{tot} &= \langle\Psi^{tot}|\hat{H}^{tot}|\Psi^{tot}\rangle \approx \langle\Psi^{act}\Psi^{env}|\hat{H}^{act} + \hat{H}^{env} + \hat{H}^{int}|\Psi^{act}\Psi^{env}\rangle \\ &\approx \langle\Psi^{act}|\hat{H}^{act} + \hat{G} + E^{env}|\Psi^{act}\rangle \ , \end{aligned} \tag{2.11}$$

with the reaction field:

$$\hat{G} = \langle\Psi^{env}|H^{int}|\Psi^{env}\rangle \ , \tag{2.12}$$

and

$$E^{env} = \langle\Psi^{env}|H^{env}|\Psi^{env}\rangle \ . \tag{2.13}$$

For a polarizable environment its wavefunction and density depends on a potential arising from the active subsystem, which therefore leads to a dependence of the reaction field on the electron density of the active subsystem:

$$\hat{G} = \hat{G}\left[\left(\Psi^{act}\right)^* \Psi^{act}\right] \ . \tag{2.14}$$

A minimization with respect to the wavefunction of the active subsystem leads then to a nonlinear Schrödinger equation.

$$\left(\hat{H}^{act} + 2\hat{G}\left[\left(\Psi^{act}\right)^* \Psi^{act}\right] + E^{env}\right)|\Psi^{act}\rangle = E^{act,env}|\Psi^{act}\rangle \tag{2.15}$$

This non-linear equation is then solved in an iterative cycle of optimizing the wavefunction of the subsystem and calculation of the density and reaction field until self-consistency is reached. Since this process is similar to the SCF procedure, an integration of the reaction field is straightforward. Figure 2.1 pictures

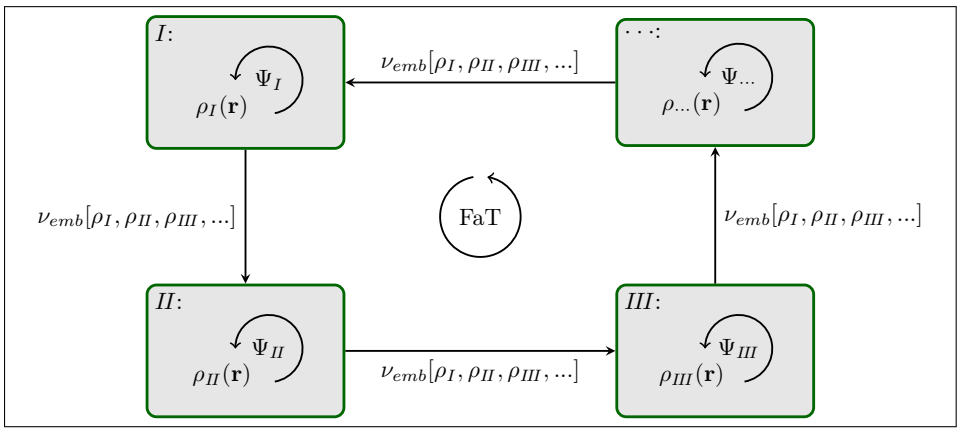


Figure 2.1: Schematic of the FDE reaction field procedure.

the iterative process of the self consistent reaction field optimization for a FDE formulation with an arbitrary number of subsystems. Here, the wavefunction of a subsystem is optimized in the presence of the embedding potential, which is calculated using the electron densities of the individual subsystems, and then the electron density of the active subsystem is calculated from the now optimized wavefunction, which in turn is used to calculate the embedding potential of the next subsystem. If this is done for each subsystem one FaT iteration is completed and more FaT iterations will be done until the system reaches self consistency. With only two subsystems, the procedure is similar to most embedding methods, where only the density of the embedded species and the environment are optimized alternately.

While the calculation of the reaction field is straight forward within self consistent field (SCF) methods, like Hartree-Fock (HF) and Density Functional Theory (DFT), multiple approaches exist for post-HF methods. Figure 2.2 displays the three most common approaches, the perturbation to the energy (PTE) approach, the perturbation to the energy and density (PTED) approach and the post-SCF reaction field scheme. The simplest approach of these three is the PTE approach. Here the HF reaction field is used without further changes in the following post-HF calculations and neglects any correlation contributions to the reaction field. This approach is consistent for the MP2 method but constitutes an additional approximation beyond MP2[126]. A more complete reaction field is obtained with the PTED approach. Here the correlated density is

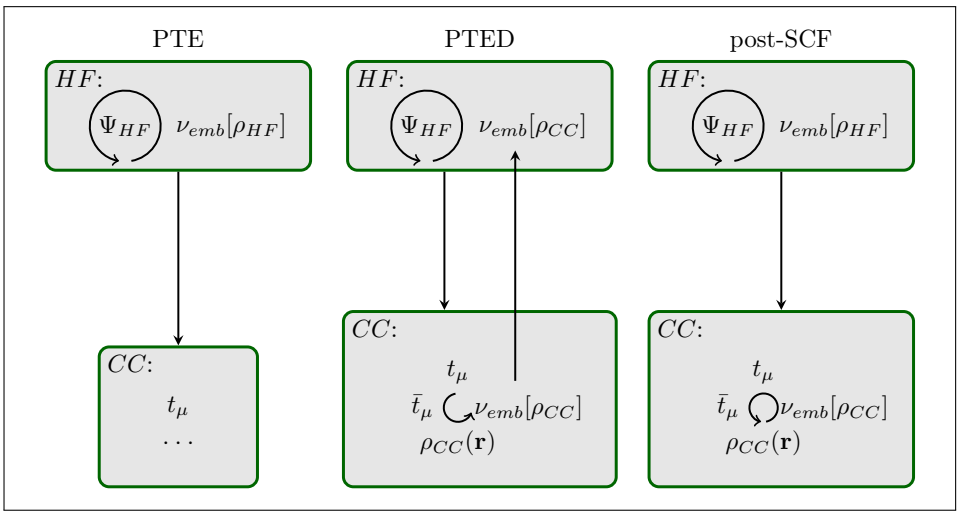


Figure 2.2: Schematic of different reaction field schemes for correlated wavefunctions.

used to calculate an improved reaction field potential which is then used for the Hartree-Fock calculation in the next SCRF cycle, i.e. the next macro-iteration over the HF and post-HF calculation. In case of FDE this is done inside the FaT iterations, which might lead to a large number of post-HF calculations. Since post-HF methods such as second-order correlation methods are usually computational demanding, this scheme is not feasible for all systems. The third model, the post-SCF reaction field, represents a compromise between PTE and PTED. Here the correlation contributions to the reaction field are computed within the post-HF calculation but not reintroduced back into the HF calculation. In case of FDE this means that within the FaT iterations only HF is used to reach a self consistent reaction field and afterwards the post-HF method is used and changes resulting from the correlated density are only added there. In this way most of the correlation contributions are included in the reaction field without the need to perform multiple post-HF calculations for the active subsystem. The post-SCF scheme was originally developed for PE and COSMO calculations[127–129], where a post-HF calculation is much more computationally demanding than the update of the reaction field. Here the reduction of the number of post-HF calculation resulted in a huge time save compared to the PTED scheme.

Further complications because of the non-linearity of the effective Schrödinger

equation for the active system arise for excitation energies. The simplest approach is to use the frozen solvent approximation[130]. Here all changes in the environment with respect to the excitation of the embedded species are neglected and the excited state is determined within the reaction field of the ground state. This leaves the nuclear and electronic response of the environment unaccounted for. The nuclear response is compared to the electronic response slow and in case of fast changes in the density of the embedded species, i.e. in vertical excitations, only the electronic response is considered fast enough, while the nuclei stay in their initial configuration for a short time scale. The electronic response beyond the frozen solvent approximation consists of two main effects. First the interaction of the transition density of the embedded species with the dynamic response of the reaction field, which is called excitonic coupling and second the polarization of the environment upon the change in the electron density of the embedded species[10,131]. The first effect is important for bright electronic transitions with large transition dipole moments and is included within the linear response theory of cFDE, but the second effect is not taken into account within linear response theory[64,131]. In state-specific excitation energy calculations the polarization of the environment can be taken into account by performing the FaT iterations for the excited state, but the state specific approach has the problem that the excited states are no longer orthogonal to each others leading to problems in the vicinity of conical intersections and avoided crossings[132].

For the COSMO embedding model different reaction field models have been studied for excitation energy calculations in ref. 132. They could show that statistically only a minor difference in excitation energies exists between the post-SCF and PTED approach for second-order correlation methods. Therefore the post-SCF reaction field model is used in the following FDE implementation to reduce the computational demand.

2.3 Implementation of the FDE method

The starting point of the project was a FDE implementation by Laricchia et al.[133] within the TURBOMOLE program package. Originally developed to study the accuracy of the FDE method if hybrid density functionals are used to

relax the density of each subsystem, the implementation was limited to systems consisting of two subsystems and could only be used with the DFT method within *dscf* program. Later on the implementation was used to develop new kinetic and exchange-correlation density functionals[134–137] and to study the accuracy of the FDE method for different kinds of density functionals[133,137–139] and for ground-state charge-transfer complexes[140]. During that time it was further extended to be used together with the localized Hartree-Fock method[141] and for subsystems with non-integer particle numbers[142].

The implementation uses the decomposable approximation for the nonadditive energy contributions, where the embedding potential is given as:

$$\nu^i_{emb}(\mathbf{r}) = \frac{\partial E^{tot}_{kxc}[\rho_{tot}(\mathbf{r})]}{\partial \rho_i(\mathbf{r})} - \frac{\partial E^{i}_{kxc}[\rho_{i}(\mathbf{r})]}{\partial \rho_i(\mathbf{r})} + \int \frac{\rho_{env}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}' + \nu^N_{env}(\mathbf{r}) \ . \quad (2.16)$$

However, within the FDE procedure the embedding potential is computed in the basis of the active subsystem i from the integrated contributions:

$$F^{emb}_{i,\alpha\beta} = \langle \alpha | \nu^i_{emb}(\mathbf{r}) | \beta \rangle = F^{kxc}_{tot,\alpha\beta} - F^{kxc}_{i,\alpha\beta} + F^{coul}_{env,\alpha\beta} \ , \quad (2.17)$$

with α and β as the atomic orbitals (AOs) of subsystem i. Hereby the contributions F^{kxc}_{tot} and F^{coul}_{env} are obtained from a SCF iteration in the supermolecular basis and their monomolecular submatrix is used to compute the embedding potential. After I extended the implementation in a first work package to an arbitrary number of subsystems m, the implementation proceeds according to the following scheme:

1. Compute isolated subsystems and save their KXC contributions to the Fock Operator (F^{kxc}_i).

2. Start freeze-and-thaw iteration $n = 1$ for subsystem $i = 1$:

   (a) Compute KXC-contributions to the Fock Operator of the total system and save it in the basis of each subsystem (F^{kxc}_{tot}).

   (b) Compute Coulomb contributions to the embedding potential F^{coul}_{env} from a supermolecular calculation of the environment.

   (c) Compute embedding potential for subsystem i according to equation 2.17, relax the density of subsystem i and save F^{kxc}_i.

(d) If $i = m$ continue with step 2e; else set $i = i + 1$ and repeat from step 2b

(e) Compute total energy with embedded MOs.

(f) If $n = n_{max}$ or the maximum change of the dipole moment within all subsystems is below the convergence threshold continue with step 3; else set $n = n + 1$ and repeat from step 2a.

3. Calculate properties.

The steps 2a, 2b and 2e require each one SCF iteration within the supermolecular basis of the system even for a monomolecular calculation. This leads in case of a monomolecular FDE calculation for a system consisting of m subsystems to $n \times (m + 2)$ supermolecular SCF iterations, where n is the number of performed freeze-and-thaw iterations. In case of a partitioning into many subsystems this leads to a high number of supermolecular SCF iterations, which can be even higher than the necessary amount of SCF iterations in a conventional supermolecular DFT calculation. On the other hand, partitioning into a small number of subsystems leads to almost no time savings, up to even longer computation times due to the remaining supermolecular SCF iterations.

In order to demonstrate the flaws of this implementation with respect to large systems, computational wall times have been determined for various systems of acetone in water, which differ in the number of water molecules around the acetone. The structures of the used cluster can be found in appendix C.

The computational wall times of these systems have been determined for conventional supermolecular Kohn-Sham-DFT calculations and for monomolecular FDE calculations with different numbers of defined subsystems on two Intel(R) Xeon(R) Gold 6230 CPUs with 2.10GHz using OpenMP parallelization on 40 cores in total. Subsystems have been defined in three different ways. First all water molecules have been defined as one subsystem (2sub), second each shell of additional water molecules has been defined as one subsystem (shellsub) and third every water molecule was defined as a separate subsystem (maxsub). The def2-TZVP basis set[143] was used for all atoms and the DFT calculations have been performed with the PBE exchange-correlation density functional[144] and a grid size of 3. FDE embedding calculations used in addition the LC94 ki-

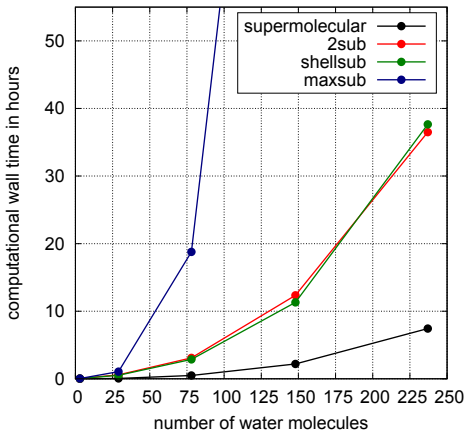


Figure 2.3: Computational wall times in hours with respect to the number of water molecules in the environment. Wall times of conventional supermolecular KS-DFT calculations are shown in black. Monomolecular FDE calculations with all water molecules as one subsystem (2sub) in red, with each molecule as a separate subsystem (maxsub) in blue and with each shell of additional water molecules as a separate subsystem (shellsub) in green.

netic density functional[37] for the embedding potential and the FaT iterations have been performed until the dipole moments of all subsystems reached the convergence threshold of 0.005 Debye.

Figure 2.3 displays the computational wall times against the number of water molecules in the system. One can see that the computational wall time increases much faster with the number of water molecules for the FDE calculations compared to the supermolecular calculation. This is especially the case if each water molecule is defined as a separate subsystem, which results from the many supermolecular SCF iterations that are performed. Even though the FDE calculations can save a lot of computation time in following property calculations due to the monomer basis, such a steep scaling with the number of subsystems is not suited for large environments.

In a second work package I extended the implementation to a purely monomolecular procedure. Here the embedding potential is calculated on an supermolecular integration grid ,which is generated at the beginning of the FDE calculation, and uses only the densities and electrostatic potentials of the other subsystems. The embedding potential at the grid point $\mathbf{r}_g$ of subsystem i in a system of m

subsystems is then given as:

$$\nu^i_{emb}(\mathbf{r}_g) = \frac{\partial E_{kxc}[\sum_{j=1}^{m} \rho_j(\mathbf{r}_g)]}{\partial \rho_i(\mathbf{r}_g)} - \frac{\partial E_{kxc}[\rho_i(\mathbf{r}_g)]}{\partial \rho_i(\mathbf{r}_g)} + \sum_{j \neq i}^{m} \left(\nu^J_j(\mathbf{r}_g) + \nu^N_j(\mathbf{r}_g)\right) \ , \tag{2.18}$$

where ν^J_j is the electrostatic potential of the electron density of subsystem j and the FDE calculation proceeds according to the following scheme:

1. Generate supermolecular integration grid.
2. Compute isolated subsystems and save their densities ($\rho_i(\mathbf{r})$) and electrostatic potential ($\nu^J_i(\mathbf{r}) + \nu^N_i(\mathbf{r})$) on the grid.
3. Start freeze-and-thaw iteration $n = 1$ for subsystem $i = 1$:
   (a) Start macroiterations for subsystem i:
      i. Compute embedding potential for subsystem i according to equation 2.18.
      ii. Relax the density of subsystem i and save the relaxed density $\rho_i(\mathbf{r})$ and electrostatic potential ($\nu^J_i(\mathbf{r}) + \nu^N_i(\mathbf{r})$).
   (b) If $i = m$ continue with step 3c; else set $i = i + 1$ and repeat from step 3a
   (c) If $n = n_{max}$ or the maximum change of all subsystem energies is below the convergence threshold continue with step 4; else set $n = n + 1$ and repeat from step 2a.
4. Calculate properties.

Following this procedure the FDE method was implemented for DFT and the HF method in the *dscf* program, together with the resolution of identity (RI) approximation into the *ridft* program and for second-order correlation methods into the *ricc2* program of the TURBOMOLE program package. The second-order correlation methods have been implemented for the ground state within the PTE and post-SCF reaction field scheme and for excitation energies within the FDEu approach and the frozen solvent approximation. The implementation also employs the OpenMP and Message-Passing Interface (MPI) parallelization

schemes and was validated by comparisons with the quantum chemistry software KOALA [145].

Chapter 3

Tackling the electron spill-out problem

This project part investigates electron spill-out in excited states with the FDE method and tackles this nonphysical behavior by introducing all-electron pseudopotentials for the environment atoms to model missing repulsive effects. Since embedding methods like QM/MM suffer from electron spill-out in excited state calculations, because of the missing Pauli repulsion in the description of the interaction, an investigation of this problem for the FDE method is only reasonable if one is interested in excitation energies. Furthermore, the question arises whether an additional treatment of the Pauli repulsion with all-electron pseudopotentials, as it has been done in polarizable embedding (PE) [146], introduces errors due to double counting of repulsive terms with the non-additive kinetic energy contributions in the FDE formalism.

The project part has been performed under the supervision of Prof. Dr. Christof Hättig and PD Dr. Sebastian Höfener. Implementation of the purely monomolecular FDE method for the ground state has been performed by me based on the implementation by Laricchia et al. [133]. All numbers, graphics and the all-electron pseudopotentials of bromine and iodine have been created by me. The first draft of the manuscript was written by me and further developed together with the co-authors.

The following publication is reprinted with permission from *J. Chem. Theory*

Comput. 2022, **18**, 3, 1737–1747; **Copyright © 2022 American Chemical Society**. Its corresponding supporting information can be found in section A.1. This publication is also available online: `https://doi.org/10.1021/acs.jctc.1c00732`

JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

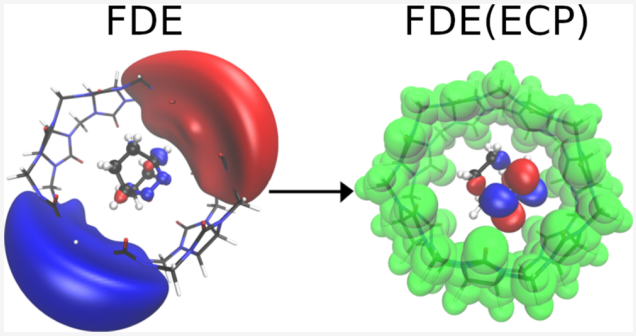
Employing Pseudopotentials to Tackle Excited-State Electron Spill-Out in Frozen Density Embedding Calculations

Robert S. Treß,* Christof Hättig, and Sebastian Höfener*

Cite This: https://doi.org/10.1021/acs.jctc.1c00732 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: In frozen density embedding (FDE), the properties of a target molecule are computed in the presence of an effective embedding potential, which accounts for the attractive and repulsive contributions of the environment. The formally exact embedding potential, however, is in practice calculated using explicit kinetic-energy functionals for which the resulting potentials are in many cases not repulsive enough to account fully for Pauli repulsion by the electrons of the environment and to compensate thereby the strong electron−nuclear attraction. For the excited states on the target molecule, this leads to charge spill-out when diffuse basis functions are included, which allow that valence electrons are excited to those regions of the environment where the strong nuclear attraction is not sufficiently compensated by repulsive contributions. To reduce this insufficiency, we propose in the present work the inclusion of atomic all-electron pseudopotentials for all environment atoms on top of the conventional embedding potential. In the current work, the pseudopotentials are applied for computing vertical excitation energies of local excited states in complex systems employing the second-order algebraic diagrammatic construction (ADC(2)) scheme. The proposed approach leads to significantly reduced charge spill-out and an improved agreement of FDE and supermolecular calculations in the frozen solvent approximation. In particular, when diffuse functions are employed, the mean absolute deviation (MAD) is reduced from 0.27 to 0.05 eV for the investigated cases.



Downloaded via BOCHUM LIBRARIES on February 3, 2022 at 07:36:52 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

1. INTRODUCTION

One of the everlasting problems of quantum chemistry is the large increase of computation time with system size, making the calculation of large biomolecules or complex systems very expensive or even impossible tasks. One way to decrease the computation time of complex systems are embedding methods, where the focus is set to only a part of the system and the influence of the environment onto this part is approximated. A popular class of embedding methods are QM/MM methods,[1−5] where the environment is described through point charges and force field methods. One representative is the polarizable embedding[6] (PE) method. Here, the polarization of the MM region is additionally taken into account through point dipole−dipole polarizabilities, fluctuating charges, or Drude oscillators.

A problem of QM/MM methods is that they miss purely quantum mechanical interactions without classical counterparts such as Pauli repulsion and nonadditive energy contributions that become important when the electron densities of the active part and the environment start to overlap. For this task, a quantum mechanical description of the environment is needed, also denoted as QM/QM embedding. One method belonging to this class is frozen density embedding (FDE).[7] In FDE, the total system is divided into small quantum mechanically described subsystems, and the interaction between them is expressed through density functional theory (DFT). While FDE was developed in the framework of DFT, it was combined with wave function-based methods already in the 1990s.[8,9] Recently, it has been extended to methods suitable for dynamic correlation and excited states.[10−14] A response theory formulation within FDE was first proposed for three overlapping partitions in 2004[15] and later extended to a general subsystem DFT response approach, also denoted as coupled FDE (FDEc).[16,17]

While the FDE approach is formally exact, in practice approximations similar to DFT must be made, including the use of approximate functionals in the interaction term, which for example are not able to capture long-range dispersion effects.[18] Additionally, the interaction term requires an explicit approximate semilocal kinetic-energy functional. The latter is an important repulsive contribution, in particular close to the nuclei of the environment where the strong Coulomb

Received: July 21, 2021

ACS Publications © XXXX The Authors. Published by American Chemical Society

attraction has to be compensated to prevent electron spill-out from the active subsystem into these regions.[19,20]

In the present work, furthermore the uncoupled (FDEu) scheme using the monomer basis is applied to avoid computational steps that involve the whole supermolecular system and thereby provides a route for calculations on much larger systems. While the coupled FDEc treatment enables one to treat excited states extending to environment subsystems, in the FDEu approximation used in the current work only excited states localized on one subsystem are accessible.

The missing Pauli repulsion originating from the orthogonality condition is with numerically accurate implementations usually not a problem for electronic ground states[21] but can still cause severe problems for excitation energies,[22,23] especially in combination with diffuse basis functions that are often essential for obtaining the excited state accurately. In such cases, it can change the energetic order of the excited states and induce low-lying artificial charge-transfer states, that is, an excited-state electron spill-out problem. One way to improve the repulsion is through embedding potential corrections such as long-distance corrections for the kinetic-energy component[24] or an energy correction[25] similar to the DFT-D[26,27] approach for dispersion. The orthogonality of the orbitals between the subsystems can be restored with projection-based embedding.[28−31] A first derivation of subsystem TDDFT with projection-based embedding has been reported in ref 32. However, in this work wrong assumptions were made, but refs 33−36 have shown that supermolecular excitation energies and properties can be restored if coupling matrix contributions are correctly taken into account. Because of the orthogonalization for the whole supersystem, however, these approaches come at a comparatively high computational cost. The description by the embedding potential can also be improved through the potential reconstruction scheme,[37] leading to improvements in TDDFT excitation energies of systems with hydrogen-bonding interactions.[38] However, the necessary supermolecular calculations make this method quite expensive for large environments.

We aim to address the problem of missing Pauli repulsion by reintroducing it into the embedding potential through effective core potentials, following the strategy of ref 39 used for PE, resulting in a significant improvement in the accuracy of excitation energies within the Algebraic Diagrammatic Construction through the second-order, ADC(2), model.[40−42] Effective core potentials (ECP) come with a small cost and are already implemented in many quantum chemical programs. Using ECPs to model chemically inactive electrons is a well-known approach and commonly used for heavy elements to replace inner electrons and to model the most important relativistic effects in otherwise nonrelativistic calculations.[43] In combination with embedding methods, ECPs have first been used to model cations in embedded cluster calculations to analyze defects in ionic crystals.[19,44] In particular, in the latter context, the idea of representing less important parts of the system by model potentials has also been extended to include valence electrons, for example, in the (fragment) ab initio model potential approach of Seijo and co-workers.[45,46] Similar ideas are also used to represent atoms or functional groups at the QM−MM boundary in QM/MM approaches.[47−50]

Because FDE does not use point charges and multipoles to approximate the embedding potential and includes repulsive contributions originating from the exchange and kinetic energy, missing intersystem Pauli repulsion is for ground-state FDE calculations less an issue[21] than for the QM/MM approaches mentioned above. Probably for this reason, ECPs have to the best of our knowledge until now not been used for the computation of excitation energies within the FDEu model, although the missing intersystem Pauli repulsion is much more important for excited states where energetically higher-lying orbitals are occupied, which are either themselves more diffuse than those occupied in the ground state or are energetically close to such diffuse orbitals. In the current work, we assess the performance of ECPs fitted to reproduce the virtual orbital energies of noble gases atoms for mitigating the problems from missing Pauli repulsion for FDEu at the level of ADC(2) as the correlated wave function method to compute excitation energies within the frozen solvent approximation,[51] that is, neglecting the response of the solvent. The latter allows one to compare the FDE results directly with those from ADC(2) calculations that use a supermolecular Hartree−Fock reference but allow only excitations out of orbitals located on the active subsystem. Besides the general performance of the approach, an important question will be whether including such ECPs simultaneously with the nonadditive kinetic and exchange energy contributions, that are part of the FDE scheme, will lead to double counting of repulsive interactions.

2. THEORY

Frozen density embedding (FDE) as used in combination with wave function methods goes back to subsystem DFT, in which the total density of a system is decomposed into the subsystem densities:

$$\rho(\mathbf{r}) = \sum_{i=1}^{m} \rho_i(\mathbf{r}) \tag{1}$$

where m denotes the number of subsystems, typically corresponding to the number of molecules. On the basis of this decomposition of the density, the energy of the total system is written as the sum of the energy contributions of the individual m subsystems plus one interaction contribution, expressed using the densities of the subsystems:

$$E_{\text{tot}} = \sum_{i=1}^{m} E_i[\rho_i(\mathbf{r})] + E_{\text{int}}[\rho_1(\mathbf{r}), ..., \rho_m(\mathbf{r})] \tag{2}$$

Formally, the interaction energy is defined therefore as the difference between the individual energies and the total energy:

$$E_{\text{int}}[\rho_1(\mathbf{r}), ..., \rho_m(\mathbf{r})] = E_{\text{tot}}[\rho_1(\mathbf{r}), ..., \rho_m(\mathbf{r})] - \sum_{i=1}^{m} E_i[\rho_i(\mathbf{r})] \tag{3}$$

The main advantage of the decomposition and sorting is rooted in the description of the interaction term as a functional of the subsystem densities, as the decomposition allows for employing different methods for the individual subsystems in general and the use of wave function methods in particular. If wave function methods are used, for example, for all subsystems, one can replace the density functional for the energy of these individual subsystems in eq 2 with a functional of the wave function. Note, however, that the interaction contribution is nevertheless expressed in terms of the subsystem densities; see eq 3:

https://doi.org/10.1021/acs.jctc.1c00732
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

$$E_{\text{tot}} = \sum_{i=1}^{m} E_i[\Psi_i(\mathbf{r})] + E_{\text{int}}[\rho_1(\mathbf{r}), ..., \rho_m(\mathbf{r})] \tag{4}$$

The motivation for this is that the interaction energy can then be expressed in terms of the densities, independent of the methods chosen for the individual subsystems:

$$\begin{aligned} E_{\text{int}}[\rho_1(\mathbf{r}), \rho_2(\mathbf{r}), ..., \rho_m(\mathbf{r})] = &-\sum_{i}^{m}\sum_{j\neq i}^{m} \int \rho_i(\mathbf{r}) \nu_j^{\text{N}}(\mathbf{r})\, \mathrm{d}\mathbf{r} \\ &+ \sum_{i}^{m}\sum_{j\neq i}^{m} \iint \frac{\rho_i(\mathbf{r})\rho_j(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\, \mathrm{d}\mathbf{r}\, \mathrm{d}\mathbf{r}' \\ &+ E_{\text{kxc}}^{\text{nadd}}[\rho_1(\mathbf{r}), ..., \rho_m(\mathbf{r})] + E_{\text{NN}} \end{aligned} \tag{5}$$

where ν_j^{N} is the potential of the nuclei of subsystem j. $E_{\text{kxc}}^{\text{nadd}}$ contains the nonadditive contributions of the kinetic and exchange-correlation energy and is given as

$$\begin{aligned} E_{\text{kxc}}^{\text{nadd}}[\rho_1(\mathbf{r}), ..., \rho_m(\mathbf{r})] = &\quad T[\rho_1(\mathbf{r}) + ... + \rho_m(\mathbf{r})] \\ &- \sum_{i}^{m} T^i[\rho_i(\mathbf{r})] + E_{\text{xc}}[\rho_1(\mathbf{r}) + ... + \rho_m(\mathbf{r})] \\ &- \sum_{i}^{m} E_{\text{xc}}^i[\rho_i(\mathbf{r})] \end{aligned} \tag{6}$$

Note that the kinetic contributions need to be computed using an explicit, that is, approximate, kinetic-energy functional. Deriving equations for the energy, or wave function, of a subsystem i leads to the conventional equations for the energy, or wave function, of a subsystem, which in the embedding case contains additional terms due to the interaction energy. These additional terms based on the interaction energy are commonly denoted as embedding potential and are given by

$$\nu_{\text{emb}}^i(\mathbf{r}) = -\sum_{j\neq i}^{m} \nu_j^{\text{N}}(\mathbf{r}) + \sum_{j\neq i}^{m} \int \frac{\rho_j(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\, \mathrm{d}\mathbf{r}' + \nu_{\text{kxc}}^i(\mathbf{r}) \tag{7}$$

where ν_{kxc}^i contains the nonadditive contributions obtained as the functional derivative:

$$\begin{aligned} \nu_{\text{kxc}}^i(\mathbf{r}) = &\left.\frac{\delta(T[\rho]+E_{\text{xc}}[\rho])}{\delta\rho}\right|_{\rho=\rho_{\text{tot}}} \\ &- \left.\frac{\delta(T[\rho]+E_{\text{xc}}[\rho])}{\delta\rho}\right|_{\rho=\rho_i} \end{aligned} \tag{8}$$

The embedding potential given in eq 7 is used in the present work in ground-state Hartree−Fock, ground-state second-order Møller−Plesset perturbation theory (MP2), as well as the response equations for the second-order adiabatic-diagrammatic construction (ADC(2)) scheme by including it in the Fock operator of subsystem i:

$$\hat{F}^{\text{FDE}} = \hat{F}^{\text{vac}} + \nu_{\text{emb}}^i(\mathbf{r}) \tag{9}$$

where $\hat{F}^{\text{vac}}$ is the conventional vacuum Fock operator. No additional changes to the code are needed for MP2 within the perturbation on energy[52] (PTE) reaction field scheme and for ADC(2) if used within the frozen solvent approximation.[51]

Intermolecular interactions can be decomposed into different contributions, see ref 53, of which one important contribution is the repulsive exchange (Pauli) repulsion between the electrons of two fragments having the same spin. Note, of course, that only the total intermolecular interaction is an observable, but the individual contributions are not. For example, if the exchange repulsion is not included as, for example, in mixed quantum-mechanics/molecular mechanics (QM/MM) methods, attractive contributions might dominate too strongly so that the intersystem distance of two fragments is too short. On the basis of this analysis, effective core potentials were introduced in polarizable embedding.[39]

In FDE, repulsive effects beyond the pure electrostatic contributions are included via the nonadditive exchange-correlation and kinetic terms contained in the embedding potential; see eq 7.[22] The literature reveals, however, that the contributions beyond the pure electrostatic embedding are not large enough in case of strong interactions; for example, in case of potential-energy surfaces, FDE is not repulsive enough.[25] Furthermore, for weakly interacting complexes, an "electron-leak" problem in FDE calculations occurs due to the strongly attractive nuclear potential, which is not compensated sufficiently by the kinetic functional.[24,54] This drawback might be rooted in the approximate kinetic-energy functionals.[20] As a solution, however, the use of effective core potentials was proposed.[19] To the best of our knowledge, ECPs were only pursued in FDE in terms of core orbitals relevant for electronic ground states.

In the present work, we focus on charge spill-out during electronic excitations. If diffuse basis functions are included, which is often necessary for accurate excitation energies, electronic excitations might lead to a charge redistribution such that electron density is moved to environment regions that exhibit a strong nuclear attraction where the repulsion is not sufficiently accounted for by the potential ν_{kxc}^i. The resulting unphysical attraction is significantly reduced when all-electron ECPs are introduced to model the Pauli repulsion, which leads to an improved description of the interaction in particular in regions that are close to nuclei. We therefore add the potential from all-electron ECPs on all atoms in the environment $\hat{\nu}_{\text{ECP}}^i$ to the FDE Fock operator of subsystem i, leading to

$$\hat{F}^{\text{FDE(ECP)}} = \hat{F}^{\text{FDE}} + \hat{\nu}_{\text{ECP}}^i \tag{10}$$

In difference to the local potential $\nu_{\text{kxc}}^i(\mathbf{r})$, the ECP potential $\hat{\nu}_{\text{ECP}}^i$ is nonlocal as it employs projection operators to act on specific parts of the one-electron Hilbert space. This facilitates the modeling of repulsion effects from the Pauli exclusion principle. Using all-electron pseudopotentials instead of effective core potentials that model only core orbitals leads to a description of the Pauli repulsion with all electrons of the environment and not just those near the nuclei. This is very important because the nonbonded interactions between subsystems are expected to happen primarily between valence electrons. A simple set of all-electron pseudopotentials that reproduces the virtual orbital energies of the noble gases and can be used for the elements in the same period has been developed for He, Ne, and Ar in ref 39 and for Kr and Xe in the current work (vide infra). These all-electron pseudopotentials are expected to model approximately the interaction with the valence region of an atom in a molecule if it can be assumed that (for main group elements) the octet rule is fulfilled, which should be the case if the environment has no low-lying electronically excited states. For a given angular momentum, these pseudopotentials include all components and not only selected contributions. Therefore, they are no

https://doi.org/10.1021/acs.jctc.1c00732
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

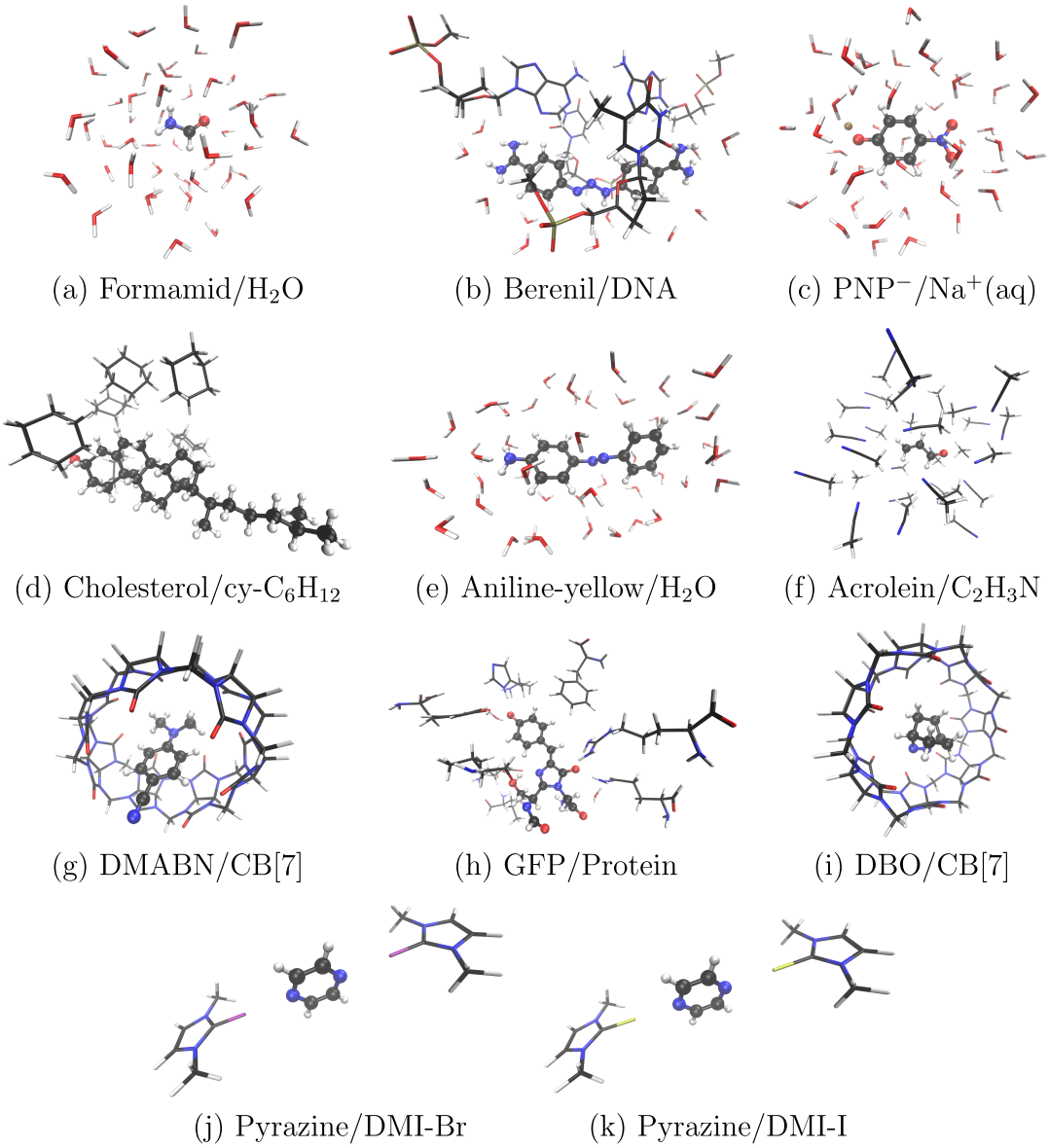


Figure 1. Active/environment total structures of the systems used as a test set. The analyzed chromophore is displayed with balls and sticks, while the environment is displayed as licorice. Systems (a)−(i) are taken from ref 39, while systems (j) and (k) have been constructed in the present work.

longer effective core potentials, but can still be calculated with the available ECP subroutines without further manipulation. To highlight the nature of their computation, we will continue to use the shorthand notation ECP, even though they refer to more than just the core orbitals.

In the FDE approach, the densities of all subsystems are kept frozen except for the active subsystem whose density or wave function is relaxed in the presence of the frozen environment. Subsequently, the role of the active subsystem is shifted to the next subsystem and its density or wave function is optimized accordingly. This is repeated for every subsystem. One cycle in which all subsystems had once the role of the active system is denoted one "Freeze-and-Thaw" (FaT) cycle. The FaT cycles can be repeated until self-consistent convergence for the total system is obtained.

This leads to the following computational protocol for a FDE procedure with m subsystems:

(1) Optimize isolated subsystem densities and save these densities and their electrostatic potentials.

(2) Start the current Freeze-and-Thaw (FaT) cycle using i = 1:

(a) Freeze all subsystems except subsystem i.

(b) Relax the density of the active subsystem i in the presence of the relaxed embedding potential.

(c) Save the density and electrostatic potential of subsytem i.

(d) If $i = m$ exit; otherwise, set $i = i + 1$ and go to step (2a).

(3) If the solution is (within a preset threshold) self-consistent, proceed with step (4); otherwise go to step (2).

(4) Compute the target properties, for example, ADC(2) excitation energies, of the desired subsystems.

While in supermolecular calculations electronic excitations are possibly delocalized over many atoms, FDEu constrains electronic excitations by construction to one subsystem. This

https://doi.org/10.1021/acs.jctc.1c00732
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

qualitative difference leads to quantitative differences when the comparison of FDE and supermolecular reference results is sought. To provide reference results that are obtained in a spirit similar to FDE, we have implemented the frozen solvent ADC(2) scheme, employing the PNO-ADC(2) code,[55] which uses localized molecular orbitals (LMOs) for the occupied space. In this scheme, the ground-state first-order amplitudes and the excited-state response vectors are restricted to single and double excitations out of LMOs localized on those atoms belonging to the target subsystem in the FDE calculation. In the following, this scheme will be denoted as subsystem-ADC(2).

3. COMPUTATIONAL DETAILS

All calculations in this work were carried out with a local development version of the TURBOMOLE[56] program package. The FDE calculations were performed in the monomer basis of each subsystem. For the embedding potential, the PBE[57] exchange-correlation and the LC94[58] kinetic functional were used together with a supersystem integration grid created from the atomic integration grids of all atoms using grid size 3. We continued with FaT cycles until the electronic energies of all subsystems were converged within a threshold of 5.0×10^{-6} H. Subsequently, excitation energies were computed using the ADC(2) method in the frozen core approximation and the frozen solvent scheme,[51] that is, skipping the response of the environment.

For the atoms H−Ar, we used the all-electron pseudopotentials for He, Ne, and Ar reported in ref 39. For iodine and bromine, we used the all-electron pseudopotentials optimized for krypton and xenon as part of the present work. Details can be found in the Supporting Information. FDE calculations with the pseudopotentials are denoted as FDE(ECP) in the following.

Reference results are computed using the subsystem-ADC(2) approach with a tight PNO truncation threshold of 10^{-9}, using IBO localization. All ADC(2) calculations employed the frozen core approximation for all orbitals with energy below −3 hartree.

3.1. Ethylene−Tetrafluoroethylene Dimer. The geometries of the two molecules ethylene and tetrafluoroethylene (TFE) have been optimized independently of each other with DFT/B3LYP[59] employing D_{2h} symmetry each. These two monomers were then placed coplanar on top of each other, for which the lowest 10 excitation energies have been computed using ADC(2) for distances between 5.5 and 18 bohr. For this system, we used the d-aug-cc-pVTZ basis set,[60] and the densities in the FDE calculations were computed at the Hartree−Fock level of theory.

3.2. Embedding Test Set. For a more representative analysis of the FDE(ECP) approach for excitation energies, we employ a test set consisting of 11 systems; see Figure 1. The structures of systems (a)−(i) were taken from ref 39. These systems contain chromophores in different chemical environments such as water, organic solvents, biomolecular matrixes, and molecular container. They have furthermore been shown to be sensitive to electron spill-out in QM/MM excitation studies. To this set were added systems (j) and (k) to include test cases for bromine and iodine as examples for atoms from periods 4 and 5. The structures for these two systems were optimized using DFT/B3LYP[59] with a grid size of 3 and the cc-pVTZ[61] basis set for the pyrazine subsystem and cc-pVDZ for its environment. This test set allows one to investigate the robustness of the FDE method against electron spill-out and the efficiency of the FDE(ECP) approach in various environments.

The first two vertical excitation energies of the chromophores in the presence of the environment have been computed using the Dunning's cc-pVTZ and aug-cc-pVTZ[62] basis sets for the chromophore and the cc-pVDZ basis set for the environment. For FDE calculations, the HF density was used for the chromophore and the DFT/PBE density employing grid 3 for the environment subsystems. System (h) and (i) have also been computed employing the revapbek[63] and Thomas−Fermi kinetic functional[64,65] for the embedding potential. The results can be found in the Supporting Information. Because the difference in excitation energies between the investigated kinetic energies is comparably smaller than the difference from the reference, they are not discussed further. As in ref 39, for system (h), only the first excitation energy is included.

The potentials for the PE and PE(ECP) calculations were taken for systems (a)−(i) from ref 39. For system (j) and (k), we used a IAO-D3 potential; that is, multipole moments are obtained from an IAO analysis of each subsystem, and the polarizabilities are taken from the DFT-D3 correction parameters.[27] Excitation energies were computed neglecting the response of the environment and using the frozen core approximation.

Subsystem-ADC(2) excitation energies of systems (h) and (i) were also computed with different localization methods, for example, the Boys[66,67] method, the second moment[68] method with an exponent of 2, and the fourth moment[69] method; see the Supporting Information. The different localization methods result in only insignificant changes in the excitation energies and will not be discussed further.

4. RESULTS

4.1. Ethylene−TFE Dimer. We have used the ethylene−TFE dimer to study the effect of different contributions from the intermolecular interaction on excitation energies as a function of the interaction strength, which for this system can easily be tuned by changing the distance between the molecules. The different effects from the ground-state potential (without London dispersion) and the remaining contributions as excitonic coupling and excited-state response on the interaction-induced shift of the excitation energies can be seen in Figure 2, which shows the first eight excitation energies of the ethylene−TFE dimer from supermolecular and subsystem-ADC(2) calculations. One can see that at large distances for all eight excitations the results obtained with the two approaches converge to the same values, as it should be as long as none of the excited states is a long-range charge-transfer state that would be missing in the subsystem-ADC(2) calculations. If the two molecules get closer, the differences between both approaches increase. Major differences are in particular seen for the higher lying states and at short distances (below about 7.086 bohr). A big effect can be seen in the third excitation energy where both approaches end up with a difference of 1.52 eV at a distance of 5.5 bohr. This illustrates that supermolecular calculations are not a helpful reference for validating the accuracy of individual contributions of uncoupled embedding approaches for excitation energies.

Figure 3a and b shows the first four excitation energies of, respectively, the ethylene and TFE subsystems within the ethylene−TFE dimer employing FDE, FDE(ECP), and

https://doi.org/10.1021/acs.jctc.1c00732
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

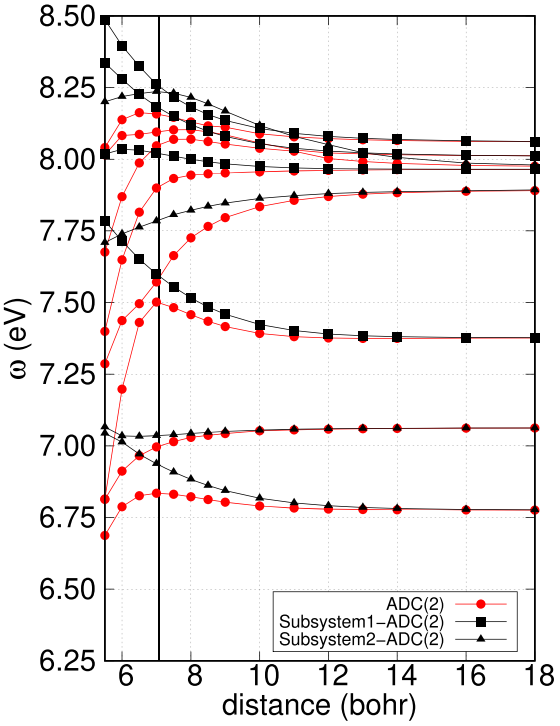


Figure 2. First eight excitation energies of the ethylene−TFE system with respect to the distance between ethylene and TFE for a conventional calculation as compared to the corresponding excitations with the subsystem-ADC(2) approach, both employing a supermolecular basis. A vertical bar is shown at 7.086 bohr.

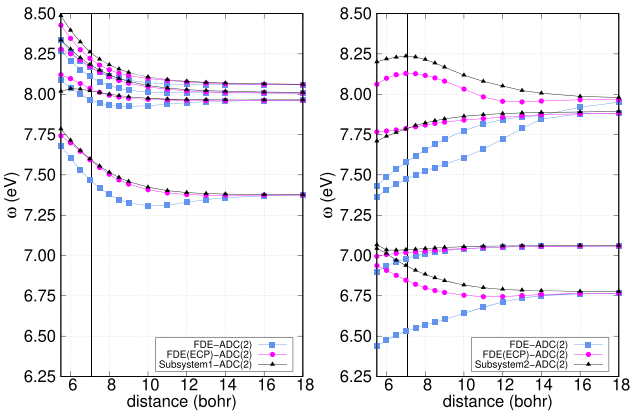


Figure 3. Excitation energies of the subsystems within the ethylene−TFE system with respect to the distance between both subsystems for FDE, FDE(ECP), and subsystem-ADC(2). A vertical bar is shown at 7.086 bohr.

subsystem-ADC(2) as the reference. This figure shows that for long distances the FDE results converge for all states to the subsystem-ADC(2) result, corresponding to a vanishing embedding potential without artifacts in the long-range region.

For most excitations on ethylene, see Figure 3a, the FDE curves are in good agreement with those from subsystem-ADC(2), including intermediate and short distances. However, for the first and the fifth excited states, FDE yields too low of excitation energies for intermediate distances at about 10−12 bohr (curves with boxes). Adding the pseudopotentials to the embedding potential, that is, using FDE(ECP), leads to major improvements for the problematic cases (curves with circles), although FDE(ECP) is not able to provide in all cases accurate shifts for short distances below 7.086 bohr in all cases. In these cases, exchange effects that cannot be captured in the FDE scheme and modeled only very roughly by the ECPs are most likely the reason. The significant improvements when going from FDE to FDE(ECP) support the analysis that the strong attraction is not completely compensated by the kinetic energy functional and the description can be improved using simple pseudopotentials, although at short distances some deviations with respect to the reference subsystem-ADC(2) show up due to remaining errors in the description of Pauli repulsion and exchange effects.

In case of the TFE subsystem of the ethylene−TFE dimer, Figure 3b reveals even more severe problems for the conventional FDE approach. For most of the excitations, significant deviations are obtained between FDE (curve with boxes) and subsystem-ADC(2) (curve with triangles) at distances below 10−12 bohr. For all excited states in this case, excitation energies obtained with the conventional FDE approach decrease with shorter distances, while for some states the subsystem-ADC(2) reference results in an increase, yielding a qualitatively wrong behavior with conventional FDE. To give an example, in case of the fourth excitation, the FDE curve (circles) deviates already clearly from the reference value at a distance as large as 16 bohr. Adding the pseudopotentials to the embedding potential improves the excitation energies significantly. The results are not as close to the subsystem-ADC(2) reference as for the ethylene subsystem but reproduce in all cases the correct behavior of the excitation energies as a function of the distance between the two molecules.

To summarize, the results for the ethylene−TFE dimer reveal unambiguously that simple pseudopotentials increase the accuracy of the numerical outcome of FDE calculations for excitation energies significantly, especially for shorter distances, that is, stronger interactions. In particular, qualitatively wrong curves are corrected to yield a qualitatively correct description.

4.2. Embedding Test Set. Having assessed the accuracy of the FDE(ECP) method as the example of a model dimer system, we now turn to larger systems collected in the test set; see Figure 1. In the left graph of Figure 4, the deviations of the excitation energies ω obtained with FDE-ADC(2) from the subsystem-ADC(2) reference results are shown for the cc-pVTZ and aug-cc-pVTZ basis sets. The corresponding deviations for FDE(ECP)-ADC(2) are shown in the right graph of Figure 4; see also Table 1.

Figure 4 reveals that in case of the cc-pVTZ basis, which does not contain diffuse functions, the deviations are in most cases rather small. The largest deviation is 0.20 eV, and the mean absolute deviation (MAD) is as small as 0.05 eV. Introducing ECPs, the largest deviation decreases slightly to 0.19 eV, and the MAD increases to 0.06 eV. One can conclude that, in this case, the basis set without diffuse functions is not able to allow for excitations into the problematic regions of the environment so that ECPs are not required.

For the aug-cc-pVTZ basis, which contains diffuse functions, Figure 4 reveals an increased MAD of 0.27 eV and a maximum error as large as 1.84 eV. It should be pointed out that the FDE-ADC(2) method consistently underestimates all vertical transition energies resulting in a mean signed deviation (MSD) of −0.27 eV. This indicates an electron spill-out in these systems in case of conventional FDE-ADC(2). Introducing

https://doi.org/10.1021/acs.jctc.1c00732
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

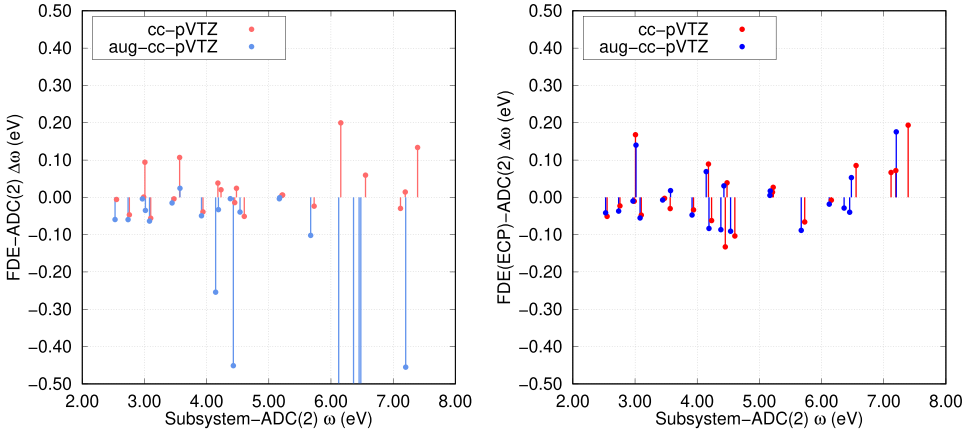


Figure 4. Deviations of the FDE-ADC(2) (left) and FDE(ECP)-ADC(2) (right) excitation energies ($\Delta\omega$) with respect to subsystem-ADC(2) excitation energies ω. Statistics for these deviations are collected in Table 1.

Table 1. Mean Signed Deviation (MSD), Mean Absolute Deviation (MAD), and Maximum Deviation (MAX) for FDE and FDE(ECP) with Respect to Subsystem-ADC(2)[a]

basis	method	MSD	MAD	MAX
cc-pVTZ	FDE-ADC(2)	+0.02	0.05	+0.20
	FDE(ECP)-ADC(2)	< +0.01	0.06	+0.19
aug-cc-pVTZ	FDE-ADC(2)	−0.27	0.27	−1.84
	FDE(ECP)-ADC(2)	> −0.01	0.05	+0.18

[a]See Figure 4. All values are in electronvolts.

ECPs, the largest deviation decreases significantly to 0.18 eV and the MAD to 0.05 eV. The magnitude of the MSD is now reduced below 0.01 eV, showing that the excitation energies are both underestimated and overestimated to a similar amount.

The impact of the ECPs on the physics of the transitions can be illustrated with natural transition orbitals (NTOs). As an example, the occupied and virtual NTOs of the second excitation of system (i) are displayed in Figure 5. The corresponding excitation energy exhibits an error of −1.42 eV in case of FDE-ADC(2) and is reduced to −0.02 eV in case of FDE(ECP)-ADC(2), making it the excitation with the largest influence of the ECPs in our test set. Figure 5 shows that the occupied NTOs differ without and with ECPs and only with ECPs is the correct NTO described, but in both cases they are well localized on the target chromophore. In case of the virtual NTOs, however, the huge influence of the ECPs is revealed. Using FDE-ADC(2) leads to a virtual NTO that is not located on the chromophore but on the environment. By virtue of the ECPs in FDE(ECP)-ADC(2), the virtual NTO is fully (re)located on the chromophore.

To analyze the influence of the ECPs upon the occupied and virtual space, one can compare the frontier orbital energies, that is, the energy of the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO), of FDE-HF and FDE(ECP)-HF. The comparison, see the Supporting Information for details, shows that in most

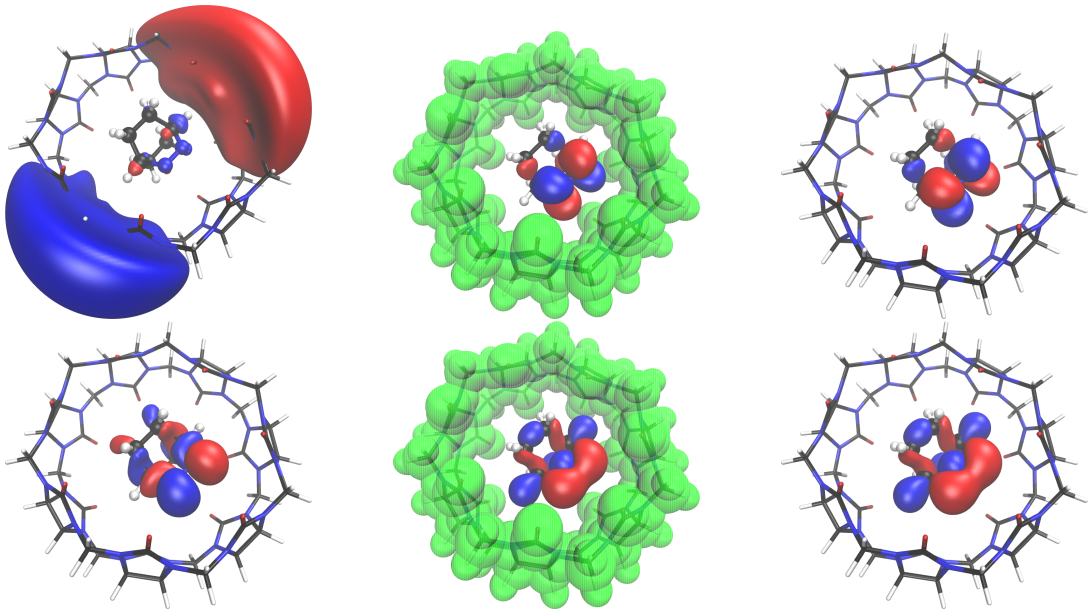


Figure 5. Natural-transition orbitals (NTOs) of the second vertical excitation of system (i). Top, virtual NTOs; bottom, occupied NTOs. Left, FDE method without ECPs; middle, FDE(ECP) method with ECPs displayed in green; right, subsystem-ADC(2).

https://doi.org/10.1021/acs.jctc.1c00732
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

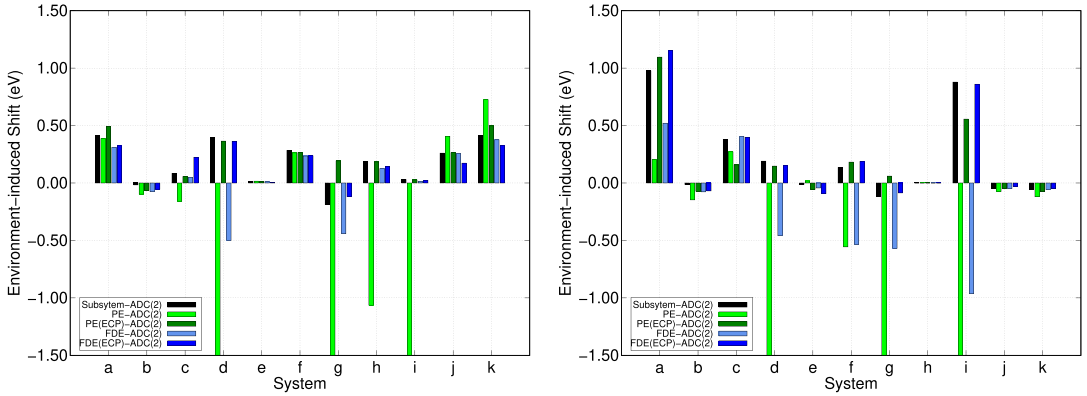


Figure 6. Environment-induced shifts of the first (left) and second (right) vertical excitation energies employing the aug-cc-pVTZ basis set. The PE-ADC(2) and PE(ECP)-ADC(2) excitation energies of systems (a)−(i) are taken from ref 39.

cases the orbital energies increase due to ECPs, which is to be expected due to the additional repulsive term. Only in one case of the test set (system (b)) are lower orbital energies obtained. However, in most cases, the increase in orbital energy is pronounced for the LUMO and in case of diffuse basis functions, while there are only a few cases in which the energy of the HOMO is increased more than the energy of the LUMO.

To give an example, the largest difference between the FDE-ADC(2) and FDE(ECP)-ADC(2) excitation energies occurs in case of system (i). In this case, the HOMO energy is in the cc-pVDZ basis set increased by 47 meV and the LUMO energy by 125 meV, while in the aug-cc-pVTZ, which contains diffuse functions, the HOMO energy is increased by 44 meV and that of the LUMO by 3291 meV.

Figure 6 shows a comparison between the frozen density embedding and the polarizable embedding (both within the frozen solvent approximation) for the environment-induced shifts in the excitation energies in the aug-cc-pVTZ basis. Comparing FDE-ADC(2) and PE-ADC(2), that is, embedding methods without pseudopotentials, reveals that for FDE the electron spill-out problem is not as large as for PE. It is even removed almost entirely for the first excitations of system (i) and (h), but still for FDE many shifts show the wrong sign, revealing that the nonadditive kinetic energy term is in many cases not repulsive enough to repel leaking electrons. However, both embedding methods employing ECPs, that is, FDE(ECP) and PE(ECP), are similarly accurate for the cases investigated in the present work. Employing the same ECPs, FDE(ECP) differs from PE(ECP) in the frozen solvent approximation mainly by the additional nonadditive functional contributions. If ECPs and the nonadditive kinetic contribution lead to a double counting of repulsive effects, an increase of the deviations from the reference values from PE(ECP) to FDE(ECP) would result, but this is not observed in the present work. This leads us to the conclusion that the Pauli repulsion originating from the orthogonality of MOs is described exclusively by the ECPs, which via projection operators shift that part of the Hilbert space that is (in the total supersystem) occupied by the electrons of the environment energetically up so that they are no longer accessible for the electrons of the active subsystem. On the other hand, there is no indication for a significant overestimation of repulsive exchange and kinetic energy contributions for those parts of the Hilbert space not occupied by the electrons of the environment by including both the local ν^i_{kxc} and the nonlocal $\hat{\nu}^i_{\mathrm{ECP}}$ potentials, in the Hamiltonian and Fock operators for the active subsystem.

5. CONCLUSION

In frozen density embedding, available kinetic energy functionals are often not repulsive enough to compensate the nuclear attraction, in particular in regions close to the nuclei. While in ground-state calculations, the missing repulsion does usually not cause problems, it becomes more visible for excited states if the basis set contains diffuse atomic orbitals that are often essential to obtain accurate excitation energies. In the present work, we have proposed an FDE(ECP) scheme in which effective core potentials are included in the embedding potential to account for the missing repulsion. The ECPs do not lead to a significant change in the occupied orbitals but have a significant impact mainly on the virtual orbitals, which are repelled by the ECPs from those parts of the Hilbert space that would be occupied by the electrons of the environment.

The calculations for the ethylene−TFE dimer at different distances revealed that conventional FDE can lead to a qualitatively wrong behavior for several excited states. Employing ECPs in the FDE(ECP) approach leads to a significant improvement, and the qualitative trend of all excited states is reproduced.

For a test set consisting of 11 systems, which are sensitive to electron spill-out, the mean absolute deviation (MAD) for the results in the aug-cc-pVTZ basis set is reduced from 0.27 to 0.05 eV and the maximum error from 1.84 to 0.18 eV when using FDE(ECP) instead of FDE.

The ECPs for bromine and iodine reported in the present work are designed in such a way that they can be used for all p-block elements within the same period. Further improvements could possibly be obtained by using additional functions and parameters in the ECP or by fitting the potentials individually for each element within a period.

Overall, the results from the present work demonstrate that FDE provides a somewhat more robust approximation for the embedding potential with a reduced unphysical electron spill-out than a PE approach with a multipolar representation of the environment's charge density. Also, FDE does not require the parametrization of static and induced multipole moments, which can be another limiting factor for the accuracy of PE models, and it accounts, in difference to the latter, also for nonadditive kinetic, exchange, and correlation contributions. However, the remaining errors in FDE due to the approximate

https://doi.org/10.1021/acs.jctc.1c00732
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

account of repulsive contributions by the kinetic energy functionals can still be sizable and can limit the accuracy of the excitation energies, in particular for diffuse basis functions. Incorporating in the FDE calculations the same simple ECPs as proposed recently for PE leads to significantly improved excitation energies. A double counting of repulsive contributions that could be anticipated when ECPs are included, in addition to the nonadditive kinetic energy contributions, is not observed.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.1c00732.

ECP parameters, orbital energy differences between FDE and FDE(ECP), calculated excitation energies, comparison of localization, comparison of kinetic functionals, and comparison with long distance correction (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Robert S. Treß − *Department of Theoretical Chemistry, Ruhr University Bochum, Bochum 44801, Germany;* orcid.org/0000-0001-8804-3547; Email: robert.tress@rub.de

Sebastian Höfener − *Institute of Physical Chemistry, Karlsruhe Institute of Technology (KIT), Karlsruhe 76131, Germany;* orcid.org/0000-0003-4504-347X; Email: hoefener@kit.edu

Author

Christof Hättig − *Department of Theoretical Chemistry, Ruhr University Bochum, Bochum 44801, Germany;* orcid.org/0000-0002-5752-2710

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.1c00732

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank TURBOMOLE GmbH for funding. C.H. acknowledges support by the Deutsche Forschungsgemeinschaft with the Priority Program SPP 1807 (HA 25888/10-1). S.H. acknowledges support by the Deutsche Forschungsgemeinschaft through SFB 1249 "N-Heteropolycycles as Functional Materials" (project B07).

■ REFERENCES

(1) Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976**, *103*, 227−249.

(2) Senn, H. M.; Thiel, W. QM/MM methods for biomolecular systems. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198−1229.

(3) Brunk, E.; Rothlisberger, U. Mixed quantum mechanical/molecular mechanical molecular dynamics simulations of biological systems in ground and electronically excited states. *Chem. Rev.* **2015**, *115*, 6217−6263.

(4) List, N. H.; Olsen, J. M. H.; Kongsted, J. Excited states in large molecular systems through polarizable embedding. *Phys. Chem. Chem. Phys.* **2016**, *18*, 20234−20250.

(5) Morzan, U. N.; Alonso de Armino, D. J.; Foglia, N. O.; Ramirez, F.; Gonzalez Lebrero, M. C.; Scherlis, D. A.; Estrin, D. A. Spectroscopy in complex environments from QM-MM simulations. *Chem. Rev.* **2018**, *118*, 4071−4113.

(6) Olsen, J. M.; Aidas, K.; Kongsted, J. Excited states in solution through polarizable embedding. *J. Chem. Theory Comput.* **2010**, *6*, 3721−3734.

(7) Wesolowski, T. A.; Warshel, A. Frozen density functional approach for ab initio calculations of solvated molecules. *J. Phys. Chem.* **1993**, *97*, 8050−8053.

(8) Govind, N.; Wang, Y.; Da Silva, A.; Carter, E. Accurate ab initio energetics of extended systems via explicit correlation embedded in a density functional environment. *Chem. Phys. Lett.* **1998**, *295*, 129−134.

(9) Govind, N.; Wang, Y. A.; Carter, E. A. Electronic-structure calculations by first-principles density-based embedding of explicitly correlated systems. *J. Chem. Phys.* **1999**, *110*, 7677−7688.

(10) Gomes, A. S. P.; Jacob, C. R.; Visscher, L. Calculation of local excitations in large systems by embedding wave-function theory in density-functional theory. *Phys. Chem. Chem. Phys.* **2008**, *10*, 5353−5362.

(11) Höfener, S.; Severo Pereira Gomes, A.; Visscher, L. Molecular properties via a subsystem density functional theory formulation: A common framework for electronic embedding. *J. Chem. Phys.* **2012**, *136*, 044104.

(12) Prager, S.; Zech, A.; Aquilante, F.; Dreuw, A.; Wesolowski, T. A. First time combination of frozen density embedding theory with the algebraic diagrammatic construction scheme for the polarization propagator of second order. *J. Chem. Theory Comput.* **2016**, *144*, 204103.

(13) Prager, S.; Zech, A.; Wesolowski, T. A.; Dreuw, A. Implementation and application of the frozen density embedding theory with the algebraic diagrammatic construction scheme for the polarization propagator up to third order. *J. Chem. Theory Comput.* **2017**, *13*, 4711−4725.

(14) Daday, C.; König, C.; Neugebauer, J.; Filippi, C. Wavefunction in density functional theory embedding for excited states: Which wavefunctions, which densities? *ChemPhysChem* **2014**, *15*, 3205−3217.

(15) Casida, M. E.; Wesolowski, T. A. Generalization of the Kohn-Sham equations with constrained electron density formalism and its time-dependent response theory formulation. *Int. J. Quantum Chem.* **2004**, *96*, 577−588.

(16) Neugebauer, J. Couplings between electronic transitions in a subsystem formulation of time-dependent density functional theory. *J. Chem. Phys.* **2007**, *126*, 134116.

(17) Neugebauer, J. On the calculation of general response properties in subsystem density functional theory. *J. Chem. Phys.* **2009**, *131*, 084104.

(18) Kristyán, S.; Pulay, P. Can (semi) local density functional theory account for the London dispersion forces? *Chem. Phys. Lett.* **1994**, *229*, 175−180.

(19) Stefanovich, E. V.; Truong, T. N. Embedded density functional approach for calculations of adsorption on ionic crystals. *J. Chem. Phys.* **1996**, *104*, 2946−2955.

(20) Jacob, C. R.; Wesolowski, T. A.; Visscher, L. Orbital-free embedding applied to the calculation of induced dipole moments in CO2···X (X = He, Ne, Ar, Kr, Xe, Hg) van der Waals complexes. *J. Chem. Phys.* **2005**, *123*, 174104.

(21) Dułak, M.; Wesolowski, T. A. On the electron leak problem in orbital-free embedding calculations. *J. Chem. Phys.* **2006**, *124*, 164101.

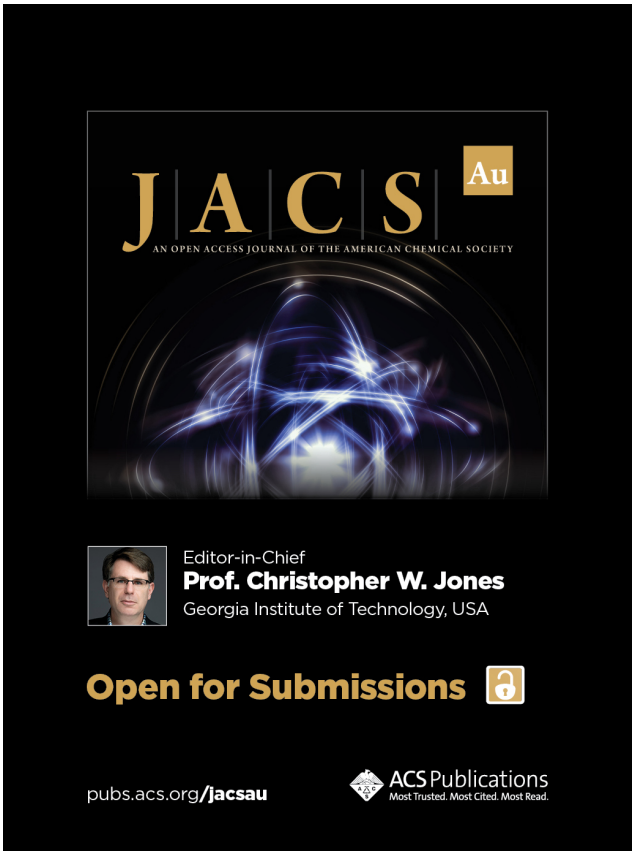
(22) Fradelos, G.; Wesolowski, T. A. Importance of the Intermolecular Pauli Repulsion in Embedding Calculations for Molecular Properties: The Case of Excitation Energies for a Chromophore in Hydrogen-Bonded Environments. *J. Phys. Chem. A* **2011**, *115*, 10018−10026.

(23) Zech, A.; Ricardi, N.; Prager, S.; Dreuw, A.; Wesolowski, T. A. Benchmark of excitation energy shifts from frozen-density embedding theory: introduction of a density-overlap-based applicability threshold. *J. Chem. Theory Comput.* **2018**, *14*, 4028−4040.

(24) Jacob, C. R.; Beyhan, S. M.; Visscher, L. Exact functional derivative of the nonadditive kinetic-energy bifunctional in the long-distance limit. *J. Chem. Phys.* **2007**, *126*, 234116.
(25) Schlüns, D.; Klahr, K.; Mück-Lichtenfeld, C.; Visscher, L.; Neugebauer, J. Subsystem-DFT potential-energy curves for weakly interacting systems. *Phys. Chem. Chem. Phys.* **2015**, *17*, 14323−14341.
(26) Grimme, S. Accurate description of van der Waals complexes by density functional theory including empirical corrections. *J. Comput. Chem.* **2004**, *25*, 1463−1473.
(27) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
(28) Chulhai, D. V.; Jensen, L. Frozen Density Embedding with External Orthogonality in Delocalized Covalent Systems. *J. Chem. Theory Comput.* **2015**, *11*, 3080−3088.
(29) Manby, F. R.; Stella, M.; Goodpaster, J. D.; Miller, T. F., III A simple, exact density-functional-theory embedding scheme. *J. Chem. Theory Comput.* **2012**, *8*, 2564−2568.
(30) Goodpaster, J. D.; Barnes, T. A.; Manby, F. R.; Miller, T. F., III Accurate and systematically improvable density functional theory embedding for correlated wavefunctions. *J. Chem. Phys.* **2014**, *140*, 18A507.
(31) Barnes, T. A.; Goodpaster, J. D.; Manby, F. R.; Miller, T. F., III Accurate basis set truncation for wavefunction embedding. *J. Chem. Phys.* **2013**, *139*, 024103.
(32) Chulhai, D. V.; Jensen, L. External orthogonality in subsystem time-dependent density functional theory. *Phys. Chem. Chem. Phys.* **2016**, *18*, 21032−21039.
(33) Tölle, J.; Böckers, M.; Neugebauer, J. Exact subsystem time-dependent density-functional theory. *J. Chem. Phys.* **2019**, *150*, 181101.
(34) Niemeyer, N.; Tölle, J.; Neugebauer, J. Approximate versus Exact Embedding for Chiroptical Properties: Reconsidering Failures in Potential and Response. *J. Chem. Theory Comput.* **2020**, *16*, 3104−3120.
(35) Scholz, L.; Tölle, J.; Neugebauer, J. Analysis of environment response effects on excitation energies within subsystem-based time-dependent density-functional theory. *Int. J. Quantum Chem.* **2020**, *120*, e26213.
(36) Tölle, J.; Böckers, M.; Niemeyer, N.; Neugebauer, J. Inter-subsystem charge-transfer excitations in exact subsystem time-dependent density-functional theory. *J. Chem. Phys.* **2019**, *151*, 174109.
(37) Fux, S.; Jacob, C. R.; Neugebauer, J.; Visscher, L.; Reiher, M. Accurate frozen-density embedding potentials as a first step towards a subsystem description of covalent bonds. *J. Chem. Phys.* **2010**, *132*, 164101.
(38) Artiukhin, D. G.; Jacob, C. R.; Neugebauer, J. Excitation energies from frozen-density embedding with accurate embedding potentials. *J. Chem. Phys.* **2015**, *142*, 234101.
(39) Marefat Khah, A.; Reinholdt, P.; Olsen, J. M. H.; Kongsted, J.; Hättig, C. Avoiding electron spill-out in QM/MM calculations on excited states with simple pseudopotentials. *J. Chem. Theory Comput.* **2020**, *16*, 1373−1381.
(40) Schirmer, J. Beyond the random-phase approximation: A new approximation scheme for the polarization propagator. *Phys. Rev. A* **1982**, *26*, 2395.
(41) Trofimov, A.; Schirmer, J. An efficient polarization propagator approach to valence electron excitation spectra. *J. Phys. B: At. Mol. Opt. Phys.* **1995**, *28*, 2299.
(42) Schirmer, J.; Trofimov, A. Intermediate state representation approach to physical properties of electronically excited molecules. *J. Chem. Phys.* **2004**, *120*, 11449−11464.
(43) Dolg, M.; Cao, X. Relativistic pseudopotentials: their development and scope of applications. *Chem. Rev.* **2012**, *112*, 403−480.
(44) Luaña, V.; Pueyo, L. Simulation of ionic crystals: The ab initio perturbed-ion method and application to alkali hydrides and halides. *Phys. Rev. B* **1990**, *41*, 3800.
(45) Barandiarán, Z.; Seijo, L. The ab initio model potential representation of the crystalline environment. theoretical study of the local distortion on NaCl: Cu+. *J. Chem. Phys.* **1988**, *89*, 5739−5746.
(46) Swerts, B.; Chibotaru, L. F.; Lindh, R.; Seijo, L.; Barandiaran, Z.; Clima, S.; Pierloot, K.; Hendrickx, M. F. Embedding fragment ab initio model potentials in CASSCF/CASPT2 calculations of doped solids: implementation and applications. *J. Chem. Theory Comput.* **2008**, *4*, 586−594.
(47) von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. Variational optimization of effective atom centered potentials for molecular properties. *J. Chem. Phys.* **2005**, *122*, 014113.
(48) Zhang, Y.; Lee, T.-S.; Yang, W. A pseudobond approach to combining quantum mechanical and molecular mechanical methods. *J. Chem. Phys.* **1999**, *110*, 46−54.
(49) Slavíček, P.; Martínez, T. J. Multicentered valence electron effective potentials: A solution to the link atom problem for ground and excited electronic states. *J. Chem. Phys.* **2006**, *124*, 084107.
(50) DiLabio, G. A.; Hurley, M. M.; Christiansen, P. A. Simple one-electron quantum capping potentials for use in hybrid QM/MM studies of biological molecules. *J. Chem. Phys.* **2002**, *116*, 9578−9584.
(51) Corni, S.; Cammi, R.; Mennucci, B.; Tomasi, J. Electronic excitation energies of molecules in solution within continuum solvation models: Investigating the discrepancy between state-specific and linear-response methods. *J. Chem. Phys.* **2005**, *123*, 134512.
(52) Valle, F. J. O. D.; Tomasi, J. Electron correlation and solvation effects. I. Basic formulation and preliminary attempt to include the electron correlation in the quantum mechanical polarizable continuum model so as to study solvation phenomena. *Chem. Phys.* **1991**, *150*, 139−150.
(53) Hopffgarten, M. v.; Frenking, G. Energy decomposition analysis. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 43−62.
(54) Jacob, C. R.; Neugebauer, J. Subsystem density-functional theory. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2014**, *4*, 325−362.
(55) Helmich, B.; Hättig, C. A pair natural orbital implementation of the coupled cluster model CC2 for excitation energies. *J. Chem. Phys.* **2013**, *139*, 084114.
(56) Development version of TURBOMOLE 2019, a development of the University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989−2007, TURBOMOLE GmbH, since 2007; available from http://www.turbomole.com (accessed Apr 23, 2021).
(57) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865.
(58) Lembarki, A.; Chermette, H. Obtaining a gradient-corrected kinetic-energy functional from the Perdew-Wang exchange functional. *Phys. Rev. A* **1994**, *50*, 5328.
(59) Becke, A. D. Density-functional thermochemistry. I. The effect of the exchange-only gradient correction. *J. Chem. Phys.* **1992**, *96*, 2155−2160.
(60) Woon, D. E.; Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. IV. Calculation of static electrical response properties. *J. Chem. Phys.* **1994**, *100*, 2975−2988.
(61) Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007−1023.
(62) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96*, 6796−6806.
(63) Laricchia, S.; Fabiano, E.; Constantin, L.; Della Sala, F. Generalized gradient approximations of the noninteracting kinetic energy from the semiclassical atom theory: Rationalization of the accuracy of the frozen density embedding theory for nonbonded interactions. *J. Chem. Theory Comput.* **2011**, *7*, 2439−2451.
(64) Thomas, L. H. The calculation of atomic fields. *Math. Proc. Camb. Philos. Soc.* **1927**, *23*, 542−548.
(65) Fermi, E. Statistical method to determine some properties of atoms. *Rend. Accad. Naz. Lincei* **1927**, *6*, 5.

https://doi.org/10.1021/acs.jctc.1c00732
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

(66) Boys, S. F. Construction of some molecular orbitals to be approximately invariant for changes from one molecule to another. *Rev. Mod. Phys.* **1960**, *32*, 296.
(67) Foster, J.; Boys, S. Canonical configurational interaction procedure. *Rev. Mod. Phys.* **1960**, *32*, 300.
(68) Jansík, B.; Høst, S.; Kristensen, K.; Jørgensen, P. Local orbitals by minimizing powers of the orbital variance. *J. Chem. Phys.* **2011**, *134*, 194104.
(69) Høyvik, I.-M.; Jansik, B.; Jørgensen, P. Orbital localization using fourth central moment minimization. *J. Chem. Phys.* **2012**, *137*, 224114.



https://doi.org/10.1021/acs.jctc.1c00732
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Chapter 4

Combining PNOs with the FDE method

For large subsystems, second-order correlation methods may be too demanding even within the framework of the FDE method. The use of PNOs can reduce the costs such that the calculations become feasible even if relatively large subsystems are needed in the context of FDE and thereby increase the applicability of FDE calculations with correlated wavefunction methods. However, the use of PNOs introduces in the results a (non-systematic) truncation error and the question arises if (or how much) this affects the accuracy of environment-induced shifts, one of the most important properties in the context of theoretical spectroscopy. This project part investigates the performance and the accuracy of this combination for tetraazaperopyrenes (TAPPs) within their molecular crystals.

It was performed under the supervision of Prof. Dr. Christof Hättig and PD Dr. Sebastian Höfener. Implementation of the contributions for second-order methods within the post-SCF reaction scheme for the *ricc2* and *pnoccsd* programs of the TURBOMOLE program package was performed by me, exploiting subroutines and structures already present from the COSMO and PE implementations in these programs. Jing Liu provided the unit cells of the TAPP molecular crystals, which I used to generate the clusters used in this work. All numbers and graphics have been created by me. The first draft of the manuscript was written by me and further developed together with the co-authors.

The here presented manuscript has been been submitted to the *Journal of Chemical Physics* by **AIP Publishing** on the 24th of May 2022. Its corresponding supporting information can be found in section A.2.

Pushing the limits: efficient wavefunction methods for excited states in complex systems using frozen-density embedding

Robert S. Treß,[1, a)] Jing Liu,[2] Christof Hättig,[1] and Sebastian Höfener[2, b)]

[1)]*Department of Theoretical Chemistry, Ruhr University Bochum, Bochum*

[2)]*Institute of Physical Chemistry, Karlsruhe Institute of Technology (KIT), 76131 Karlsruhe, Germany*

(Dated: 16 May 2022)

Frozen density embedding (FDE) is an embedding method for complex environments that is simple for users to set up. It reduces the computation time by dividing the total system into small subsystems and approximating the interaction by a functional of their densities. Its combination with wavefunction methods is, however, limited to small or medium-sized molecules because of the steep scaling in computation time of these methods. To mitigate this limitation, we present a combination of the frozen density embedding (FDE) approach with pair natural orbitals (PNOs) in the TURBOMOLE software package. It combines the uncoupled FDE (FDEu) approach for excitation energy calculations with efficient implementations of second-order correlation methods in the ricc2 and pnoccsd programs. The performance of this combination is tested for tetraazaperopyrene (TAPP) molecular crystals. It is shown that the PNO truncation error on the environment-induced shifts are significantly smaller than the shifts themselves and, thus, that the local approximations of PNO-based wavefunction methods can without the loss of relevant digits be combined with the FDE method. Computational wall times are presented for two TAPP systems. The scaling of the wall times are compared to conventional supermolecular calculations and demonstrate large computational savings for the combination of FDE and PNO-based methods. Additionally the behavior of excitation energies with the system size is investigated. It is found that for all investigated TAPPs in the current study the excitation energies converge quickly with the size of the embedding environment.

[a)]Electronic mail: robert.tress@rub.de

[b)]Electronic mail: hoefener@kit.edu

I. INTRODUCTION

The computation of accurate excitation energies usually requires large and diffuse basis sets and accurate correlated electronic structure methods beyond TDDFT with semi-local functionals. This can lead to enormous computation times for complex systems consisting of many molecules. Embedding methods can reduce the computation time by approximating the interactions of the surrounding molecules with the chromophore under study. For molecules in solution, continuum solvation models like the polarizable continuum model[1] (PCM), the conductor-like screening model[2] (COSMO), or the solvation model based on density[3] (SMD) are popular approaches. They require only a small amount of computational effort, but are limited in their applicability to situations where the description of the environment by a homogeneous continuum is sufficient. An alternative to continuum solvation models is to use a classical, i.e. molecular mechanics and electrostatic, description of the environment which leads to the QM/MM and QM/MM-pol methods[4–6] as e.g. the polarizable embedding[7–10] (PE) method. These methods require for each atom or atom type a number of parameters that need to be pre-determined. Depending on the complexity of the environment and the details of the QM/MM or QM/MM-pol method this can limit the accuracy and lead to complicated work flows that are often only worthwhile if used in repeated calculation, e.g. for geometry optimizations or MD trajectories.

The frozen density embedding (FDE) method avoids such parameterization steps which makes its application to complex environments more straight forward. In the FDE method the whole physical system is divided into subsystems that are treated separately with quantum mechanical methods and their interaction is described through their electron densities using in addition to the exact Coulomb interaction functionals for the kinetic, exchange and correlation contributions to the interaction. For subsystems of equal size, the FDE embedding can achieve a linear scaling with the number of subsystems if combined with a fast multipole method for the Coulomb interaction. While the FDE method is in principle exact, in practice the kinetic and exchange-correlation functionals have to be approximated. Also the v-representability of the optimized density of a subsystem is not guaranteed through an interacting Hamiltonian[11]. One drawback of these approximations is that strong interactions between subsystems are usually not well described, which renders the division of a molecule into several subsystems unreasonable without further measures like introducing capping groups as in the 3-FDE scheme[12]. Also, strong couplings between two close molecules can only be described within the coupled FDE (FDEc) method[13], so that for the

uncoupled FDE (FDEu) method[14] both molecules have to be considered as one subsystem. Those two cases can lead to subsystem sizes that no longer allow to use conventional implementations of higher-order methods such as coupled cluster (CC).

One way to reduce computation costs for CC methods is to reduce the virtual space by introducing pair natural orbitals (PNO). During the generation of the PNOs their contributions to the electron correlation energy[15,16] are estimated and PNOs with insignificant contributions are then neglected. This can lead to enormous reductions in computation time and storage requirements for large systems. Thus, PNO-CC methods are also an attractive option to overcome the problem that arises in FDE for large subsystems.

In the current contribution we present an extension of the FDE implementation[17] in the TURBOMOLE program package[18] for the calculation of excitation energies with the correlated second-order methods CIS(D), ADC(2), and CC2. The embedding potential can now be calculated using only the basis functions of the active subsystem, FDE calculations are no longer limited to two subsystems, the resolution of identity approximation can be used and the OMP and MPI parallelization schemes implemented in TURBOMOLE can now be exploited. For correlated wavefunction and response calculations the FDE implementation is combined with the ricc2 and pnoccsd programs[19–22] that exploit the resolution-of-the-identity (RI) and the PNO approximation to reduce the computational costs and storage requirements to allow FDE calculations with larger subsystems than it would have been possible with conventional implementations.

In the next section we give a brief overview of the FDE approach as it is used for the current work and how it is combined with coupled cluster and response theory and the pair natural orbital approximation for the calculation of excitation energies. In Sections 3 and 4 we give the computational details and discuss the results of benchmark calculations that have been carried out to analyze the performance of the implemented approach. In particular, to calculate excitation energies of tetraazaperopyrenes (TAPP) in their molecular crystals and compare wall times of FDE calculations with different subsystem sizes with those of supermolecular calculations for different sizes of the total computational models and investigate the convergence of the excitation energies with the size of the embedding environment.

II. THEORY AND IMPLEMENTATION

A. Frozen density embedding

The FDE method is based on subsystem DFT, where the total electron density is described as the sum of subsystem electron densities:[23–25]

$$\rho_{\mathrm{tot}}(\mathbf{r}) = \sum_{i=1}^{m} \rho_i(\mathbf{r}). \tag{1}$$

The total number of subsystems m is usually equal to the number of molecules in the investigated system. Using this decomposition of the total electron density leads to the following expression of the total energy:

$$E_{\mathrm{tot}} = \sum_{i=1}^{m} E_i[\rho_i] + E_{\mathrm{int}}[\rho_1(\mathbf{r}), ..., \rho_m(\mathbf{r})], \tag{2}$$

where $E_i[\rho_i(\mathbf{r})]$ is the energy expression of subsystem i and E_{int} the interaction between the different subsystems. This decomposition has the advantage that each subsystem can in general be described with a different method and in particular also with wavefunction methods. In the latter case the density functional $E_i[\rho_i]$ is for these subsystems replaced by the respective wavefunction functional $E_i[\Psi_i]$. The interaction contribution, however, is nevertheless also for these subsystems described through the electron densities. Employing wavefunction methods for all subsystems this leads to the following energy expression:

$$E_{\mathrm{tot}} = \sum_{i=1}^{m} E_i[\Psi_i] + E_{\mathrm{int}}[\rho_1(\mathbf{r}), ..., \rho_m(\mathbf{r})]. \tag{3}$$

This is motivated by the fact that, regardless of the methods chosen, the interaction is then determined only through the electron densities of the subsystems as

$$\begin{aligned} E_{\mathrm{int}}[\rho_1(\mathbf{r}), ..., \rho_m(\mathbf{r})] = &\sum_{i=1}^{m}\sum_{j\neq i}^{m} \int\int \frac{\rho_i(\mathbf{r})\rho_j(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r} d\mathbf{r}' \\ &- \sum_{i=1}^{m}\sum_{j\neq i}^{m} \int \rho_i(\mathbf{r}) v_j^N(\mathbf{r}) d\mathbf{r} \\ &+ E_{\mathrm{kxc}}^{\mathrm{nadd}}[\rho_1(\mathbf{r}), ..., \rho_m(\mathbf{r})] + E_{\mathrm{NN}}, \end{aligned} \tag{4}$$

where v_j^{N} is the potential originating from the nuclei in subsystem j and E_{NN} the nuclear repulsion contribution to the interaction energy. $E_{\mathrm{kxc}}^{\mathrm{nadd}}$ contains the nonadditive contributions of the kinetic

and exchange-correlation energy, which is given as

$$E_{\mathrm{kxc}}^{\mathrm{nadd}}[\rho_1(\mathbf{r}),...,\rho_m(\mathbf{r})] = T[\rho_1(\mathbf{r})+...+\rho_m(\mathbf{r})] - \sum_{i=1}^{m} T[\rho_i(\mathbf{r})] + E_{\mathrm{xc}}[\rho_1(\mathbf{r})+...+\rho_m(\mathbf{r})] - \sum_{i=1}^{m} E_{\mathrm{xc}}[\rho_i(\mathbf{r})] \; . \tag{5}$$

To keep $E_{\mathrm{kxc}}^{\mathrm{nadd}}$ a pure density functional that is free from any information on the wavefunctions, the kinetic contributions must be computed with an explicit kinetic energy density functional. The derivation of the equations for the energy, or wavefunction, for a subsystem i leads now to the conventional, i.e. non-embedded, equations with a Hamiltonian augmented by an additional local potential that arises from the interaction energy. This additional potential is commonly denoted embedding potential and is given by

$$v_{\mathrm{FDE}}^{i}(\mathbf{r}) = \sum_{j\neq i}^{m} \int \frac{\rho_j(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}' - \sum_{j\neq i}^{m} v_{\mathrm{N}}^{j}(\mathbf{r}) + v_{\mathrm{kxc}}^{i}(\mathbf{r}) \; , \tag{6}$$

where v_{kxc}^{i} contains the functional derivatives of the nonadditive contributions. As described in previous work[26], additional all-electron pseudopotentials are added to the environment in order to account more completely for the Pauli repulsion between the electrons of a subsystem i and its embedding environment, leading to the following Fock Operator:

$$F^{\mathrm{FDE(ECP)}} = F^{\mathrm{vac}} + v_{\mathrm{FDE}}^{i} + v_{\mathrm{ECP}}^{i} \; , \tag{7}$$

where v_{ECP}^{i} is the sum of the all-electron ECPs on all atoms of the environment for subsystem i, and an additional contribution to the interaction energy which consists of the sum of the expectation values of all subsystems i for the respective potentials v_{ECP}^{i}. With that the Hartree-Fock and, for DFT, the Kohn-Sham equations remain unchanged apart from using an augmented potential. For self-consistent field (SCF) methods like Hartree-Fock and DFT the ground-state equations that follow from the minimization of the total energy is solved self-consistently in freeze-and-thaw cycles[27].

The embedding potential is calculated for the active subsystem from the electron densities and electrostatic potentials of the subsystems in the environment on a supermolecular integration grid, that is generated by merging the atomic integration grids of all atoms in the total system. The integrals are evaluated only for the basis functions of the currently active subsystem. The FDE extension is implemented within the dscf and ridft programs for Hartree-Fock and DFT

employing OMP and MPI parallelization schemes and was validated by comparisons with the quantum chemistry software KOALA[28].

Within the perturbation of the energy (PTE) reaction field scheme, where the embedding potential is kept fixed after reached self-consistency at the Hartree-Fock level, no further changes to the code have to be made for post Hartree-Fock ground state calculations.

As for other reaction field methods that allow for a polarization of the environment, one can devise for post-SCF methods, in particular in connection with response theory, different reaction field coupling schemes[29–32], which have all their advantages and disadvantages for different applications. We use for the current implementation the post-SCF coupling scheme that in the ricc2 and pnoccsd programs of TURBOMOLE is also used for the PE and COSMO embeddings.[8,10,33]

In the post-SCF coupling scheme the reaction field for the ground-state reference wavefunction is determined fully self-consistently with the reference wavefunction and its density. The post-SCF part of the wavefunction, i.e. for coupled cluster and related methods, the amplitudes and Lagrange multipliers, are then determined in a reaction field that includes the respective contributions to the density. The post-SCF contributions to the density and the reaction field are, however, not fed back into the freeze-and-thaw cycles for the reference wavefunction. This reduces the costs relative to coupling schemes that determine all wavefunctions and densities in a reaction field that has been equilibrated with the post-SCF, e.g. coupled cluster density. But equally important, this coupling scheme avoids a functional dependence of the reference wavefunction on post-SCF contributions to the density and thereby allows for a more consistent formulation of response and perturbation theories.

In the current work we use FDEu where the response functions for each subsystem are decoupled by neglecting inter-system contributions. Within the post-SCF coupling scheme this is consistent with determining the reaction field for post-SCF correlation and response calculations from the unperturbed self-consistent field reference densities of the environment and the full post-SCF density of the active subsystem that is currently under consideration.

B. FDE contributions in post-Hartree-Fock wavefunction methods

For the coupled cluster (CC) in HF or DFT ansatz we follow the derivation in Ref. 34 and 35 and use a variational formulation in terms of a Lagrangian. For simplicity we restrict the equations to the closed-shell case and use the second quantization formalism. For an isolated system, the CC

Lagrangian is given by

$$L(\bar{t}_\mu, t_\mu) = \langle \mathrm{HF}|H|\mathrm{CC}\rangle + \sum_\mu \bar{t}_\mu \langle \mu|\exp(-T)H|\mathrm{CC}\rangle = \langle \Lambda|H|\mathrm{CC}\rangle \tag{8}$$

where $|\mathrm{CC}\rangle = \exp(T)|\mathrm{HF}\rangle$ is the CC wavefunction, $T = \sum_\mu t_\mu \tau_\mu$ the cluster operator with the excitation operators τ_μ and amplitudes t_μ, while $\langle\mu|$ is an element of the projection manifold and $\bar{t}_\mu$ the Lagrange multipliers. The latter are combined with $\langle \mathrm{HF}|$ to $\langle\Lambda| = \langle \mathrm{HF}| + \sum_\mu t_\mu \langle\mu|$. Note that if the right hand side of Eq. 8 becomes complex, only the real part should be used. The density obtained from the above Lagrangian are the so-called "unrelaxed" coupled-cluster densities since they do not account for a response (relaxation) of the orbitals.

The Lagrangian for the total energy of the whole system takes the same form as the expression for the total energy in Eq. 3, but with $E_i[\Psi_i]$ replaced by L_i for those subsystems that are treated at the CC level. For simplicity we use in the following for the active subsystem a lower index act and collect all other subsystems in the environment and write L_{tot} as the sum of the Lagrangian for the active system (L_{act}), the sum of the Lagrangians for the subsystems in the environment (L_{env}) and the interaction contribution (E_{int}):[34,35]

$$L_{\mathrm{tot}} = L_{\mathrm{act}} + L_{\mathrm{env}} + E_{\mathrm{int}} \ . \tag{9}$$

For FDEu with the post-SCF coupling scheme the interaction energy is for the coupled cluster and response part of the calculations defined as $E_{\mathrm{int}}[\rho_{\mathrm{act}}^{\mathrm{CC}}, \rho_{\mathrm{env}}^{\mathrm{SCF}}]$, where ρ_i^{SCF} can be a Hartree-Fock or DFT density and $\rho_{\mathrm{act}}^{\mathrm{CC}}$ the (unrelaxed) coupled cluster density obtained from the coupled cluster Lagrangian, Eq. (8), of the active system. The CC amplitude equations for the embedded system are obtained from the stationarity condition for the total Lagrangian, Eq. (9), with respect to the Lagrange multipliers:

$$0 = \frac{\delta L}{\delta \bar{t}_{\mu_i}} = \Omega_{\mu_i}^{\mathrm{vac}} + \sum_{pq} \langle p|v_{\mathrm{int}}|q\rangle \langle \mu_i|\exp(-T)E_{pq}|\mathrm{CC}\rangle \ , \tag{10}$$

with $\Omega_{\mu_i}^{\mathrm{vac}}$ as the vacuum contribution to the residual of the amplitude equations and

$$v_{\mathrm{int}}[\rho_{\mathrm{act}}^{\mathrm{CC}}, \rho_{\mathrm{env}}^{\mathrm{SCF}}] = v_{\mathrm{FDE}}^i[\rho_{\mathrm{act}}^{\mathrm{CC}}, \rho_{\mathrm{env}}^{\mathrm{SCF}}] + v_{\mathrm{ECP}}^i \ . \tag{11}$$

At this point we introduce the reaction field operator,

$$G = \sum_{pq} \langle p|v_{\mathrm{int}}|q\rangle E_{pq} \ , \tag{12}$$

and write equation 10 as:

$$0 = \frac{\delta L}{\delta \bar{t}_{\mu_i}} = \Omega^{\mathrm{vac}}_{\mu_i} + \langle \mu_i | \exp(-T) G | \mathrm{CC} \rangle \ . \tag{13}$$

The contributions arising from G are straight forward to evaluate as they have the same structure as the contributions arising from the Fock operator in the amplitude equations for an isolated system.

Similarly, the Lagrange multipliers are obtained from stationarity condition of the total Lagrangian with respect to the amplitudes:

$$0 = \frac{\delta L}{\delta t_{\mu_i}} = \langle \Lambda | \left[H, \tau_{\mu_i} \right] | \mathrm{CC} \rangle + \langle \Lambda | \left[G, \tau_{\mu_i} \right] | \mathrm{CC} \rangle \ . \tag{14}$$

In response theory for isolated systems, excitation energies are usually derived as poles of the response functions. This can in principle also be done for subsystem formulations. However, similar as in other embedding approaches with polarizable environments, the contributions to the Lagrangian arising from the interaction contributions that are quadratic in the multipliers introduce an additional coupling between the cluster amplitudes and the multipliers[34]. As a consequence, the poles no longer occur at the eigenvalues of the CC Jacobian, but at the eigenvalues of the full electronic Hessian with the amplitudes and multipliers considered as variables on equal footing. This leads to significant computational overhead while the numerical effects on excitation energies are expected to be small. We thus follow the computational strategy that has been successfully employed in the CC/MM schemes[36] and the related PE-CC and PCM- and COSMO-CC approaches[32,37–39] and neglect this coupling and determine the excitation energies of the embedded system as eigenvalues of its Jacobian. From the definition of the Jacobian,

$$A_{\mu_i \nu_j} = \frac{\delta^2 L_{\mathrm{tot}}}{\delta \bar{t}_{\mu_i} \delta t_{\nu_j}} = A^{\mathrm{vac}}_{\mu_i \nu_j} + A^{\mathrm{FDE}}_{\mu_i \nu_j} \ , \tag{15}$$

one obtains with the above choice for E_{int} for the FDE contribution:

$$\begin{aligned} A^{\mathrm{FDE}}_{\mu_i \nu_j} &= \langle \mu_i | \left[G, \tau_{\nu_j} \right] | \mathrm{CC} \rangle \\ &+ \frac{1}{2} \sum_{pq,rs} (pq | \, \omega_{\mathrm{int}} \, | rs) \, \langle \mu_i | \exp(-T) E_{pq} | \mathrm{CC} \rangle \langle \Lambda | [E_{rs}, \tau_{\nu_j}] | \mathrm{CC} \rangle \ . \end{aligned} \tag{16}$$

The second term includes the so called kernel contributions, which received their name due to their analogy to the exchange correlation kernel in DFT-based response theory:

$$(pq | \, \omega_{\mathrm{int}} \, | rs) = \int \int \phi^\dagger_p(\mathbf{r}') \phi_q(\mathbf{r}') \left[\left. \frac{\delta^2 E_{\mathrm{kxc}}}{\delta \rho(\mathbf{r}') \delta \rho(\mathbf{r})} \right|_{\rho_{\mathrm{tot}}} - \left. \frac{\delta^2 E_{\mathrm{kxc}}}{\delta \rho(\mathbf{r}') \delta \rho(\mathbf{r})} \right|_{\rho_{\mathrm{act}}} \right] \phi^\dagger_r(\mathbf{r}) \phi_s(\mathbf{r}) d\mathbf{r} d\mathbf{r}' \ . \tag{17}$$

The excitation energies for the active subsystem are obtained by solving the right eigenvalue for the Jacobian:

$$\mathbf{A}\mathbf{R}_k = \omega_k \mathbf{R}_k \ . \tag{18}$$

C. The second-order methods CC2, ADC(2), and CIS(D)

The approximate coupled-cluster singles and doubles model CC2 was introduced[40] as an approximation to the coupled-cluster singles and doubles model CCSD where the iterative character of coupled cluster is retained, but only terms that are at most of second order are kept in the Lagrangian. For this truncation the electron fluctuation potential, $W = H - F$, and the doubles amplitudes and multipliers are, as in Møller-Plesset perturbation theory, counted as first order. Due to their special role in response theory, the singles amplitudes and multipliers are treated as being of zeroth-order. This leads to the Lagrangian:

$$\begin{aligned} L^{\mathrm{CC2}} =& \langle \mathrm{HF}|H\exp(T)|\mathrm{HF}\rangle + \sum_{\mu_1} \bar{t}_{\mu_1} \langle \mu_1|\exp(-T_1)(H+[H,T_2])\exp(T_1)|\mathrm{HF}\rangle \\ &+ \sum_{\mu_2} \bar{t}_{\mu_2} \langle \mu_2|\exp(-T_1)H\exp(T_1) + [F,T_2]|\mathrm{HF}\rangle \ . \end{aligned} \tag{19}$$

The CC2 model gives ground-state and excitation energies for single excitations that are complete through second-order in the fluctuation potential.

For FDE, the second-order approximation must be generalized to the interaction contribution in the Lagrangian. In this context, the intersystem electron-electron interaction should in the perturbation expansion be treated similar as the intrasystem electron-electron interactions. A natural choice is to count the interaction potential that arises from the SCF densities as being of zeroth order, on equal footing with the Coulomb and exchange contributions in the vacuum Fock matrix, and all intersystem electron-electron interactions in contributions beyond the mean field potential as being of first order. In case of CC2 the unrelaxed density matrix[40] given as:

$$\begin{aligned} D_{pq}^{\mathrm{CC2}} =& \langle \mathrm{HF}|\exp(-T_1)E_{pq}\exp(T_1)|\mathrm{HF}\rangle \\ &+ \sum_{\mu_1} \bar{t}_{\mu_1} \langle \mu_1|\exp(-T_1)(E_{pq}+[E_{pq},T_2])\exp(T_1)|\mathrm{HF}\rangle \\ &+ \sum_{\mu_2} \bar{t}_{\mu_2} \langle \mu_2|\exp(-T_1)[E_{pq},T_2]\exp(T_1)|\mathrm{HF}\rangle \ , \end{aligned} \tag{20}$$

depends in the last term on the doubles multipliers and doubles amplitudes, which is therefore of second order, and gives in the interaction contribution to L_{tot} only rise to contributions that

are of third order and beyond that are not consistent with a second-order model. As discussed in Ref. 8 in the framework of polarizable embedding, also the contribution to D^{CC2} that is proportional to the doubles amplitudes and singles multipliers introduces in reaction field approaches contributions beyond second order and causes for a consistent second-order model complications. It is, therefore, skipped for CC2. The interaction term is thus in the post-SCF coupling scheme for CC2 evaluated with a CCS-like density matrix which includes in addition to the Hartree-Fock term only contributions from the singles amplitudes and multipliers, which are taken from the FDE-CC2 calculation:

$$\begin{aligned} D_{pq}^{\mathrm{CCS}} =& \langle \mathrm{HF}|\exp(-T_1)E_{pq}\exp(T_1)|\mathrm{HF}\rangle \\ &+ \sum_{\mu_1} \bar{t}_{\mu_1} \langle \mu_1|\exp(-T_1)E_{pq}\exp(T_1)|\mathrm{HF}\rangle \ . \end{aligned} \tag{21}$$

The total energy Lagrangian for CC2 in HF or DFT thus becomes:

$$L_{\mathrm{tot}} = L_{\mathrm{act}}^{\mathrm{CC2}} + E_{\mathrm{int}}[\rho_{\mathrm{act}}^{\mathrm{CCS}}, \rho_{\mathrm{env}}^{\mathrm{SCF}}] + L_{\mathrm{env}}^{\mathrm{SCF}} \ . \tag{22}$$

This leads to the following amplitude equations:

$$0 = \langle \mu_1|\exp(-T_1)\left(H^{\mathrm{emb}} + \left[H^{\mathrm{emb}}, T_2\right] + G^{\Delta}\right)\exp(T_1)|\mathrm{HF}\rangle \tag{23}$$

$$0 = \langle \mu_2|\exp(-T_1)H^{\mathrm{emb}}\exp(T_1) + \left[F^{\mathrm{emb}}, T_2\right]|\mathrm{HF}\rangle \ , \tag{24}$$

with

$$G^{\Delta} = G - G^{\mathrm{SCF}} = \sum_{pq} \langle p| v_{\mathrm{int}}\left[\rho_{\mathrm{act}}^{\mathrm{CCS}}, \rho_{\mathrm{env}}^{\mathrm{SCF}}\right] - v_{\mathrm{int}}\left[\rho_{\mathrm{act}}^{\mathrm{SCF}}, \rho_{\mathrm{env}}^{\mathrm{SCF}}\right]|q\rangle E_{pq} \tag{25}$$

$$= \sum_{pq} \langle p| v_{\mathrm{kxc}}\left[\rho_{\mathrm{act}}^{\mathrm{CCS}}, \rho_{\mathrm{env}}^{\mathrm{SCF}}\right] - v_{\mathrm{kxc}}\left[\rho_{\mathrm{act}}^{\mathrm{SCF}}, \rho_{\mathrm{env}}^{\mathrm{SCF}}\right]|q\rangle E_{pq} \ , \tag{26}$$

where G^{SCF} is the SCF reaction field operator calculated from the SCF densities. In Eqs. 23 and 24 we included the SCF reaction field in F^{emb} and H^{emb} which are defined as the respective vacuum operators plus the SCF interaction potential:

$$F^{\mathrm{emb}} = F^{\mathrm{vac}} + v_{\mathrm{int}}[\rho_{\mathrm{act}}^{\mathrm{SCF}}, \rho_{\mathrm{env}}^{\mathrm{SCF}}] \ . \tag{27}$$

This is motivated by the fact that the SCF reaction field is included via the Fock matrix obtained from the preceding FDE calculation at the HF or DFT level. In the post-SCF part F^{emb} remains unchanged, only G^{Δ} depends on the cluster amplitudes and Lagrange multipliers. We note that

in the equation for the doubles amplitudes, Eq. (24), in the first term on the right hand side from $H^{\mathrm{emb}} = F^{\mathrm{emb}} + W$ only the electron fluctuation potential W contributes.

The equations for the Lagrange multipliers become:

$$\begin{aligned} 0 =& \langle \mathrm{HF}| \left[H^{\mathrm{emb}} + G^{\Delta}, \tau_{\mu_1} \right] \exp(T) |\mathrm{HF}\rangle \\ &+ \sum_{\mu_1} \bar{t}_{\mu_1} \langle \mu_1 | \exp(-T_1) \left(\left[H^{\mathrm{emb}} + G^{\Delta}, \tau_{\mu_1} \right] + \left[\left[H^{\mathrm{emb}}, T_2 \right], \tau_{\mu_1} \right] \right) \exp(T_1) |\mathrm{HF}\rangle \\ &+ \sum_{\mu_2} \bar{t}_{\mu_2} \langle \mu_2 | \exp(-T_1) \left[H^{\mathrm{emb}}, \tau_{\mu_1} \right] \exp(T_1) |\mathrm{HF}\rangle \end{aligned} \tag{28}$$

and

$$\begin{aligned} 0 =& \langle \mathrm{HF}| \left[H^{\mathrm{emb}}, \tau_{\mu_2} \right] \exp(T) |\mathrm{HF}\rangle \\ &+ \sum_{\mu_1} \bar{t}_{\mu_1} \langle \mu_1 | \exp(-T_1) \left[H^{\mathrm{emb}}, \tau_{\mu_2} \right] \exp(-T_1) |\mathrm{HF}\rangle \\ &+ \sum_{\mu_2} \bar{t}_{\mu_2} \langle \mu_2 | \left[F^{\mathrm{emb}}, \tau_{\mu_2} \right] |\mathrm{HF}\rangle \ . \end{aligned} \tag{29}$$

With the restriction to the CCS-like density for the interaction energy, for the Jacobian explicit contributions from the embedding beyond those automatically included via F^{emb} arise only in the singles-singles block.

$$\begin{aligned} A^{\mathrm{FDE}}_{\mu_1 \nu_1} =& \langle \mu_1 | \exp(-T_1) \left[\left(H^{\mathrm{emb}} + \left[H^{\mathrm{emb}}, T_2 \right] + G^{\Delta} \right), \tau_{\nu_1} \right] \exp(T_1) |\mathrm{HF}\rangle \\ &+ \sum_{pqrs} (pq|\, \omega^{\mathrm{CCS}}_{\mathrm{int}} \,|rs) \, \langle \Lambda_1 | [E_{rs}, \tau_{\nu_1}] \exp(T_1) |\mathrm{HF}\rangle \langle \mu_1 | \exp(-T_1) E_{pq} \exp(T_1) |\mathrm{HF}\rangle \end{aligned} \tag{30}$$

The eigenvalue problem for the excitation energies, Eq. (18), is solved using iterative direct techniques which only needs the possibility to compute contractions of $\mathbf{A}$ with trial vectors $\mathbf{b}$. For the FDE contributions this is implemented by first setting up a transition density

$$\rho^b(\mathbf{r}) = \sum_{rs} \phi^{\dagger}_r(\mathbf{r}) \phi_s(\mathbf{r}) D^b_{rs} \tag{31}$$

$$\mathbf{D}^b = \begin{pmatrix} -\bar{\mathbf{t}}_1^{\dagger} \mathbf{b}_1 & 0 \\ 2\mathbf{b}_1 & \bar{\mathbf{t}}_1 \mathbf{b}_1^{\dagger} \end{pmatrix} , \tag{32}$$

which is then used to evaluate the integrals on the supermolecular grid,

$$G^b_{\mu\nu} = \int \chi_{\mu}(\mathbf{r}') \chi_{\nu}(\mathbf{r}') \int \left[\left. \frac{\delta^2 E_{\mathrm{kxc}}}{\delta\rho(\mathbf{r}')\delta\rho(\mathbf{r})} \right|_{\rho_{\mathrm{tot}}} - \left. \frac{\delta^2 E_{\mathrm{kxc}}}{\delta\rho(\mathbf{r}')\delta\rho(\mathbf{r})} \right|_{\rho_{\mathrm{act}}} \right] \rho^b(\mathbf{r}) d\mathbf{r} d\mathbf{r}' \ , \tag{33}$$

for the atomic orbitals χ_μ. These integrals are then transformed to the molecular orbital basis, which allows to formulate the FDE contribution to $\sigma = \mathbf{Ab}$ as

$$\sigma_{\mu_1}^{\mathrm{FDE}} = \sum_{\nu_1} A_{\mu_1\nu_1}^{\mathrm{FDE}} b_{\nu_1} = \langle \mu_1 | \exp(-T_1) \left(\left[G^{\Delta}, b_1 \right] + G^b \right) \exp(T_1) | \mathrm{HF} \rangle \ , \tag{34}$$

with the one-electron operators $b_1 = \sum_{ab} b_{ai} E_{ai}$ and $G^b = \sum_{pq} G^b_{pq} E_{pq}$. The latter equation is evaluated similar to other one-electron contributions.

The algebraic diagrammatic construction is a propagator approach[41] which uses a very different theoretical framework for the derivation of response functions than it is used in coupled cluster response theory. At the second-order level[42], ADC(2), the secular matrix from which the excitation energies are obtained as eigenvalues is, however, technically closely related to that for the iterative variant of the configuration interaction singles model with a second-order doubles correction[43], CIS(D_∞), and the Jacoby matrix for CC2:[44] The secular matrix for CIS(D_∞) is obtained by replacing in the CC2 Jacoby matrix the CC2 ground-state singles and doubles amplitudes by those from first-order Møller-Plesset perturbation theory. Similar to the CC2 Jacobian, it is non-Hermitian. The secular matrix for ADC(2) is equal to the Hermitian part of that for CIS(D_∞).

The excitation energies for configuration interaction singles, CIS, with a perturbative second-order doubles correction[45], CIS(D), can be written as:

$$\omega^{\mathrm{CIS(D)}} = \mathbf{E}_1^{\mathrm{CIS}} \left(\mathbf{A}_{11}^{\mathrm{CIS}(D_\infty)} - \mathbf{A}_{12}^{\mathrm{CIS}(D_\infty)} \left[\mathbf{A}_{22}^{\mathrm{CIS}(D_\infty)} - \omega^{\mathrm{CIS}} \mathbf{1}_{22} \right]^{-1} \mathbf{A}_{21}^{\mathrm{CIS}(D_\infty)} \right) \mathbf{E}_1^{\mathrm{CIS}} \ , \tag{35}$$

where ω^{CIS} and $\mathbf{E}^{\mathrm{CIS}}$ are the CIS eigenvalues and eigenvectors and $\mathbf{A}_{kl}^{\mathrm{CIS}(D_\infty)}$ the respective blocks of the CIS(D_∞) secular matrix for the excitation classes k, l. Note that for excitation energies CIS is equivalent to coupled cluster singles, CCS, if the reference wavefunction fulfills the Brillouin condition.

The same relations can be used to derive from the FDE expressions for CC2 in HF or DFT the respective expressions for ADC(2), CIS(D_∞), and CIS(D). All three methods require the ground-state amplitudes from first-order Møller-Plesset perturbation theory. The post-SCF coupling scheme used in the current work becomes at this level identical with the PTE scheme. The first-order singles amplitudes vanish and the first-order doubles amplitudes are determined by the equation:

$$0 = \langle \mu_2 | [F^{\mathrm{emb}}, T_2^{(1)}] + W | \mathrm{HF} \rangle \ . \tag{36}$$

The last equation differs from equation 24 only in that the fluctuation potential is not similarity-transformed with $\exp(T_1)$.

D. Pair natural orbital approximation of excitation energies

The implementation of the CIS(D) excitation energies with pair natural orbitals has been described in Ref. 20 and its extension to the iterative methods ADC(2) and CC2 in Ref. 21. Here, only a brief summary on the use of the pair natural orbitals in the calculation of excitation energies is given.

The second-order methods for excitation energies have in common that the computationally most costly steps are related to the doubles amplitudes for the ground (t_{μ_2}) and excited states (E_{μ_2}) and two-electron integrals of the type $(ai|bj)$ with two occupied (i, j) and two virtual (a, b) MO indeces, where the orbitals can be dressed with ground- and/or excited-states singles amplitudes. For localized occupied molecular orbitals (LMOs) the latter integrals are only non-negligible if the virtual MOs a and b are close to the LMOs i and j and their magnitude decays as R^{-3} with the distance between the centers of i and j.

This sparsity is exploited in the pair natural orbital approximation by determining for each LMO pair ij a set of PNOs that are defined as linear combinations of virtual orbitals

$$|\bar{a}\rangle = \sum_a |a\rangle d^{ij}_{a\bar{a}} \,, \tag{37}$$

with the coefficient $d^{ij}_{a\bar{a}}$ chosen such that they diagonalize the correlation contribution D^{ij} of pair ij to the one-electron density matrix:

$$\sum_{ab} d^{ij}_{a\bar{a}} D^{ij}_{ab} d^{ij}_{b\bar{b}} = n^{ij}_{\bar{a}} \delta_{\bar{a}\bar{b}} \,, \tag{38}$$

with

$$D^{ij}_{ab} = \frac{2}{N_{ij}} \sum_c \left(u^{ij}_{ac} t^{ij}_{bc} + u^{ij}_{ca} t^{ij}_{cb} \right), \tag{39}$$

$$N_{ij} = 1 + \frac{2}{1+\delta_{ij}} \sum_{ab} u^{ij}_{ab} t^{ij}_{ab}, \tag{40}$$

$$u^{ij}_{ab} = 2t^{ij}_{ab} - t^{ij}_{ba} \tag{41}$$

and t^{ij}_{ab} as the MP2 doubles amplitudes. The eigenvalues $n^{ij}_{\bar{a}}$ are natural orbital occupation numbers and provide a measure for the importance of the corresponding PNO. Thus, all PNOs with a natural occupation number lower than a chosen PNO threshold value T_{PNO} are discarded. For localized LMOs ij the $n^{ij}_{\bar{a}}$ decay rapidly with the index $\bar{a}$ towards zero which then leads to a large reduction of the virtual space for ij. In the second-order correlation methods CIS(D), ADC(2), and CC2, the

double amplitudes are then used only within the PNO basis, and the integrals in which they are contracted are also projected onto it.

For the excited-state doubles amplitudes a state-specific PNO basis is constructed for each excited state to account for the different characters of the excitations, which can be localized on a single chromophore group or exhibit long-range charge-transfer character in case of non local excitation. The excited state PNOs are obtained by a similar procedure as for the ground-state but with t_{ab}^{ij} replaced by the first order excited-state doubles amplitudes[20] E_{ab}^{ij}:

$$\tilde{D}_{ab}^{ij} = \frac{2}{\tilde{N}_{ij}} \sum_{c} \left(\tilde{u}_{ac}^{ij} E_{bc}^{ij} + \tilde{u}_{ca}^{ij} t_{cb}^{ij} \right) . \tag{42}$$

For further technical details on the generation of the PNO coefficients with a reduced scaling of the operation count with the system size and the truncation of the set of significant LMOs pairs we refer to Refs. 20–22.

All FDE contributions have also been implemented within the pnoccsd program and have been tested by comparisons to the ricc2 program in the limit of a very tight PNO truncation threshold. With that, FDE is available in the pnoccsd program within the frozen density approximation and the post SCF reaction field scheme in combination with PNOs.

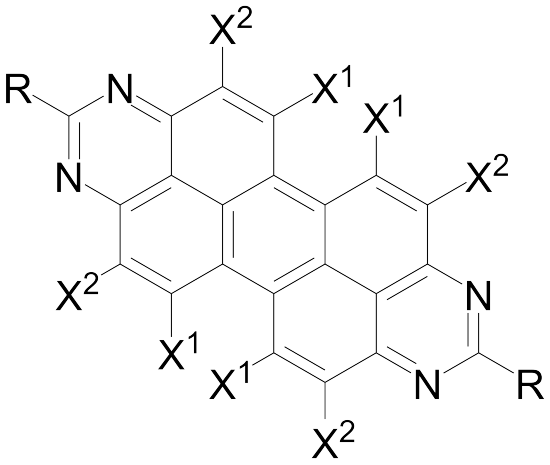
III. COMPUTATIONAL DETAILS

A. Geometries

To test the performance of the FDE implementation, clusters were created from tetraazaperopyrene (TAPP) molecular crystals. The general structure of the TAPPs is shown in Figure 1 together with the TAPP derivatives used for the current study.

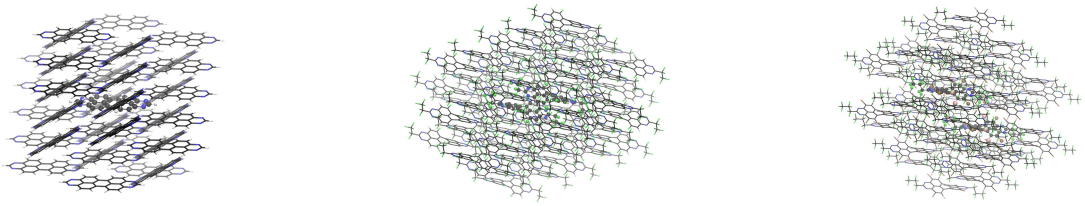
Starting point for constructing the TAPP clusters were solid-state geometries for selected TAPP compounds available in the Cambridge Crystallographic Data Centre (CCDC). We have used the CCDC entries 747374 (Refs. 46 and 47) and 1980431 (Refs. 48 and 49) in case of H-TAPP-H and *all*-C_3F_7-TAPP-Cl_4Br_4, respectively. In case of monoclinic *bay*-C_2F_5-TAPP-H_4F_4 the structure from CCDC entry 1908781 (Refs. 50 and 51) is used in which the chlorine atoms are replaced with fluorine atoms. Starting from these initial geometries, the bulk phase was optimized using the CP2K program[52] with periodic DFT methods. Both, the cell parameters and the crystal structures, have been optimized with the PBE functional[53,54] and the TZVP-MOLOPT-GTH[55] basis set together with the GTH-PBE pseudopotential[56]. The DFT-D3 correction has been applied to account

TAPP



Compound	X^1	X^2	R
(**1**) H-TAPP-H	H	H	H
(**2**) *bay*-C_2F_5-TAPP-H_4F_4	F	H	C_2F_5
(**3**) *all*-C_3F_7-TAPP-Cl_4Br_4	Cl	Br	C_3F_7

FIG. 1: Chemical structure of the investigated TAPP compounds



(a) H-TAPP-H (b) *bay*-C_2F_5-TAPP-H_4F_4 (c) *all*-C_3F_7-TAPP-Cl_4Br_4

FIG. 2: Structures of the largest extracted clusters for each TAPP.

for dispersion[57]. Based on the fully optimized bulk phases, the clusters have been extracted using Avogadro[58] and VMD[59] such that only those molecules are included which are entirely located within a specified radius. A detailed analysis for a broad range of TAPP compounds, including cell structure or electronic gaps as well as crystal stability, will be published elsewhere[60].

With this method clusters have been created with a radii of 8, 10, 12 and 15 Å for compound (**1**), with radii of 10, 12, 15 and 20 Å for compound (**2**), and with radii of 10, 13, 15 and 20 Å for compound (**3**) around a single central molecule. Additional clusters have been created for the compounds (**2**) and (**3**) with radii of 10, 15 and 20 Å around a dimer of two stacked molecules. The largest clusters of each compound are shown in Figure 2.

Additionally to the TAPPs, two systems from previous work[26] are used to investigate the influence of the kernel contributions on excitation energies: the Green Fluorescence Protein (GFP) system and the 2,3-Diazabicyclo[2.2.2]oct-2-en (DBO) within Cucurbit[7]uril (CB[7]), both shown in Figure 3.

TAPP



(a) GFP (b) DBO/CB[7]

FIG. 3: Structures of systems used to investigate the post-SCF contributions. The examined chromophore is displayed with balls and sticks, while the environment is displayed as licorice.

B. FDE and excitation energy calculations

All FDE and excitation energy calculations were carried out with a development version of the TURBOMOLE program package[18]. For all TAPP systems we used the def2-TZVP basis set[61,62] for the atoms of the central molecule and the def2-SVP basis set[61,62] for the molecules in the environment. All FDE calculations used for each subsystem the respective monomer basis. For the current study, the first step is a self-consistent FDE calculation with all subsystems treated at the Hartree-Fock level. The resolution of identity approximation for the Coulomb potential[63], RI-J, was employed for all subsystems except the central one, using the respective optimized auxiliary basis sets from Refs. 62. The embedding potential was calculated with the PBE exchange-correlation[53] and revAPBEk kinetic energy functionals[64] on a supermolecular integration grid constructed from the atomic integration grids of all atoms using grid size 3. In addition, the all-electron pseudopotentials[65] ecp-all have been used to prevent electron spill-out effects[26]. The freeze and thaw iterations have been carried out until the electronic energies of all subsystems were converged within a threshold of $5.0 \cdot 10^{-6}$ H. The frozen core approximation has been applied for all orbitals with energies below -3.0 Hartree. Except for test calculations (*vide infra*), all PNO-based calculations were performed with a PNO truncation threshold of 10^{-8} and used intrinsic bond orbital (IBO) localization for the occupied orbitals.[66] All post Hartree-Fock calculations used for the resolution-of-the-identity approximations in the correlation contributions and the excitation energy calculations the respective auxiliary basis sets[67] for RI-CC.

TABLE I: The first two excitation energies ω of the GFP system without (w/o) and with (w/) kernel contributions using the FDEu(ECP)-CCS method. All in eV.

	ω		$\Delta\omega$
	w/o kernel	w/ kernel	
S_1	3.661	3.662	$0.66 \cdot 10^{-3}$
S_2	4.538	4.552	$13.65 \cdot 10^{-3}$

IV. RESULTS

A. Evaluation of the kernel contributions

The kernel contributions originate from a non-linear dependence of the embedding potential on the electron density in the active subsystem. Therefore, this contributions should be larger when the distances between the active and surrounding subsystems are small resulting in increased overlap of subsystem electron densities. To evaluate their importance, we computed the kernel contributions for a GFP system and for DBO within CB[7] as shown in Figure 3. The GFP system has been chosen because of hydrogen bonds between the subsystems and also because of the in general small distances between them. In previous studies[10,65] DBO in CB[7] was found to be sensitive to electron charge spill-out in electronically excited states which then causes a strong interaction between the subsystems. Therefore excitation energies have been computed for the first two excitations within the CCS model using either the frozen solvent[68] approximation (w/o kernel) or the post-SCF reaction field scheme (w/ kernel). For these systems the aug-cc-pVTZ basis[69] for the chromophore and the cc-pVDZ basis[70] for the environment has been used together with the lc94 kinetic energy functional[71] for the embedding potential.

As shown in Table I and II the kernel contributions are in general rather small. In case of DBO in CB[7] the largest kernel contribution is only 0.64 meV, which is negligible. For the GFP system a contribution of 13.65 meV is observed for the second excitation, which shows that the kernel contributions can be significant. This difference between the kernel contributions indicates its importance for systems with hydrogen bonds between subsystems. Since the TAPPs which are studied in the following do not exhibit hydrogen bonds between the molecules, only insignificant contributions from the kernel are expected. To check this presumption, we computed the kernel contributions for compound (**2**). For most excitations they are below 1 meV; the largest

TABLE II: The first two excitation energies of the DBO/CB[7] system with and without kernel contributions using the FDEu(ECP)-CCS method. All in eV.

	ω		$\Delta\omega$
	w/o kernel	w/ kernel	
S_1	3.682	3.683	$0.64 \cdot 10^{-3}$
S_2	6.711	6.711	$0.38 \cdot 10^{-3}$

contribution obtained is 3.47 meV for the 14th excitation.

Since the kernel contributions to the excitations of interest are only minor for the TAPPs, we decided to employ the frozen solvent approximation in the following.

B. PNO truncation error in FDE calculations

A major task of embedding methods in the framework of theoretical spectroscopy is to model environment-induced shifts in the spectra. These shifts are often small, so it is necessary to be careful that an approximation such as the truncation of the PNOs does not introduce errors in these shifts that are larger than the shifts themselves. In order to ensure that PNO-based methods are able to qualitatively describe environment-induced shifts for electronic valence excitations, the PNO truncation error on the environment shift was investigated. Therefore, the first four excitation energies of the isolated molecule and the molecule within the environment was computed with and without PNO approximation for compounds (**1**) and (**2**) within the CC2 model. The clusters with radii of 15 Å for compound (**1**) and 20 Å for compound (**2**) around the target molecule were in these calculations used to model the environment.

For compound (**2**) we can see in Table III that the PNO truncation errors on the excitation energies increases with the excitation energy and reaches for S_4 a value of 59.21 meV. The environment-induced shift for the first two excitations are relatively small, below 0.1 eV, so that the PNO truncation errors on the excitation energies are even slightly larger than the environment shifts. However, most of the PNO truncation errors consists of a mostly systematic blue shift and, thus, the PNO truncation error on the environment shift is nevertheless much smaller than the environment shift itself, in most cases by an order of magnitude.

For compound (**1**) the truncation errors on the excitation energies (see Table IV) are lower than

TABLE III: Excitation energy ω, environment induced shift $\Delta_{\text{env}}\omega$ and the corresponding truncation errors $\Delta_{\text{trunc}}\omega$ and $\Delta_{\text{trunc}}(\Delta_{\text{env}}\omega)$ in eV for the first four excitations of (**2**) within the largest tested cluster.

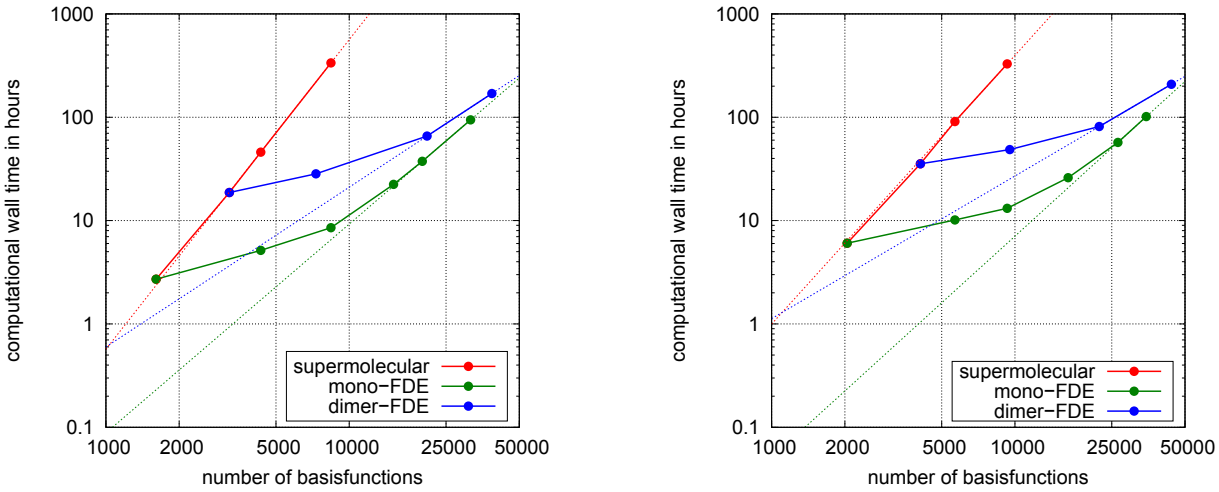
	ω	$\Delta_{\text{trunc}}\omega$	$\Delta_{\text{env}}\omega$	$\Delta_{\text{trunc}}(\Delta_{\text{env}}\omega)$
method	PNO-CC2	PNO-CC2	FDE(ECP)-PNO-CC2	FDE(ECP)-PNO-CC2
reference	-	CC2	PNO-CC2	FDE(ECP)-CC2
S_1	2.964	$+18.25\cdot10^{-3}$	$+13.16\cdot10^{-3}$	$+4.64\cdot10^{-3}$
S_2	3.057	$+29.71\cdot10^{-3}$	$+25.92\cdot10^{-3}$	$-1.19\cdot10^{-3}$
S_3	3.191	$+48.83\cdot10^{-3}$	$+157.51\cdot10^{-3}$	$-12.10\cdot10^{-3}$
S_4	3.208	$+59.21\cdot10^{-3}$	$+141.88\cdot10^{-3}$	$-27.92\cdot10^{-3}$

TABLE IV: Excitation energy ω, environment induced shift $\Delta_{\text{env}}\omega$ and the corresponding truncation errors $\Delta_{\text{trunc}}\omega$ and $\Delta_{\text{trunc}}(\Delta_{\text{env}}\omega)$ in eV for the first four excitations of (**1**) within the largest tested cluster.

	ω	$\Delta_{\text{trunc}}\omega$	$\Delta_{\text{env}}\omega$	$\Delta_{\text{trunc}}(\Delta_{\text{env}}\omega)$
method	PNO-CC2	PNO-CC2	FDE(ECP)-PNO-CC2	FDE(ECP)-PNO-CC2
reference	-	CC2	PNO-CC2	FDE(ECP)-CC2
S_1	3.060	$-2.60\cdot10^{-3}$	$+9.77\cdot10^{-3}$	$+0.19\cdot10^{-3}$
S_2	3.235	$+14.47\cdot10^{-3}$	$+26.78\cdot10^{-3}$	$-8.07\cdot10^{-3}$
S_3	3.238	$+8.80\cdot10^{-3}$	$+161.31\cdot10^{-3}$	$+12.38\cdot10^{-3}$
S_4	3.241	$+9.42\cdot10^{-3}$	$+167.02\cdot10^{-3}$	$+2.66\cdot10^{-3}$

the environment-induced shifts and the effect of the PNO truncation on the environment shifts are also for this compound much smaller than the shifts themselves. For the first excitation energy the PNO truncation error on the environment shift is even below 1 meV. We thus conclude that a PNO truncation threshold of 10^{-8} is sufficient to ensure a qualitatively correct description of the environment-induced shifts for the investigated TAPPs.

TAPP



(a) *bay*-C_2F_5-TAPP-H_4F_4 (b) *all*-C_3F_7-TAPP-Cl_4Br_4

FIG. 4: Computational wall time versus the number of basis function for (**2**) (left) and for (**3**) (right). The dashed lines represent the functions $f(N) = k \times N^a$ that were numerically fitted to the last two data points of each method. Further information to the fit can be found in the supporting information.

C. Timings and scaling with system size

The computational performance of the implementation was evaluated for compounds (**2**) and (**3**) by determining wall times of conventional supermolecular and FDE calculations for various system sizes with either one or two molecules as the target subsystem for the FDE calculations. The supermolecular calculations were performed with the same mix of basis sets as the FDE calculations together with the RI-J approximation. All computational wall times have been determined for ground state (FDE-)Hartree-Fock calculations followed by a PNO-CIS(D) excitation energy calculation of four excitations on two Intel(R) Xeon(R) Gold 6230 CPUs with 2.10GHz using OMP parallelization on 40 cores in total with a maximum available memory of 192 Gigabytes. Figure 4 shows the obtained wall times against the number of basis functions of the total system on a logarithmic scale.

For compound (**2**) the computation time of the FDE calculations with a monomer as the target subsystem initially increases sublinear with the number of basis functions N, but increases to $k \times N^{1.98}$, i.e. quadratically with the overall system size. This behavior arises because for small systems the computation time is dominated by the computation of the target subsystem includ-

ing the excitation energies. Additional subsystems in the environment do not lead to a significant increase in computation time. As the size of the total system increases, also the integration grid increases, resulting in higher computational costs for the embedding potential, especially the electrostatic potential. This, along with the linear increase in computation time for the FDE method with the number of subsystems, leads for the current implementation to a quadratic scaling with the number of basis functions above a certain system size. In this case, the point is reached between 8.4k and 15.2k basis functions. It should be pointed out that this behavior with respect to the basis functions only occurs because the system size is also increased. An increase in basis functions without any change in system size will lead to a similar behavior as in supermolecular calculations. The quadratic scaling with the overall system size can be avoided with fast multipole methods, but this is beyond the scope of the present work.

For the FDE calculations with a TAPP dimer as target subsystem, only the lower scaling behavior is observed within the range of system sizes used here, but a similar behavior for larger environments as for the FDE calculations with the monomer as target subsystem is expected.

The computation times for the supermolecular calculations exhibit even with the PNO-based implementation a steeper scaling compared to the FDE calculations. Here, the computation time is dominated by the calculation of the excitation energies. From the available data points we can deduce an approximate power law of $k \times N^{2.99}$ for the wall time of these calculations in agreement with the scaling reported in Ref. 22, although the effective scaling varies with the system size, structure and the basis set depending on the onset of integral screening and PNO truncation.

The different scaling of the computation time with the number of basis functions between the FDE method and supermolecular calculations allows the computation of much larger systems with the FDE approach. This can be shown by the largest system for which the wall times were determined for both methods and which consists of 8.4k basis functions and 528 atoms. For this system, the FDE calculation was 39 times faster than the supermolecular calculation, saving 327 h of wall time.

Compound (**3**) (see Figure 4b) shows a similar behavior for all methods. The wall times of FDE calculations with a monomer as the target subsystem show at the beginning a comparably slow increase in computation time with the number of basis functions and for larger systems an increase with $k \times N^{1.82}$. With a dimer as the target subsystem, again only the slow increase within this test series is visible. For the supermolecular calculations, an increase in computation time with the number of basis functions of $k \times N^{2.75}$ is observed. This is slightly lower compared to (**2**),

probably due to the larger distance between the subsystems, which leads to more negligible pairs in the calculation of the excitation energies. The largest system of compound (**3**) calculated within the supermolecular basis and with the FDE approach consists of 9.3k basis functions and 486 atoms. For this system the FDE calculation was 25 times faster than the supermolecular approach and could so save 315 h of wall time.

D. Convergence of the excitation energies with the environment size

We also investigated how large the environment needs to be for converged results for the excitation energies in the target subsystem. For this purpose, 18 excitation energies were computed for four different radii for three different molecular crystals employing the CIS(D).

Table V shows the lowest 10 excitation energies for compound (**2**). For the clusters obtained with radii 10 and 12 Å the target molecule is not surrounded by neighbor molecules on all sides. First with 15 Å the π system of the central molecule is surrounded entirely either by neighbor molecules or -C_2F_5 side chains. Up to this point, the excitation energies show variations > 10 meV. From the cluster for 15 Å and to that for 20 Å the excitation energies change at most by about 0.001 eV. We thus conclude that in this case 20 subsystems for the environment are sufficient to converge the environment-induced shifts. The largest change of 0.113 eV is found for the 9th excitation energy between none and the first four environment subsystems. For the 9th excitation, the difference in excitation energy between the isolated monomer and the monomer surrounded by 44 molecules is the highest with 0.181 eV. The environment-induced shifts are expected to become in general larger for higher excitation energies, as the orbitals into which the electrons are excited become more diffuse and have larger overlap with the electron density of neighboring molecules. With a few exceptions, the excitation energies of the TAPPs are all blue shifted with the influence of the environment.

Table VI shows the excitation energies for compound (**3**). Its geometric structure is similar to that of (**2**) only with a slightly larger side chain (-C_3F_7 instead of -C_2F_5) and we observe a similar convergence of the environment-induced shifts with the cluster size, in particular, only minor changes < 10 meV between the clusters obtained with a 15 and 20 Å radius. We thus expect that the excitation energies are converged within 10 meV with the size of the environment. The largest changes are observed between the clusters for 13 and 15 Å, indicating the importance of the additional molecules added in this step to the environment. In contrast to compound (**2**), about

TABLE V: Environment-induced shifts of the first 10 excitation energies in eV of compound (**2**) for different sizes of the environment.

	ω	$\Delta_{\text{env}}\omega$			
radius	0Å	10Å	12Å	15Å	20Å
Env.[a]	0	4	10	20	44
S_1	3.114	+0.013	+0.011	+0.027	+0.027
S_2	3.250	+0.018	+0.019	+0.005	+0.005
S_3	3.697	+0.001	+0.004	−0.027	−0.026
S_4	3.702	+0.041	+0.034	−0.002	−0.002
S_5	3.749	−0.005	−0.001	+0.077	+0.077
S_6	3.898	−0.034	−0.028	−0.029	−0.029
S_7	3.898	+0.023	+0.025	−0.003	−0.003
S_8	4.138	+0.019	+0.020	+0.004	+0.004
S_9	4.307	+0.113	+0.103	+0.182	+0.181
S_{10}	4.638	+0.019	+0.020	−0.001	+0.000

[a] Number of environment subsystems

half of the excitation energies are red shifted by the interaction with the environment. The origin for the change in this trend from (**2**) to (**3**) is most likely the larger distance between the π systems of neighboring TAPP molecules in the crystal with the larger -C_3F_7 side chain. The increased distance reduces the Pauli repulsion between the electrons in the active target subsystem and the environment which is usually in excited states larger than in the ground states. This is corroborated by the fact that most excitation energies increase between none and four environment subsystems, while the red shifts mostly happen when more distant neighbor molecules are included in the larger clusters. The latter should mainly interact electrostatically with the central molecule. The red shift of excitation energies originate usually from a decrease of the orbital energies for the virtual orbitals by favorable electrostatic interactions with the environment. Only for the second excitation a decrease of the excitation energies for the small environment is observed. This could be explained by an increase of the orbital energy for the occupied molecular orbital out of which an electron is excited. Again, the largest changes in excitation energies are observed for the higher

TABLE VI: Environment-induced shifts for the first 10 excitation energies in eV of compound (**3**) for different sizes of the environment.

	ω	$\Delta_{\text{env}}\omega$			
radius	0Å	10Å	13Å	15Å	20Å
Env.[a]	0	4	8	16	36
S_1	2.815	+0.005	+0.006	−0.027	−0.026
S_2	2.930	−0.019	−0.019	−0.021	−0.021
S_3	3.115	+0.038	+0.042	−0.019	−0.012
S_4	3.192	+0.029	+0.033	−0.031	−0.027
S_5	3.395	+0.044	+0.045	−0.007	−0.005
S_6	3.452	−0.008	−0.011	+0.044	+0.040
S_7	3.469	+0.024	+0.023	+0.050	+0.047
S_8	3.519	+0.039	+0.042	+0.033	+0.038
S_9	4.000	+0.036	+0.038	−0.054	−0.051
S_{10}	4.021	+0.045	+0.042	+0.055	+0.053

[a] Number of environment subsystems

excitations because electrons are transferred into orbitals which are more diffuse and therefore interact more strongly with the environment.

The excitation energies of (**1**) are shown in Table VII. Here we used smaller radii for the environment because the molecular crystal is packed much more densely due to the lack of fluorocarbon side chains at the positions labeled with R in Figure 1. Although the number of molecules around the target subsystem is similar to the other two molecular crystals the excitation energies still show some variation between the two largest clusters. For example the fourth excitation energy still changes by −0.045 eV. The highest changes in excitation energy between the systems without environment and 40 surrounding molecules are 0.155 eV for S_4 and 0.120 eV for S_3. The origin of long range interactions in this system could be a larger polarization of the closest neighbors by theirs neighbors because of the smaller intermolecular distances in the crystal for this compound.

TAPP

TABLE VII: Environment-induced shifts for the first 10 excitation energies in eV of compound (**1**) for different sizes of the environment.

	ω	$\Delta_{\text{env}}\omega$			
radius	0Å	8Å	10Å	12Å	15Å
Env.[a]	0	4	10	20	40
S_1	3.240	+0.015	+0.016	+0.028	+0.028
S_2	3.299	+0.050	+0.039	+0.004	+0.001
S_3	3.555	−0.012	+0.020	+0.107	+0.120
S_4	3.585	+0.041	+0.168	+0.200	+0.155
S_5	4.164	+0.054	+0.072	+0.025	+0.016
S_6	4.199	+0.066	+0.069	+0.074	+0.066
S_7	4.276	+0.027	−0.008	+0.005	+0.002
S_8	4.346	−0.003	−0.022	−0.030	+0.015
S_9	4.398	+0.078	−0.040	+0.040	+0.035
S_{10}	4.562	−0.058	−0.034	−0.038	+0.016

[a] Number of environment subsystems

E. Embedded TAPP dimer

Finally, we studied the excitation energies for embedded TAPP dimers. Table VIII lists the lowest excitation energies for a dimer of (**2**). Here we see almost no changes between the last two radii and can therefore expect that the excitation energies are converged with respect to the size of the environment. Also the changes between the radii of 10 and 15Å are rather small compared to the changes seen with the monomer as target subsystem. The largest change in an excitation energy between these radii is seen for S_{10} with -0.027 eV. In contrast to the monomer as central system all excitation energies shown in Tables VIII and IX decrease already for the smallest embedding environment. Also, the environment-induced shifts are larger than for the monomer as active central system. While the maximum difference in excitation energy for the dimer system is -0.518 eV , the maximum difference or the monomer system is only 0.113 eV. Most of the changes in the excitation energies already occurs in the step from the isolated dimer

TABLE VIII: Environment-induced shifts for the first 10 excitation energies in eV of the dimer of compound (**2**) for different sizes of the environment.

	ω	$\Delta_{\text{env}}\omega$		
radius	0Å	10Å	15Å	20Å
Env.[a]	0	6	26	52
S_1	3.115	−0.183	−0.186	−0.187
S_2	3.120	−0.125	−0.133	−0.134
S_3	3.254	−0.127	−0.116	−0.114
S_4	3.311	−0.161	−0.151	−0.150
S_5	3.705	−0.509	−0.519	−0.518
S_6	3.707	−0.422	−0.432	−0.432
S_7	3.714	−0.097	−0.104	−0.106
S_8	3.753	−0.075	−0.089	−0.091
S_9	3.766	−0.064	−0.088	−0.086
S_{10}	3.977	−0.261	−0.288	−0.288

[a] Number of environment subsystems

and to six surrounding subsystems. Further changes are comparatively small. Thus, the greatest effect on the dimer system can be achieved with a relatively small environment. The difference in the converged excitation energies of the monomer and the dimer demonstrates the effect of the missing response of the environment within the FDEu approach. Similar converged excitation energies can only be achieved with the FDEc method.

Table IX shows the excitation energies for a dimer of compound (**3**) within its molecular crystal taking different sizes of the environment into account. Between the last two radii only minor changes up to 0.008 eV in the excitation energies. As for the dimer of compound (**2**) the largest changes in the excitation energies observed between the isolated dimer and the dimer embedded in the first environment shell of nearest neighbors. Also for compound (**3**) the excitation energies have all been lowered through the influence of the environment. The largest environment-induced shifts in the excitation energies are found for for S_1 and S_2 with changes of -0.591 eV and -0.624 eV.

TABLE IX: Environment-induced shifts for the first 10 excitation energies in eV of the dimer of compound (**3**) for different sizes of the environment.

	ω	$\Delta_{\text{env}}\omega$		
radius	0Å	10Å	15Å	20Å
Env.[a]	0	6	20	44
S_1	2.829	−0.582	−0.593	−0.591
S_2	2.876	−0.615	−0.626	−0.624
S_3	2.950	−0.141	−0.166	−0.164
S_4	2.950	−0.095	−0.122	−0.120
S_5	3.147	−0.208	−0.209	−0.209
S_6	3.147	−0.199	−0.202	−0.201
S_7	3.219	−0.098	−0.130	−0.122
S_8	3.220	−0.095	−0.125	−0.117
S_9	3.514	−0.286	−0.336	−0.332
S_{10}	3.517	−0.288	−0.337	−0.333

[a] Number of environment subsystems

V. CONCLUSION

Frozen density embedding offers a straight-forward, almost black-box way to reduce the costs for calculation of properties for systems that consists of several only weakly interacting molecules. Large reduction of computational costs can in particular be obtained when a system can be divided into small subsystems. Unfortunately, this is not always possible, for example in the case of complex chromophores, which can lead to large subsystems. This can in turn lead to long computational times when these subsystems are studied, even with second-order correlation methods as e.g. CIS(D), ADC(2), or CC2.

As a solution to this problem, we presented a combination of frozen density embedding with pair natural orbital based local correlation methods within the pnoccsd program of the TURBOMOLE package. The new extended FDE implementation is applicable for an arbitrary number of subsystems and evaluated the embedding potentials only within the basis of the active subsystem,

avoiding the calculation of integrals for the supermolecular basis. At the Hartree-Fock and DFT levels is has been implemented in the dscf and ridft programs. In the latter program it can be used together with the RI approximation. At the correlated wavefunction level it has been implemented for MP2, CIS(D), ADC(2) and CC2 in the ricc2 and pnoccsd programs that exploit the resolution-of-the-identity (RI) and the pair natural orbital (PNO) approximations. Excitation energies can be computed within the frozen solvent approximation or the post-SCF reaction field coupling.

We investigated the importance of the kernel contributions for excitation energies. The latter originate from the non-linear dependence of the interaction energy on the density of the embedded subsystem. It turned out that they are most of the time small.For the TAPPs that have been used in the current work to study the computational performance of the implementation, test calculations for compound (**2**) demonstrated that the kernel contributions are insignificant and can be neglected for TAPPs.

Furthermore we investigated the effect of the PNO truncation error on environment-induced shifts in the excitation energies of an embedded subsystem. For two TAPP systems we demonstrate that the truncation error on the environment-induced shifts in valence excitation energies are significantly smaller than the shift themselves and that, thus, the additional approximations made in the PNO-based implementation has only an insignificant effect on the accuracy of the predicted environment-induced shifts.

One aim of the current work has been to reduce the computational effort for the calculation of excitation energies in large systems consisting of several weakly interaction molecules. The computational performance for the combination FDE with PNO-based local methods was tested against conventional supermolecular calculations in Sec. IV C. The results demonstrate that large savings that can be achieved with this combination. In case of compound (**2**) the FDE calculation was 39 times faster than the supermolecular calculation for the largest system computed with both methods.

Finally, the behavior of the excitation energy with respect to the size of the embedding environment was investigated for subsystems consisting of either a monomer or a dimer. This showed that the excitation energies converge for the investigated TAPPS with a relatively small number of environment subsystems, for an embedded dimer with even less subsystems than for an embedded monomer.

Overall, we demonstrate the capabilities of this combination, both in terms of the overall size of the system and in terms of the size of individual subsystems and could show that the accuracy

of the environment-induced shift are not significantly affected by the PNO approximation.

ACKNOWLEDGMENTS

The authors thank TURBOMOLE GmbH for funding. C.H. acknowledges support by the Deutsche Forschungsgemeinschaft within the Priority Program SPP 1807 (HA 25888/10-1). Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2022 - 390677874 – RESOLV. S.H. acknowledges support by the Deutsche Forschungsgemeinschaft through SFB 1249 "N-Heteropolycycles as Functional Materials" (Project B07).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available within the article and its supplementary material.

REFERENCES

[1]S. Miertuš, E. Scrocco, and J. Tomasi, Chem. Phys. **55**, 117 (1981).

[2]A. Klamt and G. Schüürmann, J. Chem. Soc., Perkin Trans. 2 , 799 (1993).

[3]A. V. Marenich, C. J. Cramer, and D. G. Truhlar, J. Phys. Chem. B **113**, 6378 (2009).

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

[5]E. Brunk and U. Rothlisberger, Chem. Rev. **115**, 6217 (2015).

[6]H. M. Senn and W. Thiel, Angew. Chem., Int. Ed. **48**, 1198 (2009).

[7]J. M. Olsen, K. Aidas, and J. Kongsted, J. Chem. Theory Comput. **6**, 3721 (2010).

[8]T. Schwabe, K. Sneskov, J. M. Haugaard Olsen, J. Kongsted, O. Christiansen, and C. Hättig, J. Chem. Theory Comput. **8**, 3274 (2012).

[9]N. H. List, J. M. H. Olsen, and J. Kongsted, Phys. Chem. Chem. Phys. **18**, 20234 (2016).

[10]A. Marefat Khah, S. Karbalaei Khani, and C. Hättig, J. Chem. Theory Comput. **14**, 4640 (2018).

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

[12]C. R. Jacob and L. Visscher, J. Chem. Phys. **128**, 155102 (2008).

[13]J. Neugebauer, J. Chem. Phys. **126**, 134116 (2007).

[14]M. E. Casida and T. A. Wesołowski, Int. J. Quantum Chem. **96**, 577 (2004).

[15]F. Neese, F. Wennmohs, and A. Hansen, J. Chem. Phys. **130**, 114108 (2009).

[16]F. Neese, A. Hansen, and D. G. Liakos, J. Chem. Phys. **131**, 064103 (2009).

[17]S. Laricchia, E. Fabiano, and F. Della Sala, J. Chem. Phys. **133**, 164111 (2010).

[18]"Development version of TURBOMOLE 2020, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from `http://www.turbomole.org` (accessed May 13, 2022).".

[19]C. Hättig and F. Weigend, J. Chem. Phys. **113**, 5154 (2000).

[20]B. Helmich and C. Hättig, J. Chem. Phys. **135**, 214106 (2011).

[21]B. Helmich and C. Hättig, J. Chem. Phys. **139**, 084114 (2013).

[22]M. S. Frank and C. Hättig, J. Chem. Phys. **148**, 134102 (2018).

[23]G. Senatore and K. R. Subbaswamy, Phys. Rev. B **34**, 5754 (1986).

[24]M. D. Johnson, K. R. Subbaswamy, and G. Senatore, Phys. Rev. B **36**, 9202 (1987).

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

[26]R. S. Treß, C. Hättig, and S. Höfener, J. Chem. Theory Comput. **18**, 1737 (2022).

[27]T. A. Wesołowski and J. Weber, Chem. Phys. Lett. **248**, 71 (1996).

[28]S. Höfener, Int. J. Quantum Chem. **121**, e26351 (2021).

[29]F. del Valle and J. Tomasi, Chem. Phys. **150**, 139 (1991).

[30]M. Caricato, B. Mennucci, J. Tomasi, F. Ingrosso, R. Cammi, S. Corni, and G. Scalmani, J. Chem. Phys. **124**, 124520 (2006).

[31]M. Caricato, J. Chem. Phys. **135**, 074113 (2011).

[32]K. Sneskov, T. Schwabe, J. Kongsted, and O. Christiansen, J. Chem. Phys. **134**, 104108 (2011).

[33]S. Karbalaei Khani, A. Marefat Khah, and C. Hättig, Phys. Chem. Chem. Phys. **20**, 16354 (2018).

[34]S. Höfener, A. Severo Pereira Gomes, and L. Visscher, J. Chem. Phys. **136**, 044104 (2012).

[35]S. Höfener, A. S. P. Gomes, and L. Visscher, J. Chem. Phys. **139**, 104106 (2013).

[36]O. Christiansen and K. V. Mikkelsen, J. Chem. Phys. **110**, 8348 (1999).

[37]R. Cammi, J. Chem. Phys. **131**, 164104 (2009).

[38]J. Kongsted, A. Osted, K. V. Mikkelsen, and O. Christiansen, J. Chem. Phys. **118**, 1620 (2003).

[39]S. K. Khani, R. Faber, F. Santoro, C. Hättig, and S. Coriani, J. Chem. Theory Comput. **15**, 1242 (2019).

[40]O. Christiansen, H. Koch, and P. Jørgensen, Chem. Phys. Lett. **243**, 409 (1995).

[41]J. Schirmer, Phys. Rev. A **26**, 2395 (1982).

[42]A. B. Trofimov and J. Schirmer, J. Phys. B: At. Mol. Opt. Phys **28**, 2299 (1995).

[43]M. Head-Gordon, M. Oumi, and D. Maurice, Mol. Phys. **96**, 593 (1999).

[44]C. Hättig, Adv. Quantum Chem. **50**, 37 (2005).

[45]M. Head-Gordon, R. J. Rico, M. Oumi, and T. J. Lee, Chem. Phys. Lett. **219**, 21 (1994).

[46]M. Matena, M. Stohr, T. Riehm, J. Bjork, S. Martens, M. S. Dyer, M. Persson, J. Lobo-Checa, K. Muller, M. Enache, H. Wadepohl, J. Zegenhagen, T. A. Jung, and L. H. Gade, "Ccdc 747374: Experimental crystal structure determination," (2010).

[47]M. Matena, M. Stöhr, T. Riehm, J. Björk, S. Martens, M. S. Dyer, M. Persson, J. Lobo-Checa, K. Müller, M. Enache, H. Wadepohl, J. Zegenhagen, T. A. Jung, and L. H. Gade, Chem. Eur. J **16**, 2079 (2010).

[48]B. A. R. Günther, S. Höfener, R. Eichelmann, U. Zschieschang, H. Wadepohl, H. Klauk, and L. H. Gade, "Ccdc 1980431: Experimental crystal structure determination," (2020).

[49]B. A. R. Günther, S. Höfener, R. Eichelmann, U. Zschieschang, H. Wadepohl, H. Klauk, and L. H. Gade, Org. Lett. **22**, 2298 (2020).

[50]B. A. R. Günther, S. Höfener, U. Zschieschang, H. Wadepohl, H. Klauk, and L. H. Gade, "Ccdc 1908781: Experimental crystal structure determination," (2020).

[51]B. A. R. Günther, S. Höfener, U. Zschieschang, H. Wadepohl, H. Klauk, and L. H. Gade, Chem. Eur. J **25**, 14669 (2019).

[52]T. D. Kühne, M. Iannuzzi, M. Del Ben, V. V. Rybkin, P. Seewald, F. Stein, T. Laino, R. Z. Khaliullin, O. Schütt, F. Schiffmann, D. Golze, J. Wilhelm, S. Chulkov, M. H. Bani-Hashemian, V. Weber, U. Borštnik, M. Taillefumier, A. S. Jakobovits, A. Lazzaro, H. Pabst, T. Müller, R. Schade, M. Guidon, S. Andermatt, N. Holmberg, G. K. Schenter, A. Hehn, A. Bussy, F. Belleflamme, G. Tabacchi, A. Glöß, M. Lass, I. Bethune, C. J. Mundy, C. Plessl, M. Watkins, J. VandeVondele, M. Krack, and J. Hutter, J. Chem. Phys. **152**, 194103 (2020).

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

[54]M. Ernzerhof and G. E. Scuseria, J. Chem. Phys. **110**, 5029 (1999).

[55]J. VandeVondele and J. Hutter, J. Chem. Phys. **127**, 114105 (2007).

[56]M. Krack, Theor. Chem. Acc. **114**, 145 (2005).

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

[58]M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, and G. R. Hutchison, J. Cheminform. **4**, 17 (2012).

[59]W. Humphrey, A. Dalke, and K. Schulten, J. Mol. Grap. **14**, 33 (1996).

[60]J. Liu and S. Höfener, “manuscript in preparation,”.

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

[62]F. Weigend, Phys. Chem. Chem. Phys. **8**, 1057 (2006).

[63]K. Eichkorn, O. Treutler, H. Öhm, M. Häser, and R. Ahlrichs, Chem. Phys. Lett. **242**, 652 (1995).

[64]S. Laricchia, E. Fabiano, L. A. Constantin, and F. Della Sala, J. Chem. Theory Comput. **7**, 2439 (2011).

[65]A. Marefat Khah, P. Reinholdt, J. M. H. Olsen, J. Kongsted, and C. Hättig, J. Chem. Theory Comput. **16**, 1373 (2020).

[66]G. Knizia, J. Chem. Theory Comput. **9**, 4834 (2013).

[67]A. Hellweg, C. Hättig, S. Höfener, and W. Klopper, Theor. Chem. Acc. **117**, 587 (2007).

[68]S. Corni, R. Cammi, B. Mennucci, and J. Tomasi, J. Chem. Phys. **123**, 134512 (2005).

[69]R. A. Kendall, T. H. Dunning, and R. J. Harrison, J. Chem. Phys. **96**, 6796 (1992).

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

[71]A. Lembarki and H. Chermette, Phys. Rev. A **50**, 5328 (1994).

[72]K. Momma and F. Izumi, J. Appl. Crystallogr. **41**, 653 (2008).

Chapter 5

Conclusions and Outlook

In the course of this PhD project a FDE implementation was developed that can be used in combination with the PNO approximation and works exclusively in the monomolecular basis, i.e. avoids any calculations with the supermolecular basis sets. The starting point of the project was the implementation by Laricchia et al.[133] within the TURBOMOLE package. The implementation was limited to two subsystem, could only be used with the local HF and DFT method within the *odft* and *dscf* programs, could only make use of the OpenMP parallelization scheme and needed supermolecular SCF iterations to compute the embedding potential. An extension of this implementation to an arbitrary number of subsystems turned out to be too computational demanding for large systems and an implementation which used only the monomolecular basis of each subsystem was aspired. The FDE implementation within the KOALA program served here as a model implementation, because second-order correlation methods could be used and the embedding potential is calculated using only the monomolecular basis. Nevertheless the KOALA implementation is limited to canonical orbitals and the OpenMP parallelization scheme. The new implementation was integrated into the *dscf*, *ridft*, *ricc2* and *pnoccsd* programs of the TURBOMOLE package. This allows to perform FDE calculations with HF, DFT or second-order correlation methods for the individual subsystems while exploiting the RI and PNO approximation. In addition the implementation makes use of the OpenMP and MPI parallelization schemes allowing to use various processor architectures.

As part of the project also the electron spill-out problem in ADC(2) excitation energy calculations of chromophores within various environments was addressed. For this a testset consisting of eleven different systems, where most of them have been known to suffer from electron spill-out if modeled with the PE method[146], was used to investigate the errors in the excitation energies compared to a subsystem-ADC(2) reference calculation. It could be demonstrated that some systems still suffer from electron spill-out even with the FDE method, although the nonadditive kinetic energy contributions provide an additional repulsive term in the embedding potential. By comparison with the PE method it was shown that the FDE method could prevent the electron spill-out for some excitations and also lower the error caused by this problem, but nevertheless errors up to 1.84 eV occurred. Following a similar approach as used for the PE method, all-electron pseudopotentials where added to the embedding potential to model the Pauli repulsion between subsystems. With this so-called FDE(ECP) approach the maximum error could be reduced to 0.18 eV and the mean absolute deviation (MAD) from 0.27 eV to 0.05 eV. Since the FDE approach has compared to the PE approach an additional repulsive contribution through the non-additive kinetic energy potential, the FDE(ECP) approach bears the risk of double counting repulsive contributions when used with all-electron pseudopotentials developed for the PE(ECP) approach. However a systematic error of the FDE(ECP) approach compared to the PE(ECP) method could not be found and therefore a double counting does not seem to be an issue with the proposed ECPs.

The efficiency and accuracy of the proposed combination of PNOs with FDE was tested in the second publication of this work. Here the excitation energies of one or two TAPP molecules within their molecular crystals were calculated for different numbers of environment molecules and their computational wall times determined. This demonstrated the capabilities of the combination of FDE with PNOs, since excitation energies of systems with up to 2592 atoms could be determined. Also the huge time savings compared to supermolecular calculations were demonstrated. For one of the largest clusters investigated with a supermolecular approach the FDE(ECP)-PNO calculation was 39 times faster. In addition, the scaling of the FDE implementation with the system

size was determined, showing a sublinear scaling for small systems, because of the huge prefactor caused by the excitation energy calculation, and a scaling with N^2 with the number of basis functions for larger systems representing the true scaling of the FDE implementation. Beside the efficiency also the accuracy of the results were investigated with respect to the truncation error caused by the PNO approximation. Here the error caused by the truncation of the PNO expansion on the environment induced shift was of special interest since this is a very important property within the framework of theoretical spectroscopy and is usually also rather small. While the truncation errors of the excitation energies themselves were similar in magnitude to the environment induced shifts, the truncation error in the shifts were an order of magnitude smaller than the shifts themselves, which ensures a qualitative correct description of the environment induced shifts even in combination with PNOs. In addition to the accuracy with respect to the truncation error also the convergence with the environment size was investigated for monomer and dimer TAPPs and could be verified.

Supported by the results of the presented study the implementation offers the opportunity to calculate excitation energies of large chromophores, like a TAPP dimer, with accurate second-order correlation methods within large and complex environments without loosing the accuracy to model environment induced shifts. Together with all-electron pseudopotentials the implementation also ensures a robustness against the electron spill-out problem.

The efficiency of this implementation could in future be even further improved with respect to large systems in two points. First thing is the N^2 scaling with respect to the size of the environment, which currently still caused by the evaluation of the electrostatic potential on the supermolecular integration grid. This could be improved using the fast multipole method (FMM) to evaluate the far-field of the electrostatic potential, which should result in a near linear scaling of the FDE implementation in combination with the RI approximation[147] (MARIJ). Another possibility to increase the efficiency of this implementation is to change the current parallelization scheme of the FDE approach. At the moment each subsystem is computed using all available processes and threads until its calculation is finished and only then the next subsystem will be calculated. In case of systems containing small subsystems, i.e. one water molecule,

a high parallelization of such a subsystem does not lead to any improvement and may even take more time than using less resources. For such cases it would be more efficient to reduce the number of resources used for small subsystems and use the saved resources to compute the density of another subsystem at the same time. Such a parallelization scheme would make it necessary to change the current implemented FDE procedure, such that the densities used to calculate the embedding potentials are only updated after a FaT iteration is completed. This ensures that the embedding potential does not depend on the update order of the densities, which would lead to different interim results for different update sequences. The current update scheme was chosen to be consistent with the KOALA program to provide comparable results for canonical CC2 calculations and therefore the subsystem densities are updated in the same order in which they are defined. For such a update scheme it is advisable to define the subsystem of the greatest interest as the last subsystem, so that its embedding potential is the most accurate. A change of the density update scheme to the one proposed above would also get rid of this dependency and the embedding potentials of all subsystems are equivalently accurate.

Chapter 6

Bibliography

[1] Severo Pereira Gomes, A.; Jacob, C. R. *Annu. Rep. Prog. Chem., Sect. C: Phys. Chem.* **2012**, *108*, 222–277.

[2] Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027–2094.

[3] Senn, H. M.; Thiel, W. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198–1229.

[4] Jacob, C. R.; Neugebauer, J. *Wiley Interdiscip. Rev.-Comput. Mol. Sci.* **2014**, *4*, 325–362.

[5] Wesołowski, T. A.; Shedge, S.; Zhou, X. *Chem. Rev.* **2015**, *115*, 5891–5928.

[6] Miertuš, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117–129.

[7] Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799–805.

[8] Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2009**, *113*, 6378–6396.

[9] Pedone, A.; Biczysko, M.; Barone, V. *ChemPhysChem* **2010**, *11*, 1812–1832.

[10] Mennucci, B.; Cappelli, C.; Guido, C. A.; Cammi, R.; Tomasi, J. *J. Phys. Chem. A* **2009**, *113*, 3009–3020.

[11] Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227–249.

[12] Thompson, M. A. *J. Phys. Chem.* **1996**, *100*, 14492–14507.

[13] Gao, J. *J. Comput. Chem.* **1997**, *18*, 1061–1071.

[14] Zhang, Y.; Lin, H.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 1378–1398.

[15] Mortier, W. J.; Van Genechten, K.; Gasteiger, J. *J. Am. Chem. Soc.* **1985**, *107*, 829–835.

[16] Anisimov, V. M.; Lamoureux, G.; Vorobyov, I. V.; Huang, N.; Roux, B.; MacKerell, A. D. *J. Chem. Theory Comput.* **2005**, *1*, 153–168.

[17] Olsen, J. M.; Aidas, K.; Kongsted, J. *J. Chem. Theory Comput.* **2010**, *6*, 3721–3734.

[18] Sneskov, K.; Schwabe, T.; Kongsted, J.; Christiansen, O. *J. Chem. Phys.* **2011**, *134*, 104108.

[19] Schwabe, T.; Olsen, J. M. H.; Sneskov, K.; Kongsted, J.; Christiansen, O. *J. Chem. Theory Comput.* **2011**, *7*, 2209–2217.

[20] Wesołowski, T. A.; Warshel, A. *J. Phys. Chem.* **1993**, *97*, 8050–8053.

[21] Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133–A1138.

[22] Wesołowski, T. A. *Chem. Phys. Lett.* **1999**, *311*, 87–92.

[23] Wesołowski, T.; Muller, R. P.; Warshel, A. *J. Phys. Chem.* **1996**, *100*, 15444–15449.

[24] Wesołowski, T.; Warshel, A. *J. Phys. Chem.* **1994**, *98*, 5183–5187.

[25] Hong, G.; Strajbl, M.; Wesolowski, T. A.; Warshel, A. *J. Comput. Chem.* **2000**, *21*, 1554–1561.

[26] Wesołowski, T. A.; Weber, J. *Chem. Phys. Lett.* **1996**, *248*, 71–76.

[27] Zbiri, M.; Atanasov, M.; Daul, C.; Garcia-Lastra, J. M.; Wesolowski, T. A. *Chem. Phys. Lett.* **2004**, *397*, 441–446.

[28] Zbiri, M.; Daul, C. A.; Wesolowski, T. A. *J. Chem. Theory Comput.* **2006**, *2*, 1106–1111.

[29] Galván, I.; Sánchez, M.; Martín, M.; Olivares del Valle, F.; Aguilar, M. *Comput. Phys. Commun.* **2003**, *155*, 244–259.

[30] Coutinho, K.; Georg, H.; Fonseca, T.; Ludwig, V.; Canuto, S. *Chem. Phys. Lett.* **2007**, *437*, 148–152.

[31] Thomas, L. H. The calculation of atomic fields. Math. Proc. Camb. Philos. Soc. 1927; pp 542–548.

[32] Fermi, E. *Rend. Accad. Naz. Lincei* **1927**, *6*, 5.

[33] Weizsäcker, C. v. *Z. Phys.* **1935**, *96*, 431–458.

[34] Wesołowski, T. A. *Computational Chemistry: Reviews of Current Trends*; 2006; pp 1–82.

[35] Lee, H.; Lee, C.; Parr, R. G. *Phys. Rev. A* **1991**, *44*, 768–771.

[36] March, N. H.; Santamaria, R. *Int. J. Quantum Chem.* **1991**, *39*, 585–592.

[37] Lembarki, A.; Chermette, H. *Phys. Rev. A* **1994**, *50*, 5328.

[38] Wang, Y. A.; Govind, N.; Carter, E. A. *Phys. Rev. B* **1998**, *58*, 13465–13471.

[39] Choly, N.; Kaxiras, E. *Solid State Commun.* **2002**, *121*, 281–286.

[40] García-Aldea, D.; Alvarellos, J. E. *J. Chem. Phys.* **2008**, *129*, 074103.

[41] Wesołowski, T. A.; Weber, J. *Int. J. Quantum Chem.* **1997**, *61*, 303–311.

[42] Jacob, C. R.; Beyhan, S. M.; Visscher, L. *J. Chem. Phys.* **2007**, *126*, 234116.

[43] van Leeuwen, R.; Baerends, E. J. *Phys. Rev. A* **1994**, *49*, 2421–2431.

[44] Kadantsev, E. S.; Stott, M. J. *Phys. Rev. A* **2004**, *69*, 012502.

[45] Wang, Y.; Parr, R. G. *Phys. Rev. A* **1993**, *47*, R1591–R1593.

[46] Colonna, F.; Savin, A. *J. Chem. Phys.* **1999**, *110*, 2828–2835.

[47] Zhao, Q.; Morrison, R. C.; Parr, R. G. *Phys. Rev. A* **1994**, *50*, 2138–2142.

[48] Yang, W.; Wu, Q. *Phys. Rev. Lett.* **2002**, *89*, 143002.

[49] Heaton-Burgess, T.; Bulat, F. A.; Yang, W. *Phys. Rev. Lett.* **2007**, *98*, 256401.

[50] King, R. A.; Handy, N. C. *Phys. Chem. Chem. Phys.* **2000**, *2*, 5049–5056.

[51] Jacob, C. R. *The Journal of Chemical Physics* **2011**, *135*, 244102.

[52] Govind, N.; Wang, Y.; da Silva, A.; Carter, E. *Chem. Phys. Lett.* **1998**, *295*, 129–134.

[53] Govind, N.; Wang, Y. A.; Carter, E. A. *J. Chem. Phys.* **1999**, *110*, 7677–7688.

[54] Klüner, T.; Govind, N.; Wang, Y. A.; Carter, E. A. *Phys. Rev. Lett.* **2001**, *86*, 5954–5957.

[55] Höfener, S.; Severo Pereira Gomes, A.; Visscher, L. *J. Chem. Phys.* **2012**, *136*, 044104.

[56] Höfener, S.; Severo Pereira Gomes, A.; Visscher, L. *J. Chem. Phys.* **2013**, *139*, 104106.

[57] Wesołowski, T. A. *Phys. Rev. A* **2008**, *77*, 012504.

[58] Jacob, C. R.; Visscher, L. *J. Chem. Phys.* **2006**, *125*, 194104.

[59] Jacob, C. R.; Neugebauer, J.; Visscher, L. *J. Comput. Chem.* **2008**, *29*, 1011–1018.

[60] Neugebauer, J.; Louwerse, M. J.; Belanzoni, P.; Wesolowski, T. A.; Baerends, E. J. *J. Chem. Phys.* **2005**, *123*, 114101.

[61] García-Lastra, J.; Wesolowski, T.; Barriuso, M.; Aramburu, J.; Moreno, M. *J. Condens. Matter Phys.* **2006**, *18*, 1519.

[62] Casida, M. E.; Wesołowski, T. A. *Int. J. Quantum Chem.* **2004**, *96*, 577–588.

[63] Wesolowski, T. A. *J. Am. Chem. Soc.* **2004**, *126*, 11444–11445.

[64] Neugebauer, J. *J. Chem. Phys.* **2007**, *126*, 134116.

[65] Fradelos, G.; Kaminski, J. W.; Wesolowski, T. A.; Leutwyler, S. *J. Phys. Chem. A* **2009**, *113*, 9766–9771.

[66] König, C.; Schlüter, N.; Neugebauer, J. *J. Chem. Phys.* **2013**, *138*, 034104.

[67] Pavanello, M. *J. Chem. Phys.* **2013**, *138*, 204118.

[68] Neugebauer, J.; Louwerse, M. J.; Baerends, E. J.; Wesolowski, T. A. *J. Chem. Phys.* **2005**, *122*, 094115.

[69] Neugebauer, J.; Jacob, C. R.; Wesolowski, T. A.; Baerends, E. J. *J. Phys. Chem. A* **2005**, *109*, 7805–7814.

[70] Kovyrshin, A.; Neugebauer, J. *Chem. Phys.* **2011**, *391*, 147–156.

[71] Neugebauer, J.; Baerends, E. J. *J. Phys. Chem. A* **2006**, *110*, 8786–8796.

[72] König, C.; Neugebauer, J. *ChemPhysChem* **2012**, *13*, 386–425.

[73] Neugebauer, J. *ChemPhysChem* **2009**, *10*, 3148–3173.

[74] Neugebauer, J. *Phys. Rep.* **2010**, *489*, 1–87.

[75] Severo Pereira Gomes, A.; Jacob, C. R.; Visscher, L. *Phys. Chem. Chem. Phys.* **2008**, *10*, 5353–5362.

[76] Khait, Y. G.; Hoffmann, M. R. *J. Chem. Phys.* **2010**, *133*, 044107.

[77] Daday, C.; König, C.; Valsson, O.; Neugebauer, J.; Filippi, C. *J. Chem. Theory Comput.* **2013**, *9*, 2355–2367.

[78] Dułak, M.; Wesolowski, T. A. *J. Chem. Phys.* **2006**, *124*, 164101.

[79] Fradelos, G.; Wesolowski, T. A. *J. Phys. Chem. A* **2011**, *115*, 10018–10026.

[80] Chulhai, D. V.; Jensen, L. *J. Chem. Theory Comput.* **2015**, *11*, 3080–3088.

[81] Manby, F. R.; Stella, M.; Goodpaster, J. D.; Miller III, T. F. *J. Chem. Theory Comput.* **2012**, *8*, 2564–2568.

[82] Chulhai, D. V.; Jensen, L. *Phys. Chem. Chem. Phys.* **2016**, *18*, 21032–21039.

[83] Tölle, J.; Böckers, M.; Neugebauer, J. *J. Chem. Phys.* **2019**, *150*, 181101.

[84] Tölle, J.; Böckers, M.; Niemeyer, N.; Neugebauer, J. *J. Chem. Phys.* **2019**, *151*, 174109.

[85] Niemeyer, N.; Tölle, J.; Neugebauer, J. *J. Chem. Theory Comput.* **2020**, *16*, 3104–3120.

[86] Scholz, L.; Tölle, J.; Neugebauer, J. *Int. J. Quantum Chem.* **2020**, *120*, e26213.

[87] Pulay, P. *Chem. Phys. Lett.* **1983**, *100*, 151–154.

[88] Christiansen, O.; Manninen, P.; Jørgensen, P.; Olsen, J. *J. Chem. Phys.* **2006**, *124*, 084103.

[89] Yang, J.; Kurashige, Y.; Manby, F. R.; Chan, G. K. L. *J. Chem. Phys.* **2011**, *134*, 044123.

[90] Yang, J.; Chan, G. K.-L.; Manby, F. R.; Schütz, M.; Werner, H.-J. *J. Chem. Phys.* **2012**, *136*, 144105.

[91] Meyer, W. *Int. J. Quantum Chem.* **1971**, *5*, 341–348.

[92] Meyer, W. *J. Chem. Phys.* **1973**, *58*, 1017–1035.

[93] Meyer, W. *Theoret. Chim. Acta* **1974**, *35*, 277–292.

[94] Meyer, W.; Rosmus, P. *J. Chem. Phys.* **1975**, *63*, 2356–2375.

[95] Hay, P. J. *J. Chem. Phys.* **1973**, *59*, 2468–2476.

[96] Ahlrichs, R.; Lischka, H.; Staemmler, V.; Kutzelnigg, W. *J. Chem. Phys.* **1975**, *62*, 1225–1234.

[97] Ahlrichs, R.; Driessler, F.; Lischka, H.; Staemmler, V.; Kutzelnigg, W. *J. Chem. Phys.* **1975**, *62*, 1235–1247.

[98] Ahlrichs, R.; Keil, F.; Lischka, H.; Kutzelnigg, W.; Staemmler, V. *J. Chem. Phys.* **1975**, *63*, 455–463.

[99] Ahlrichs, R.; Lischka, H.; Zurawski, B.; Kutzelnigg, W. *J. Chem. Phys.* **1975**, *63*, 4685–4694.

[100] Ahlrichs, R.; Driessler, F. *Theoret. Chim. Acta* **1975**, *36*, 275–287.

[101] Meyer, W.; Hariharan, P. C.; Kutzelnigg, W. *J. Chem. Phys.* **1980**, *73*, 1880–1897.

[102] Staemmler, V.; Jaquet, R. *Theoret. Chim. Acta* **1981**, *59*, 487–500.

[103] Taylor, P. R. *J. Chem. Phys.* **1981**, *74*, 1256–1270.

[104] Neese, F.; Wennmohs, F.; Hansen, A. *J. Chem. Phys.* **2009**, *130*, 114108.

[105] Neese, F.; Hansen, A.; Liakos, D. G. *J. Chem. Phys.* **2009**, *131*, 064103.

[106] Vahtras, O.; Almlöf, J.; Feyereisen, M. *Chem. Phys. Lett.* **1993**, *213*, 514–518.

[107] Feyereisen, M.; Fitzgerald, G.; Komornicki, A. *Chem. Phys. Lett.* **1993**, *208*, 359–363.

[108] Tew, D. P.; Helmich, B.; Hättig, C. *J. Chem. Phys.* **2011**, *135*, 074107.

[109] Tew, D. P.; Hättig, C. *Int. J. Quantum Chem.* **2013**, *113*, 224–229.

[110] Schmitz, G.; Hättig, C.; Tew, D. P. *Phys. Chem. Chem. Phys.* **2014**, *16*, 22167–22178.

[111] Helmich, B.; Hättig, C. *J. Chem. Phys.* **2011**, *135*, 214106.

[112] Helmich, B.; Hättig, C. *Computational and Theoretical Chemistry* **2014**, *1040-1041*, 35–44.

[113] Helmich, B.; Hättig, C. *J. Chem. Phys.* **2013**, *139*, 084114.

[114] Frank, M. S.; Schmitz, G.; Hättig, C. *Mol. Phys.* **2017**, *115*, 343–356.

[115] Frank, M. S.; Schmitz, G.; Hättig, C. *J. Chem. Phys.* **2020**, *153*, 034109.

[116] Krause, C.; Werner, H.-J. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7591–7604.

[117] Schmitz, G.; Helmich, B.; Hättig, C. *Mol. Phys.* **2013**, *111*, 2463–2476.

[118] Bernard, Y. A.; Dułak, M.; Kamiński, J. W.; Wesołowski, T. A. *J. Phys. A: Math. Theor.* **2008**, *41*, 055302.

[119] Wesołowski, T. A.; Chermette, H.; Weber, J. *J. Chem. Phys.* **1996**, *105*, 9182–9190.

[120] Wesołowski, T. A. *J. Chem. Phys.* **1997**, *106*, 8516–8526.

[121] de Silva, P.; Wesolowski, T. A. *The Journal of Chemical Physics* **2012**, *137*, 094110.

[122] Götz, A. W.; Beyhan, S. M.; Visscher, L. *J. Chem. Theory Comput.* **2009**, *5*, 3161–3174.

[123] Fux, S.; Jacob, C. R.; Neugebauer, J.; Visscher, L.; Reiher, M. *J. Chem. Phys.* **2010**, *132*, 164101.

[124] Beyhan, S. M.; Götz, A. W.; Jacob, C. R.; Visscher, L. *J. Chem. Phys.* **2010**, *132*, 044114.

[125] Jacob, C. R.; Visscher, L. *J. Chem. Phys.* **2008**, *128*, 155102.

[126] Ángyán, J. G. *Chem. Phys. Lett.* **1995**, *241*, 51–56.

[127] Schwabe, T.; Sneskov, K.; Haugaard Olsen, J. M.; Kongsted, J.; Christiansen, O.; Hättig, C. *J. Chem. Theory Comput.* **2012**, *8*, 3274–3283.

[128] Karbalaei Khani, S.; Marefat Khah, A.; Hättig, C. *Phys. Chem. Chem. Phys.* **2018**, *20*, 16354–16363.

[129] Marefat Khah, A.; Karbalaei Khani, S.; Hättig, C. *J. Chem. Theory Comput.* **2018**, *14*, 4640–4650.

[130] Corni, S.; Cammi, R.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **2005**, *123*, 134512.

[131] Lunkenheimer, B.; Köhn, A. *J. Chem. Theory Comput.* **2013**, *9*, 977–994.

[132] Karbalaei Khani, S.; Marefat Khah, A.; Hättig, C. *J. Chem. Theory Comput.* **2020**, *16*, 4554–4564.

[133] Laricchia, S.; Fabiano, E.; Della Sala, F. *J. Chem. Phys.* **2010**, *133*, 164111.

[134] Constantin, L. A.; Fabiano, E.; Laricchia, S.; Della Sala, F. *Phys. Rev. Lett.* **2011**, *106*, 186406.

[135] Laricchia, S.; Fabiano, E.; Constantin, L. A.; Della Sala, F. *J. Chem. Theory Comput.* **2011**, *7*, 2439–2451.

[136] Laricchia, S.; Constantin, L. A.; Fabiano, E.; Della Sala, F. *J. Chem. Theory Comput.* **2014**, *10*, 164–179.

[137] Śmiga, S.; Fabiano, E.; Constantin, L. A.; Della Sala, F. *J. Chem. Phys.* **2017**, *146*, 064105.

[138] Laricchia, S.; Fabiano, E.; Sala, F. D. *J. Chem. Phys.* **2012**, *137*, 014102.

[139] Śmiga, S.; Fabiano, E.; Laricchia, S.; Constantin, L. A.; Della Sala, F. *J. Chem. Phys.* **2015**, *142*, 154121.

[140] Laricchia, S.; Fabiano, E.; Sala, F. D. *J. Chem. Phys.* **2013**, *138*, 124112.

[141] Laricchia, S.; Fabiano, E.; Della Sala, F. *Chem. Phys. Lett.* **2011**, *518*, 114–118.

[142] Fabiano, E.; Laricchia, S.; Sala, F. D. *J. Chem. Phys.* **2014**, *140*, 114101.

[143] Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.

[144] Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865.

[145] Höfener, S. *Int. J. Quantum Chem.* **2021**, *121*, e26351.

[146] Marefat Khah, A.; Reinholdt, P.; Olsen, J. M. H.; Kongsted, J.; Hättig, C. *J. Chem. Theory Comput.* **2020**, *16*, 1373–1381.

[147] Sierka, M.; Hogekamp, A.; Ahlrichs, R. *J. Chem. Phys.* **2003**, *118*, 9136–9148.

Appendix A

Supporting Informations

A.1 Tackling the electron spill-out problem

The following supporting information belongs to the publication *J. Chem. Theory Comput.* 2022, **18**, 3, 1737–1747; **Copyright © 2022 The Authors. Published by American Chemical Society**. The publication is available online: `https://doi.org/10.1021/acs.jctc.1c00732` and presented in chapter 3.

Employing pseudopotentials to tackle excited-state electron spill-out in Frozen Density Embedding Calculations

Robert S. Treß,[*,†] Christof Hättig,[†] and Sebastian Höfener[*,‡]

Department of Theoretical Chemistry, Ruhr University Bochum, Bochum, and Institute of Physical Chemistry, Karlsruhe Institute of Technology (KIT), 76131 Karlsruhe, Germany

E-mail: robert.tress@rub.de; hoefener@kit.edu

[*]To whom correspondence should be addressed
[†]Department of Theoretical Chemistry, Ruhr University Bochum, Bochum
[‡]Institute of Physical Chemistry, Karlsruhe Institute of Technology (KIT), 76131 Karlsruhe, Germany

1 Parameterized ECPs of Ar, Kr and Xe

Table 1: All-electron ECP Ar ecp-all, fitted to reproduce with zero electrons the orbital energies of the 4s4p3d5s5p4d shells in the aug-cc-pV(6+d)Z basis.

ncore=	18	lmax=	3
f	0.00	2	1.0000
s-f	275.00	2	1.6410
p-f	1.90	2	0.2733
d-f	-3.40	2	0.4400

Table 2: All-electron ECP Kr ecp-all, fitted to reproduce with zero electrons the orbital energies of the 5s5p4d6s6p4f5d shells in the aug-cc-pV5Z basis.

ncore=	36	lmax=	4
g	0.00	2	1.00000
s-g	158.20	2	1.20000
p-g	9.40	2	0.49000
d-g	-78.30	2	5.95000
f-g	-53.00	2	2.00000

Table 3: Xe ecp-all. All-electron ECP were fitted to reproduce with zero electrons the orbital energies of the 6s6p4f5d7s7p shells in the aug-cc-pV5Z-PP basis together with the defpp-ecp[1] ECP. For each type of potential, a pair of coefficient and exponent was added, which is shown in the respective last line.

ncore=	54	lmax=	4
g	0.00000000	2	1.0000000
s-g	49.99796200	2	40.00518400
	281.01330300	2	17.81221400
	61.53825500	2	9.30415000
	10.70000000	2	0.52000000
p-g	67.43914200	2	15.70177200
	134.87471100	2	15.25860800
	14.66330000	2	9.29218400
	29.35473000	2	8.55900300
	50.40000000	2	2.00000000
d-g	35.43690800	2	15.18560000
	53.19577200	2	14.28450000
	9.04623200	2	7.12188900
	13.22368100	2	6.99196300
	0.08485300	2	0.62394600
	0.04415500	2	0.64728400
	-18.20000000	2	1.46000000
f-g	-23.08929500	2	20.88155700
	-30.07447500	2	20.78344300
	-0.28822700	2	5.25338900
	-0.38692400	2	5.36118800
	-109.10000000	2	2.20000000

2 HOMO LUMO Tables

Table 4: Orbital-energy differences in meV of HOMO and LUMO between FDE and FDE(ECP) using the cc-pVTZ basis set for the chromophore.

System	ΔHOMO	ΔLUMO	ΔΔHL
Formamid/H_2O	+273.5	+236.6	-36.9
Berenil/DNA	-47.4	+3.2	+50.6
PNP^-/Na^+(aq)	+647.3	+675.0	+27.6
Cholesterol/cy-C_6H_{12}	+63.4	+131.2	+67.8
Aniline-yellow/H_2O	+83.1	+23.5	-59.6
Acrolein/C_2H_3N	+9.0	+49.6	+40.6
DMABN/CB[7]	+80.6	+104.8	+24.3
GFP/Protein	+376.5	+303.0	-73.5
DBO/CB[7]	+46.8	+124.8	+77.9
Pyrazine/DMI-Br	+122.4	+168.3	+45.9
Pyrazine/DMI-I	+10.3	+25.4	+15.1

Table 5: Orbital-energy differences in meV of HOMO and LUMO between FDE and FDE(ECP) using the aug-cc-pVTZ basis set for the chromophore.

System	ΔHOMO	ΔLUMO	ΔΔHL
Formamid/H_2O	+243.1	+3527.2	+3284.1
Berenil/DNA	-67.8	+269.1	+336.9
PNP^-/Na^+(aq)	+525.2	+1869.0	+1343.8
Cholesterol/cy-C_6H_{12}	+22.2	+763.5	+741.2
Aniline-yellow/H_2O	+61.8	+1968.4	+1906.6
Acrolein/C_2H_3N	+28.4	+2415.7	+2387.3
DMABN/CB[7]	+52.2	+1686.0	+1633.8
GFP/Protein	+359.0	+1869.1	+1510.2
DBO/CB[7]	+43.7	+3290.8	+3247.1
Pyrazine/DMI-Br	+124.1	+1204.9	+1080.8
Pyrazine/DMI-I	+37.2	+826.8	+789.5

3 Excitation Energies

3.1 Subsystem-ADC(2)

Table 6: First and second excitation energies in eV computed with the Subsystem-ADC(2) approach. Excitation energies computed with the cc-pVTZ and aug-cc-pVTZ basis set are denoted "TZ" and "aTZ", respectively.

System	ΔE_{S1-TZ}	ΔE_{S2-TZ}	ΔE_{S1-aTZ}	ΔE_{S2-aTZ}
Formamid/H_2O	5.729	7.393	5.674	7.202
Berenil/DNA	2.758	3.102	2.737	3.082
PNP^-/Na^+(aq)	3.007	3.566	3.016	3.572
Cholesterol/cy-C_6H_{12}	7.119	7.194	6.363	6.452
Aniline-yellow/H_2O	2.987	4.232	2.967	4.189
Acrolein/C_2H_3N	3.941	6.555	3.918	6.480
DMABN/CB[7]	4.181	4.480	4.146	4.430
GFP/Protein	2.552	3.328	2.528	3.324
DBO/CB[7]	3.475	6.158	3.444	6.124
Pyrazine/DMI-Br	4.452	5.223	4.381	5.176
Pyrazine/DMI-I	4.605	5.212	4.537	5.170

3.2 FDE-ADC(2)

Table 7: First and second excitation energies in eV computed with the FDE-ADC(2) approach. Excitation energies computed with the cc-pVTZ and aug-cc-pVTZ basis set are denoted "TZ" and "aTZ", respectively.

System	ΔE_{S1-TZ}	ΔE_{S2-TZ}	ΔE_{S1-aTZ}	ΔE_{S2-aTZ}
Formamid/H_2O	5.705	7.526	5.572	6.747
Berenil/DNA	2.711	3.046	2.677	3.018
PNP^-/Na^+(aq)	3.101	3.673	2.981	3.596
Cholesterol/cy-C_6H_{12}	7.089	7.208	5.472	5.802
Aniline-yellow/H_2O	2.988	4.253	2.963	4.157
Acrolein/C_2H_3N	3.902	6.615	3.868	5.810
DMABN/CB[7]	4.219	4.504	3.891	3.979
GFP/Protein	2.546	3.168	2.468	3.029
DBO/CB[7]	3.471	6.358	3.429	4.282
Pyrazine/DMI-Br	4.438	5.229	4.378	5.175
Pyrazine/DMI-I	4.554	5.218	4.497	5.166

3.3 FDE(ECP)-ADC(2)

Table 8: First and second excitation energies in eV computed with the FDE(ECP)-ADC(2) approach. Excitation energies computed with the cc-pVTZ and aug-cc-pVTZ basis set are denoted "TZ" and "aTZ", respectively.

System	ΔE_{S1-TZ}	ΔE_{S2-TZ}	ΔE_{S1-aTZ}	ΔE_{S2-aTZ}
Formamid/H_2O	5.663	7.586	5.585	7.377
Berenil/DNA	2.735	3.054	2.700	3.027
PNP^-/Na^+(aq)	3.175	3.536	3.156	3.590
Cholesterol/cy-C_6H_{12}	7.185	7.266	6.335	6.412
Aniline-yellow/H_2O	2.977	4.170	2.957	4.106
Acrolein/C_2H_3N	3.908	6.641	3.871	6.533
DMABN/CB[7]	4.270	4.519	4.215	4.461
GFP/Protein	2.501	2.716	2.486	2.774
DBO/CB[7]	3.472	6.151	3.437	6.106
Pyrazine/DMI-Br	4.319	5.250	4.294	5.193
Pyrazine/DMI-I	4.501	5.226	4.446	5.175

3.4 PE-ADC(2)

Table 9: First and second excitation energies in eV computed with the PE-ADC(2) approach. Excitation energies computed with the cc-pVTZ and aug-cc-pVTZ basis set are denoted "TZ" and "aTZ", respectively.

System	ΔE_{S1-TZ}	ΔE_{S2-TZ}	ΔE_{S1-aTZ}	ΔE_{S2-aTZ}
Formamid/H_2O	5.851	7.460	5.648	6.430
Berenil/DNA	2.703	3.046	2.657	2.949
PNP^-/Na^+(aq)	2.961	3.589	2.773	3.463
Cholesterol/cy-C_6H_{12}	7.092	7.215	3.378	4.531
Aniline-yellow/H_2O	2.994	4.318	2.967	4.217
Acrolein/C_2H_3N	3.925	6.607	3.895	5.789
DMABN/CB[7]	4.525	4.692	1.177	1.323
GFP/Protein	2.579	3.559	1.279	2.555
DBO/CB[7]	3.472	6.101	0.886	1.096
Pyrazine/DMI-Br	4.553	5.207	4.529	5.149
Pyrazine/DMI-I	4.759	5.180	4.847	5.104

3.5 PE(ECP)-ADC(2)

Table 10: First and second excitation energies in eV computed with the PE(ECP)-ADC(2) approach. Excitation energies computed with the cc-pVTZ and aug-cc-pVTZ basis set are denoted "TZ" and "aTZ", respectively.

System	ΔE_{S1-TZ}	ΔE_{S2-TZ}	ΔE_{S1-aTZ}	ΔE_{S2-aTZ}
Formamid/H_2O	5.822	7.508	5.754	7.322
Berenil/DNA	2.722	3.050	2.688	3.024
PNP^-/Na^+(aq)	3.002	3.293	2.988	3.349
Cholesterol/cy-C_6H_{12}	7.190	7.255	6.336	6.408
Aniline-yellow/H_2O	2.984	4.207	2.963	4.144
Acrolein/C_2H_3N	3.929	6.628	3.897	6.527
DMABN/CB[7]	4.576	4.710	4.526	4.607
GFP/Protein	2.558	2.959	2.526	3.044
DBO/CB[7]	3.475	6.105	3.440	5.802
Pyrazine/DMI-Br	4.408	5.233	4.385	5.178
Pyrazine/DMI-I	4.660	5.198	4.619	5.148

4 Validation of used methods

This section shows the excitation energies of different variants of the methods used for systems (h) and (i). System (i) was chosen because the second excitation energy had the highest error, while system (h) was selected because it exhibited the largest changes in the excitation energy based on the localization method.

4.1 Localization methods

Table 11: First excitation energy in eV of system (h) in the aug-cc-pvTZ basis computed with the subsystem-ADC(2) approach using different localization methods.

Localization-method	ΔE_{S1-aTZ}
Boys	2.525
second moment (exponent=2)	2.526
fourth moment	2.525
IBO	2.528

Table 12: First and second excitation energies in eV of system (i) in the aug-cc-pvTZ basis computed with the subsystem-ADC(2) approach using different localization methods.

Localization-method	ΔE_{S1-aTZ}	ΔE_{S2-aTZ}
Boys	3.447	6.124
second moment (exponent=2)	3.447	6.124
fourth moment	3.447	6.124
IBO	3.444	6.124

4.2 Kinetic Functionals

Table 13: First and second excitation energies in eV of system (h) in the aug-cc-pvTZ basis computed with the FDE-ADC(2) approach using different kinetic functionals.

Kinetical functional	ΔE_{S1-aTZ}	ΔE_{S2-aTZ}
lc94	2.469	3.029
revapbek	2.468	3.094
thomas-fermi	2.482	3.150
subsystem-ADC(2)	2.528	3.324

Table 14: First and second excitation energies in eV of system (i) in the aug-cc-pvTZ basis computed with the FDE-ADC(2) approach using different kinetic functionals.

Kinetical functional	ΔE_{S1-aTZ}	ΔE_{S2-aTZ}
lc94	3.429	4.282
revapbek	3.428	4.287
thomas-fermi	3.432	4.525
subsystem-ADC(2)	3.444	6.124

4.3 Long distance Correction

The long distance correction to the kinetic-energy component of the embedding potential proposed by Jacob et al.[2] is a computational cheap approximation and lead to accurate orbital energies for the Li-H2O system. In order to test its efficiency for the here used systems the long distance correction has been used for system (i) with the proposed value of 0.1 for the α parameter. Table 15 shows the first two excitation energies for different FDE-ADC(2) approaches and the reference method subsystem-ADC(2). An improvement of 0.5 eV could be achieved for the second excitation with the long distance correction compared to an regular FDE-ADC(2) approach, but even with this improvement an error of about 1.4 eV remains. The FDE(ECP) approach on the other hand could reduce the error to about 0.2 eV. This lead us to the conclusion that a simple reduction of the embedding potential in regions with a much higher density of the environment is not enough to compensate the missing repulsion between the subsystems.

Table 15: First and second excitation energies in eV of system (i) in the aug-cc-pVTZ basis computed with different approaches. The long distance correction approach is denoted as "ldc" and has been used with a value of 0.1 for the parameter α.

Method	ΔE_{S1-aTZ}	ΔE_{S2-aTZ}
FDE-ADC(2)	3.429	4.282
FDE-ldc-ADC(2)	3.426	4.727
FDE(ECP)-ADC(2)	3.437	6.106
Subsystem-ADC(2)	3.444	6.124

References

(1) Peterson, K. A.; Figgen, D.; Dolg, M.; Stoll, H. *J. Chem. Phys.* **2007**, *126*, 124101.

(2) Jacob, C. R.; Beyhan, S. M.; Visscher, L. *J. Chem. Phys.* **2007**, *126*, 234116.

A.2 Combining PNOs with the FDE method

The following supporting information belong to the submitted manuscript presented in chapter 4 and is submitted to the *Journal of Chemical Physics* by **AIP Publishing** on the 24th of May 2022.

TAPP

I. TIMINGS

1. *bay-C_2F_5-TAPP-H_4F_4*

Basis functions	Wall time [s]
1608	9766
3216	31130
4328	165341
8408	1209191

TABLE I: Computational wall times of supermolecular calculations.

Basis functions	Wall time [s]
1608	9766
4328	18531
8408	30715
15208	80624
19968	134941
31528	339676

TABLE II: Computational wall times of FDE calculations with a monomer as the main subsystem.

Basis functions	Wall time [s]
3216	31130
7296	102148
20896	236878
38576	611086

TABLE III: Computational wall times of FDE calculations with a dimer as the main subsystem.

TAPP

2. *all-C_3F_7-TAPP-Cl_4Br_4*

Basis functions	Wall time [s]
2044	21749
4088	68903
5668	326885
9292	1181782

TABLE IV: Computational wall times of supermolecular calculations.

Basis functions	Wall time [s]
2044	21749
5668	36541
9292	47300
16540	93688
34660	365131
26506	205507

TABLE V: Computational wall times of FDE calculations with a monomer as the main subsystem.

Basis functions	Wall time [s]
4088	68903
9524	175001
22208	292556
43952	750375

TABLE VI: Computational wall times of FDE calculations with a dimer as the main subsystem.

II. FIT INFORMATION

TABLE VII: Parameters resulting from fitting $f(N) = k \times N^a$ to the last two points of the wall times of compound (**2**).

Parameter	supermolecular	mono-FDE	dimer-FDE
a	2.99618	2.02116	1.54581
k	5.84939e-10	7.62394e-08	1.38113e-05

TABLE VIII: Parameters resulting from fitting $f(N) = k \times N^a$ to the last two points of the wall times of compound (**3**).

Parameter	supermolecular	mono-FDE	dimer-FDE
a	2.59989	2.14297	1.3798
k	1.58341e-08	1.89429e-08	8.17666e-05

III. EXCITATION ENERGIES

A. FDE-Kernel contributions

radius	10Å	12Å	15Å	20Å
Env.[a]	4	10	20	44
S_1	+0.00031	+0.00029	+0.00035	+0.00032
S_2	+0.00060	+0.00055	+0.00057	+0.00049
S_3	+0.00004	+0.00007	+0.00021	+0.00005
S_4	+0.00016	+0.00008	+0.00010	+0.00011
S_5	+0.00005	+0.00007	+0.00010	−0.00009
S_6	-0.00107	−0.00151	+0.00022	+0.00001
S_7	+0.00024	+0.00016	+0.00034	+0.00032
S_8	+0.00056	+0.00044	+0.00066	+0.00047
S_9	+0.00022	+0.00059	+0.00061	+0.00079
S_{10}	+0.00204	+0.00246	+0.00009	+0.00011
S_{11}	-0.00028	−0.00028	−0.00030	−0.00040
S_{12}	+0.00037	+0.00047	+0.00045	+0.00055
S_{13}	+0.00028	−0.00008	+0.00035	+0.00025
S_{14}	+0.00244	+0.00275	+0.00347	+0.00308
S_{15}	+0.00244	+0.00275	+0.00347	+0.00308
S_{16}	+0.00024	+0.00012	+0.00066	+0.00064
S_{17}	+0.00111	+0.00119	+0.00072	+0.00068
S_{18}	+0.00015	+0.00029	+0.00016	+0.00031

[a] Number of environment subsystems

TABLE IX: The kernel contributions for the first 18 excitations energies of compound (**2**) for different sizes of the environment using the PNO-CIS(D) method in eV.

TAPP

	ω		$\Delta\omega$
	w/o kernel	w/ kernel	
S_1	3.661	3.662	$0.66 \cdot 10^{-3}$
S_2	4.538	4.552	$13.65 \cdot 10^{-3}$

TABLE X: The first two excitation energies ω of the GFP system without (w/o) and with (w/) kernel contributions using the FDEu(ECP)-CCS method. All in eV.

	ω		$\Delta\omega$
	w/o kernel	w/ kernel	
S_1	3.682	3.683	$0.64 \cdot 10^{-3}$
S_2	6.711	6.711	$0.38 \cdot 10^{-3}$

TABLE XI: The first two excitation energies of the DBO/CB[7] system with and without kernel contributions using the FDEu(ECP)-CCS method. All in eV.

TAPP

B. Excitation energy convergence

1. Monomer target subsystem

radius	0Å	10Å	12Å	15Å	20Å
Env.[a]	0	4	10	20	44
S_1	3.114	3.127	3.125	3.141	3.141
S_2	3.250	3.268	3.269	3.255	3.255
S_3	3.697	3.698	3.701	3.670	3.671
S_4	3.702	3.743	3.736	3.700	3.700
S_5	3.749	3.744	3.748	3.826	3.826
S_6	3.898	3.864	3.870	3.869	3.869
S_7	3.898	3.921	3.923	3.895	3.895
S_8	4.138	4.157	4.158	4.142	4.142
S_9	4.307	4.420	4.410	4.489	4.488
S_{10}	4.638	4.657	4.658	4.637	4.638
S_{11}	4.743	4.759	4.760	4.783	4.782
S_{12}	4.780	4.784	4.784	4.832	4.829
S_{13}	4.897	4.910	4.910	4.871	4.873
S_{14}	4.916	4.910	4.911	4.898	4.898
S_{15}	4.916	4.951	4.949	4.970	4.970
S_{16}	4.991	5.013	5.012	5.007	5.007
S_{17}	5.195	5.224	5.222	5.223	5.223
S_{18}	5.699	5.689	5.692	5.682	5.682

[a] Number of environment subsystems

TABLE XII: The first 18 excitation energies of compound (**2**) for different sizes of the environment.

TAPP

radius	0Å	10Å	13Å	15Å	20Å
Env.[a]	0	4	8	16	36
S_1	2.815	2.820	2.821	2.788	2.789
S_2	2.930	2.911	2.911	2.909	2.909
S_3	3.115	3.153	3.157	3.096	3.103
S_4	3.192	3.221	3.225	3.161	3.165
S_5	3.395	3.439	3.440	3.388	3.390
S_6	3.452	3.444	3.441	3.496	3.492
S_7	3.469	3.493	3.492	3.519	3.516
S_8	3.519	3.558	3.561	3.552	3.557
S_9	4.000	4.036	4.038	3.946	3.949
S_{10}	4.021	4.066	4.063	4.076	4.074
S_{11}	4.290	4.332	4.331	4.131	4.150
S_{12}	4.310	4.359	4.365	4.290	4.296
S_{13}	4.379	4.536	4.535	4.313	4.313
S_{14}	4.504	4.542	4.556	4.445	4.480
S_{15}	4.575	4.640	4.641	4.479	4.501
S_{16}	4.640	4.682	4.684	4.536	4.542
S_{17}	4.654	4.759	4.770	4.623	4.562
S_{18}	4.668	5.005	5.010	4.633	4.625

TABLE XIII: The first 18 excitation energies of compound (**3**) for different sizes of the environment.

[a] Number of environment subsystems

TAPP

radius	0Å	8Å	10Å	12Å	15Å
Env.[a]	0	4	10	20	40
S_1	3.240	3.255	3.256	3.268	3.268
S_2	3.299	3.349	3.338	3.303	3.300
S_3	3.555	3.543	3.575	3.662	3.675
S_4	3.585	3.626	3.753	3.785	3.740
S_5	4.164	4.218	4.236	4.189	4.180
S_6	4.199	4.265	4.268	4.273	4.265
S_7	4.276	4.303	4.268	4.281	4.274
S_8	4.346	4.343	4.324	4.316	4.331
S_9	4.398	4.476	4.358	4.438	4.433
S_{10}	4.562	4.504	4.528	4.524	4.546
S_{11}	4.594	4.640	4.628	4.541	4.577
S_{12}	4.645	4.641	4.704	4.973	4.974
S_{13}	4.849	4.915	4.928	5.132	5.141
S_{14}	5.077	5.055	5.125	5.162	5.167
S_{15}	5.163	5.170	5.192	5.220	5.223
S_{16}	5.197	5.269	5.266	5.300	5.298
S_{17}	5.331	5.339	5.341	5.325	5.319
S_{18}	5.758	5.786	5.763	5.770	5.775

[a] Number of environment subsystems

TABLE XIV: The first 18 excitation energies of compound (**1**) for different sizes of the environment.

TAPP

2. Dimer target subsystem

radius	0Å	10Å	15Å	20Å
Env.[a]	0	6	26	52
S_1	2.919	2.932	2.929	2.928
S_2	3.054	2.995	2.987	2.986
S_3	3.115	3.127	3.138	3.140
S_4	3.137	3.150	3.160	3.161
S_5	3.181	3.196	3.186	3.187
S_6	3.265	3.285	3.275	3.275
S_7	3.599	3.617	3.610	3.608
S_8	3.602	3.678	3.664	3.662
S_9	3.673	3.702	3.678	3.680
S_{10}	3.720	3.716	3.689	3.689
S_{11}	3.803	3.902	3.858	3.861
S_{12}	3.995	4.031	4.026	4.018
S_{13}	4.110	4.068	4.140	4.128
S_{14}	4.236	4.424	4.348	4.360
S_{15}	4.282	4.433	4.443	4.453
S_{16}	4.309	4.498	4.503	4.493
S_{17}	4.363	4.503	4.523	4.523
S_{18}	4.634	4.566	4.596	4.605

[a] Number of environment subsystems

TABLE XV: The first 18 excitation energies of the dimer of compound (**2**) for different sizes of the environment in eV.

TAPP

radius	0Å	10Å	15Å	20Å
Env.[a]	0	6	20	44
S_1	2.243	2.247	2.236	2.238
S_2	2.262	2.261	2.250	2.252
S_3	2.800	2.809	2.784	2.786
S_4	2.848	2.855	2.828	2.830
S_5	2.951	2.939	2.938	2.938
S_6	2.957	2.948	2.945	2.946
S_7	3.111	3.121	3.089	3.097
S_8	3.113	3.125	3.095	3.103
S_9	3.221	3.228	3.178	3.182
S_{10}	3.227	3.229	3.180	3.184
S_{11}	3.463	3.491	3.423	3.428
S_{12}	3.789	3.639	3.660	3.531
S_{13}	3.956	4.022	3.952	3.956
S_{14}	4.001	4.029	3.964	3.967
S_{15}	4.021	4.070	3.971	4.085
S_{16}	4.023	4.091	4.093	4.093
S_{17}	4.192	4.319	4.330	4.331
S_{18}	4.616	4.542	4.521	4.528

[a] Number of environment subsystems

TABLE XVI: The first 18 excitation energies of the dimer of compound (**3**) for different sizes of the environment in eV.

IV. PNO TRUNCATION ERROR

	ω		$\Delta_{\text{trunc}}\omega$
method	CC2	PNO-CC2	
S_1	2.94623	2.96448	+0.01825
S_2	3.02761	3.05732	+0.02971
S_3	3.14241	3.19124	+0.04883
S_4	3.14926	3.20847	+0.05921

TABLE XVII: The first four excitation energies of the compound (**2**) computed using canonical MOs and PNOs and the resulting truncation error in eV.

	ω		$\Delta_{\text{trunc}}\omega$
method	FDE-CC2	FDE-PNO-CC2	
S_1	2.95475	2.97764	+0.02289
S_2	3.05472	3.08324	+0.02852
S_3	3.31202	3.34875	+0.03673
S_4	3.31906	3.35035	+0.03129

TABLE XVIII: The first four excitation energies of compound (**2**) within its molecular crystal (w20) computed using canonical MOs and PNOs and the resulting truncation error in eV.

TAPP

	$\Delta_{\text{env}}\omega$		Δ_{trunc} ($\Delta_{\text{env}}\omega$)
method	FDE-CC2	FDE-PNO-CC2	
S_1	0.00852	0.01316	+0.00464
S_2	0.02711	0.02592	−0.00119
S_3	0.16961	0.15751	−0.01210
S_4	0.16980	0.14188	−0.02792

TABLE XIX: Environment induced shifts of the excitation energy of compound (**2**) within its molecular crystal (w20) computed using canonical MOs and PNOs and the resulting truncation error on the shifts in eV.

	ω		$\Delta_{\text{trunc}}\omega$
method	CC2	PNO-CC2	
S_1	3.06251	3.05991	−0.00260
S_2	3.22050	3.23497	+0.01447
S_3	3.22896	3.23776	+0.00880
S_4	3.23139	3.24081	+0.00942

TABLE XX: The first four excitation energies of the compound (**1**) computed using canonical MOs and PNOs and the resulting truncation error in eV.

	ω		$\Delta_{\text{trunc}}\omega$
method	FDE-CC2	FDE-PNO-CC2	
S_1	3.07209	3.06968	−0.00241
S_2	3.25535	3.26175	+0.00640
S_3	3.37789	3.39907	+0.02118
S_4	3.39575	3.40783	+0.01208

TABLE XXI: The first four excitation energies of compound (**1**) within its molecular crystal (w15) computed using canonical MOs and PNOs and the resulting truncation error in eV.

TAPP

	$\Delta_{\text{env}}\omega$		Δ_{trunc} ($\Delta_{\text{env}}\omega$)
method	FDE-CC2	FDE-PNO-CC2	
S_1	0.00958	0.00977	+0.00019
S_2	0.03485	0.02678	−0.00807
S_3	0.14893	0.16131	+0.01238
S_4	0.16436	0.16702	+0.00266

TABLE XXII: Environment induced shifts of the excitation energy of compound (**1**) within its molecular crystal (w15) computed using canonical MOs and PNOs and the resulting truncation error on the shifts in eV.

Appendix B

FDE-Documentation

B.1 FDE related keywords

FDE calculations can be performed within TURBOMOLE program package with the `FDE` script. But before the script can be used two additional keywords beside the ones already necessary for a conventional calculation have to be defined.

- `$frag`
- `$fde`

`$frag` is needed to define the subsystems within your system. This can either be done via `define` and choosing the *frag* submenu within the coordination setup menu or by the user itself. For a water dimer the keyword could look like this:

```
$frag
   fragment 1  charge 0 atoms 1-3
   fragment 2  charge 0 atoms 4-6
```

Beside the atoms also the charge of a subsystem has to be defined. If no charge is given within the keyword a charge of 0 is anticipated.

The second keyword `$fde` defines all options within an FDE calculation. Most of its options can be defined with the program `define` but some of the options currently have to be added by the user itself. Such keywords will be

marked in the following with *.

`$fde`

`functional` *name*

specifies the exchange-correlation energy functional used for the embedding potential. It can also be used together with the xcfun and libxc library using the same notation as in `$dft`.

`gridsize` *integer*

specifies the size of the supermolecular integration grid used to compute embedding potential.

`kin` *name*

specifies the kinetic energy functional used for the embedding potential. The kinetic energy can also be included in the exchange-correlation functional via the xcfun and libxc library.

`comp`

switches to compatibility mode and FDE calculation is performed according to the implementation by Laricchia et. al.[133].

`fdeecp`

switches to FDE(ECP) calculation. Needs in addition the definition of the embedding ECPs in `$ecpemb`.

*`fdehf`

script ignores post-SCF calculation during the FaT iterations.

*`kernel` *integer*

forces the calculation of contributions beyond PTE for post-SCF methods. 1=on and 0=of.

`zemb_point_charges`

switch to point charges for the calculation of nuclear contributions in the embedding potential. This may lead to numerical errors and is therefore not recommended.

`active` *integer*

specifies currently active subsystem.

`mode` *integer*

specifies necessary steps in each program.

*`debug`

switches to debug output.

*`multilevel`

specifies methods used for specific subsystems in a multilevel ansatz. This option currently has to refer to another file.

*`ldc` *integer*

activates the long distance correction[42] and specifies the its factor alpha.

B.2 Multilevel ansatz within FDE

A multilevel ansatz is defined such that all keywords within the control file of the supermolecular system will also be used for each subsystem and specific keywords for specific subsystems have to be defined in the multilevel reference file. Such a multilevel reference file has to start with the keyword `$multilevel` and end with `$end`. Between those two keywords lines defining the fragments followed by their specific keywords have to be added. For a system containing of three subsystems where subsystem one and two shall be calculated with the DFT method and subsystem three with the CC2 model together with the first two excitation energies the multilevel reference file would look like this:

```
$multilevel
   fragment 1-2
      $dft
         gridsize 3
         functional pbe
   fragment 3
      $ricc2
```

```
        cc2
      $excitations
        irrep=a nexc=2 npre=4 nstart=6
$end
```

It is suggested to define all necessary basis sets already in the control file of the supermolecular system to ensure a correct execution of the calculation.

B.3 The FDE script

After all the above options have been set a FDE calculation can be started by running the FDE script. The script performs all FaT iterations until either the energy/dipole convergence criteria or the maximum number of FaT iterations is reached. Both of these values and more can be set by invoking the FDE script with additional arguments. In the following all important arguments for the in this work presented implementation of the FDE script will be listed and explained:

`-help / -help / -h / ?`
: Any one of these commands prints an explanation of all valid arguments for the FDE script.

`-k / -kin` *name*
: This argument defines the kinetic energy functional used for the embedding potential. and overwrites the in the control file defined kinetic energy functional. If this argument is not used the script tries to use the kinetic energy functional defined in the control file and if no kinetic functional is defined the default kinetic energy functional `revAPBEk` is used.

`-epsilon=` *number*
: This argument defines the energy convergence threshold in Hartree of the freeze and thaw iterations. If this argument is not invoked the default value of $5.0E-6$ is used.

`-max-iter=` *integer*

This argument defines the maximum number of freeze and thaw iterations. The default value is 20.

`-max-macro` *integer*

This argument defines the maximum number of macroiterations. The default value is 50.

`-frozencore`

This arguments activates the frozen core approximation for all subsystems and defines all orbitals below the freezing point as frozen.

`-fp` *number*

This argument defines the freezing point in Hartree. All orbitals with energies below the freezing point will be defined as frozen. The default value is 3.0.

`-comp`

This argument activates the compatibility mode and the FDE calculation will be performed according to the implementation by Laricchia et. al.[133].

Appendix C

Structures

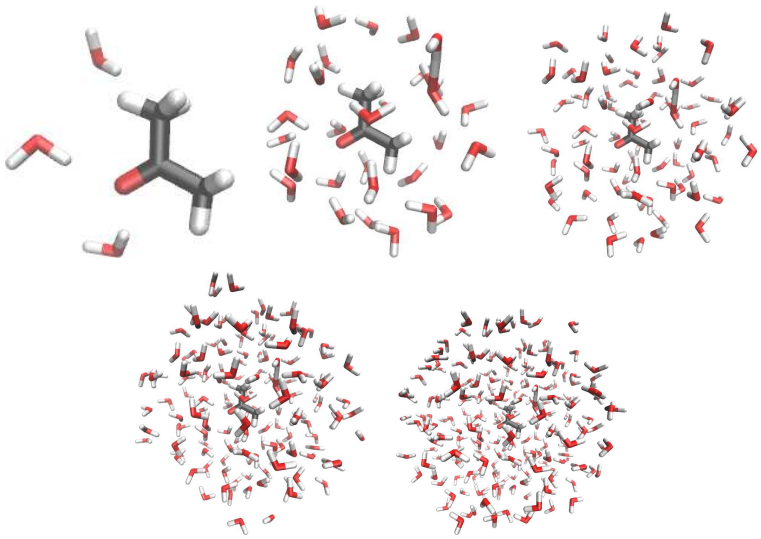


Figure C.1: Structures of the acetone in water clusters used in chapter 2.3 to determine the computational wall times of the initial FDE implementation.