Development and applications of non-empirical methods for modeling electronic structure in condensed phase

Kaminski, Jakub Wojciech

How to cite

KAMINSKI, Jakub Wojciech. Development and applications of non-empirical methods for modeling electronic structure in condensed phase. Doctoral Thesis, 2010. doi: 10.13097/archive-ouverte/unige:8507

This publication URL: https://archive-ouverte.unige.ch/unige:8507
Publication DOI: 10.13097/archive-ouverte/unige:8507

UNIVERSITÉ DE GENÈVE
Section de chimie et biochimie
Département de chimie physique

FACULTÉ DES SCIENCES
Docteur Tomasz A. Wesołowski

Development and applications of non-empirical methods for modeling electronic structure in condensed phase

THÈSE

présentée à la Faculté des sciences de l'Université de Genève
pour obtenir le grade de Docteur ès sciences, mention chimie

par

Jakub Wojciech Kamiński

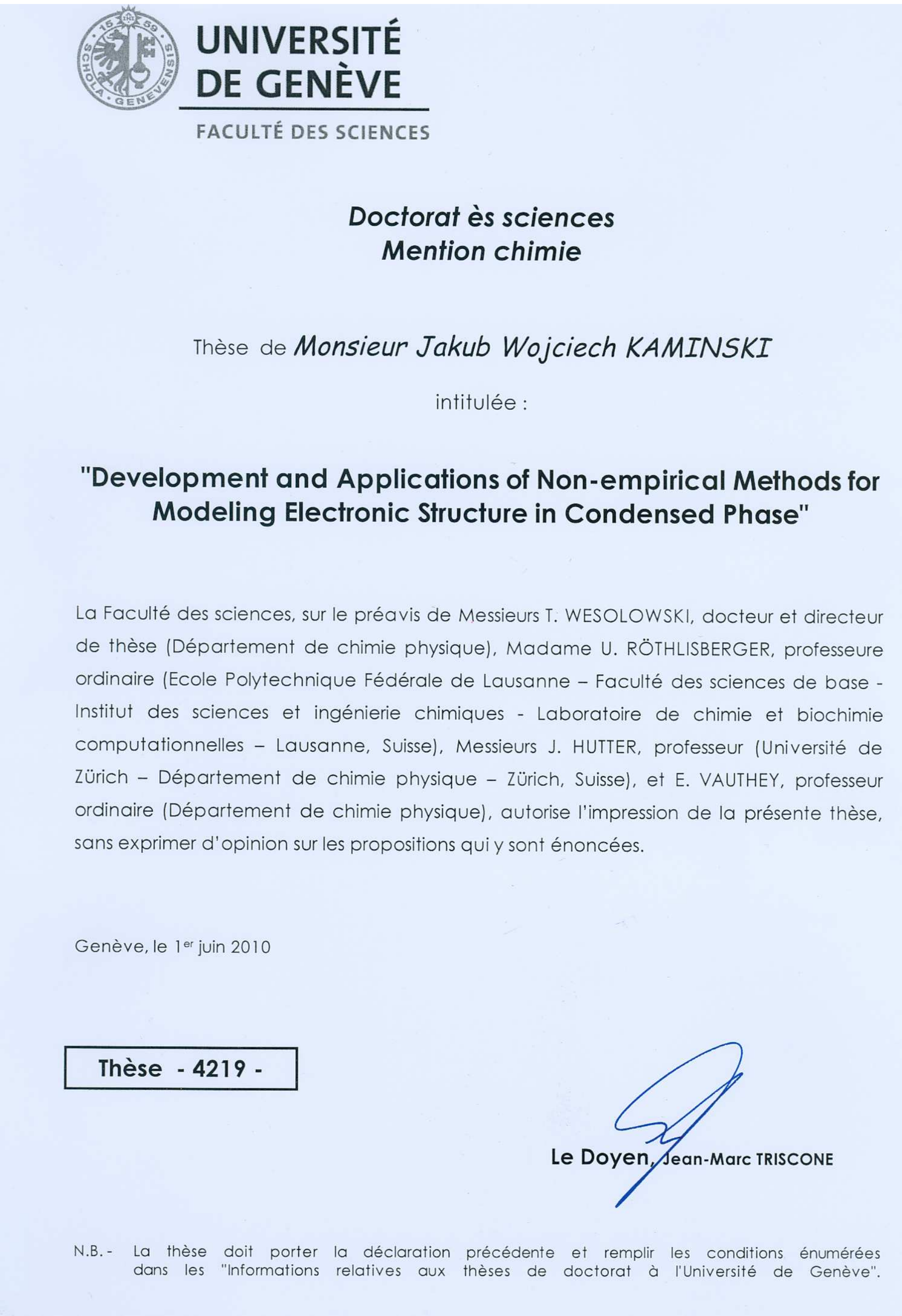
de

Pologne

Thèse N° 4219

GENÈVE

2010



UNIVERSITÉ DE GENÈVE

FACULTÉ DES SCIENCES

Doctorat ès sciences
Mention chimie

Thèse de *Monsieur Jakub Wojciech KAMINSKI*

intitulée :

"Development and Applications of Non-empirical Methods for Modeling Electronic Structure in Condensed Phase"

La Faculté des sciences, sur le préavis de Messieurs T. WESOLOWSKI, docteur et directeur de thèse (Département de chimie physique), Madame U. RÖTHLISBERGER, professeure ordinaire (Ecole Polytechnique Fédérale de Lausanne – Faculté des sciences de base - Institut des sciences et ingénierie chimiques - Laboratoire de chimie et biochimie computationnelles – Lausanne, Suisse), Messieurs J. HUTTER, professeur (Université de Zürich – Département de chimie physique – Zürich, Suisse), et E. VAUTHEY, professeur ordinaire (Département de chimie physique), autorise l'impression de la présente thèse, sans exprimer d'opinion sur les propositions qui y sont énoncées.

Genève, le 1er juin 2010

Thèse - 4219 -

Le Doyen, Jean-Marc TRISCONE

N.B. - La thèse doit porter la déclaration précédente et remplir les conditions énumérées dans les "Informations relatives aux thèses de doctorat à l'Université de Genève".

Search for your adventure!
And when you find it,
tackle it with all you have got,
all your strength, experience and passion.
Sir Edmund Hillary

Remerciements

First of all I would like to thank Doctor Tomasz Wesołowski for being a remarkable supervisor and a great person, always having time for me and patience to explain things I didn't understand. During four years of working on this this thesis I apprehend better not only science but also the world surrounding me.

I would like also to acknowledge Professor Ursula Rothlisberger, Professor Juerg Hutter and Professor Eric Vauthey for accepting invitation to be a jury member and for reading and commenting on this thesis.

Of course the work on this thesis would not be possible without support and sympathy of my friends. Everyday they filled me with positive energy and optimism during and after working hours, allowing me to enjoy every single day. I am very grateful to my big "Italian family" - Ivan for countless hours spent on biking or planning where to bike and being an amazing friend in-between, Big Giovanni for always being ready to help me with French language or with finishing a beer, Little Giovanni and Enza for always being cheerful and showing me the true passion and devotion for marvelous Italian cuisine, Francesco for being true "neapolitano" and bringing all the fun and knowledge at the same time to the group and to Laura for calming down Francesco's neapolitan character. A lot of thanks goes to Tanya for bringing light to the office everyday, Abdul for endless questions and showing me the other Indian world, Cesar for showing what does it mean to play bowling. This marvelous group of people would not be here together without Professor Laura Gagliardi. Special acknowledgements goes to Georgios for being the way he is and for his countless uncomfortable and critical questions that makes one to think twice. I could not forget also the former members of the Wesołowski and Gagliardi group - Marcin for introducing me to the software and tools I used during this Thesis, Juraj and his love to gadgets and endless discussions on peculiarities of Polish and Slovak languages, Daniel for always being ready to have fun and Eugeniusz for discussions about science in general.

Part of this Thesis have been done during cold winter days in Edmonton in Canada due to support and help from Doctor Andriy Kovalenko and Doctor Sergey Gusarov for which I am extremely grateful. My stay there would not be fun without people I have met - Luda for being my closest friend there and making the coldest days warm, Taras, Andrey, Stan, Viktor, Nikolay, Zoya, Oleksiy, Natali, John, Jiang, Qinggbin, Zsolt, for taking care of me.

I also would like to acknowledge help from Pierre-Yves who was always ready to assist me with any computer problem and bravely maintaining *wolf*, Renata and Sophie for keeping eye on my administrative paper work.

A lot of thanks goes to the friends outside of my group - Marcin, Joseph, Andrea, Piotr, Marina, Nadine, Uta, Vesna, Diego, Itana.

And last but not least my family - parents and my brother - for constant support and good word anytime I needed it.

Publications

- *"Solvatochromic Shifts Derived From Combination of Statistical-Mechanical Theory of Liquids with Orbital-Free Embedding Theory"*
  Jakub W. Kaminski, Sergey Gusarov, Tomasz A. Wesolowski, Andriy Kovalenko
  Journal of Physical Chemistry A, **114**, 6082, (2010)

- *"Cooperative Effect of Hydrogen-Bonded Chains in the Environment of a $\pi \rightarrow \pi^*$ Chromophore"*
  Georgios Fradelos, Jakub W. Kaminski, Tomasz A. Wesolowski, Samuel Leutwyler
  Journal of Physical Chemistry A, **113**, 9766, (2009)

- *"Linearized Orbital-Free Embedding Potential in Self-Consistent Calculations"*
  Marcin Dulak, Jakub W. Kaminski, Tomasz A. Wesolowski
  International Journal of Quantum Chemistry, **109**, 1886, (2009)

- *"Orbital-free effective embedding potential at nuclear cusp"*
  Juan Maria Garcia Lastra, Jakub W. Kaminski, and Tomasz A. Wesolowski
  Journal of Chemical Physics, **129**, 074107, (2008)

- *"$LiSc(BH_4)_4$, a new complex salt with discrete $Sc(BH_4)_4^-$ ions"*
  H. Hagemann, M. Longhini, J. W. Kaminski, T. A. Wesolowski, R. Cerny, N.Penin, M. H. Sorby, B.C. Hauback, G. Severa, C. M. Jensen
  Journal of Physical Chemistry A, **112**, 7551, (2008)

- *"The energy-differences based exact criterion for testing approximations to the functional for the kinetic energy of non-interacting electrons"*
  Yves A. Bernard, Marcin Dulak, Jakub W. Kaminski, and Tomasz A.

Wesolowski
Journal of Physics A: Mathematical and General, **41**, 055302 (2008)

- *"Equilibrium geometries of non-covalently bound intermolecular complexes derived from subsystem formulation of density functional theory"*
  Marcin Dulak, Jakub W. Kaminski, and Tomasz A. Wesolowski
  Journal of Chemical Theory and Computation, **3**, 735-745 (2007)

Résumé en français

Cette thèse va traiter de la théorie orbital-free embedding, qui consiste à diviser le système étudié en deux sous-systèmes - l'environnement, décrit par la densité électronique ρ_B et le système qui est englobé par ce dernier, représenté par ρ_A. Différents éléments utilisés dans la formulation de Hohenberg-Kohn-Sham de la théorie de la fonctionnelle de la densité, tels que les théorèmes de Hohenberg-Kohn [1], l'utilisation d'un système de référence basé sur des électrons non-interagissant [2] et en particulier de la fonctionnelle de la densité correspondant à l'énergie cinétique ($T_s[\rho]$) [3] sont utilisés pour définir le potentiel qui entoure le système qui est à l'étude: le potentiel environnant. Ce potentiel est *exact* dans la limite de l'utilisation de fonctionnelles exactes et est *orbitales-indépendant*, c'est-à-dire qu'il ne nécessite que de la densité électronique pour sa description [4]. La densité électronique ρ_A^{min}, correspondant à l'état fondamental pour un système d'électrons non-interagissant, est choisie de façon à minimiser l'énergie du système entier après addition d'une densité arbitraire (ρ_B) associée à l'environnement. Elle est obtenue à travers la résolution des équations à un électron (Eq. 20-21 dans la Ref. [4]) et l'équation 1 de ce manuscrit:

$$\left[-\frac{1}{2}\nabla^2 + v_{eff}^{KS}[\rho_A;\vec{r}] + v_{emb}^{KSCED}[\rho_A,\rho_B;\vec{r}]\right]\phi_i^A = \epsilon_i^A\phi_i^A \qquad i = 1, N^A \tag{1}$$

$\rho_A = 2\sum_i^{N_A}|\phi_i|^2$ et $v_{eff}^{KS}[\rho_A;\vec{r}]$ est le potentiel Kohn-Sham effectif. Le terme $v_{emb}^{KSCED}[\rho_A,\rho_B;\vec{r}]$ correspond au potentiel environnant local, qui dépend uniquement des densités électroniques ρ_A et ρ_B. Il est décrit de la manière suivante:

$$v_{eff}^{KSCED}[\rho_A,\rho_B;\vec{r}] = v_{eff}^{KS}[\rho_A+\rho_B;\vec{r}] + v_t[\rho_A,\rho_B](\vec{r}) \tag{2}$$

La composante $v_t[\rho_A, \rho_B](\vec{r})$ du potentiel environnant a pour origine la non-additivité de l'énergie cinétique et est associée à la répulsion de Pauli. La forme analytique de ce terme n'est pas connue et par conséquent cette quantité doit être approximée. Le développement d'une nouvelle approximation de $v_t[\rho_A, \rho_B](\vec{r})$ est le sujet du Chapitre 4 de cette thèse. L'analyse du comportement asymptotique de $v_t[\rho_A, \rho_B](\vec{r})$ pour des séparations intermoléculaires importante aboutit à une nouvelle condition exacte à laquelle ce potentiel doit obéir. Une nouvelle approximation tenant compte de ce paramètre est proposée. Des tests numériques sur des groupes de molécules mettant en jeu des interactions très faibles ont mis en évidence l'amélioration des résultats après introduction de cette nouvelle approximation, qui permet d'atteindre jusqu'à 50% des valeurs correspondant aux systèmes de référence. De plus, ce nouveau potentiel améliore la description des orbitales LUMO, ce qui permet de résoudre partiellement le problème lié au transfert artificiel de la densité électronique de la molécule à son environnement qui apparaît lors de l'étude de complexes chargés. L'implémentation de gradients analytiques de l'énergie au sein du formalisme "orbital-free embedding" permet d'utiliser cette méthode afin d'obtenir des géométries à l'équilibre pour des systèmes contenant différents complexes.

Dans le chapitre 5, un nouvel algorithme est proposé, pour lequel un minimum de l'énergie Born-Oppenheimer est obtenu uniquement à la fin des cycles d'optimisation. Cette approche est particulièrement utile dans le cadre de l'étude de larges systèmes. De plus, plusieurs variantes permettant l'optimisation de certains degrés de liberté ont été introduites afin de faciliter l'étude de grandes molécules. Différents complexes caractérisés par des types particuliers d'interactions ont été utilisés comme référence, ceci afin de déterminer la précision des différentes fonctionnelles d'énergie cinétique qui sont habituellement utilisées pour des calculs de type "orbital-free embedding". Mis à part les systèmes impliquant des interactions de type π-π, les autres composés à l'étude utilisant la fonctionnelle LDA ont permis d'obtenir d'excellents résultats. Les différences entre les distances intermoléculaires ainsi obtenues et celles correspondant à des méthodes ab-initio standards ne sont pas supérieures à 0.1 Å.

Afin de faciliter l'utilisation du potentiel orbital-free embedding à grande échelle, des approximations de ce dernier sont introduites dans le chapitre 4. Elles se basent sur la linéarisation du potentiel environnant autour d'une densité ρ_A^0 choisie arbitrairement. De cette manière, le potentiel environnant sera constant durant les itérations SCF et cela se traduira par une diminution du temps de calcul. La précision des résultats dépend fortement de la densité ρ_A^0 choisie. Ainsi, si cette dernière est proche de la densité ρ_A^0 réel de l'état fondamental, la qualité des résultats obtenus sera meilleure. Des analyses détaillées montrent que dans la plupart des cas, mis à part des complexes caractérisés par une importante déformation de la densité due à de fortes interactions, le choix optimal pour ρ_A^0 est la densité pour le sous-système isolé A. Les résultats obtenus en suivant ce protocole sont pratiquement identiques de ceux correspondant à des calculs se basant sur le potentiel environnant complet. Afin de contrôler les erreurs résultant de cette linéarisation, l'algorithme *splitSCF* double est utilisé. Dans ce cas, le processus SCF est divisé en deux boucles - interne, pour laquelle le potentiel orbital-free embedding linéaire est utilisé, et externe, pour laquelle ce dernier est mis à jour. Des tests numériques ont montré que seulement trois itérations pour la boucle externe sont nécessaires afin d'obtenir des résultats équivalents à ceux obtenus à travers des calculs conventionnels (quatre itérations pour des systèmes caractérisés par une déformation induite de la densité). Une autre propriété du potentiel orbital-free embedding linéaire est que ce dernier peut être facilement intégré à des codes déjà existants.

Une application pratique du potentiel orbital-free embedding sera le sujet du chapitre 7, où il sera également question de la méthode 3D-RISM-KH. Cette dernière approche fournit une description se basant sur la mécanique statistique des liquides. Cette méthode est un intermédiaire entre les modèles décrivant le solvant de manière implicite et explicite - elle ne se base pas sur une description atomique mais tient en compte la spécificité chimique des interactions entre les molécules. La méthode 3D-RISM-KH permet de diviser la molécule en différents sites et chacun d'entre eux se verra attribuer une densité électronique, qui sera utilisée dans l'équation 1 pour la description ρ_B de l'environnement. De cette façon, le potentiel environnant *moyenné* est obtenu. Ce dernier présente différents avantages lorsque le déplacement des pics du spectre d'absorption induit par le solvant est étudié. Un point important est que cette méthode ne nécessite de faire qu'un seul calcul sur une structure tenant compte des effets moyennés du solvant, au lieu de faire différents calculs coûteux sur un ensemble de configurations de l'environnement, qui sont le résultat de simulation de dynamique moléculaire. Les résultats présentés dans le chapitre 7 valide l'approche utilisée. Par exemple, l'utilisation du potentiel environnant *moyenné* dans le cas de l'aminocumarin C151 et de l'acétone, qui ont été précédemment étudiés à travers la méthode orbital-free embedding, permet d'obtenir des

résultats en accords avec les valeurs de référence et les données expérimentales. La différence est que la méthode utilisée permet de réduire considérablement le coût des calculs.

Le dernier chapitre de cette thèse concerne l'application du formalisme de la fonctionnelle de la densité Kohn-Sham afin de confirmer les études expérimentales de la structure et du spectre vibrationnelle de $\mathrm{LiSc(BH_4)_4}$. Pour une fonctionnelle d'échange-corrélation donnée, la structure de ce composé et le spectre IR ont été déterminés. Les résultats obtenus sont en parfait accord avec les données expérimentales et confirment la structure de $\mathrm{LiSc(BH_4)_4}$ obtenue à travers une expérience de diffraction sur poudre. Ces résultats attestent la validité du protocole computationnel utilisé. Ce travail tire parti des avantages de l'utilisation du potentiel orbital-free embedding pour l'étude des changements de la structure électronique des molécules dus à leurs interactions avec l'environnement. La nouvelle fonctionnelle pour le potentiel d'énergie cinétique, pour un système de particules non-interagissant obéissant à la relation exacte dérivée, a été utilisée avec à la clef une amélioration des résultats.

En résumé, de nouveaux algorithmes et des approximations facilitant l'application de ce potentiel dans la cadre d'études à larges échelles d'interactions dans la matière condensée sont présentés. La combinaison de l'orbital-free embedding avec la théorie de la mécanique statistique des liquides est introduite et est validée dans le cadre de l'étude des déplacements des pics d'absorption dus à différents solvants et pour différents chromophores.

Contents

Introduction

The central object of quantum mechanics is the Schrödinger equation. It describes the quantum state of the system and its evolution in time. The analytical solution to this equation, called wavefunction, is, however, available only for certain, well defined systems. Otherwise, only the approximate solutions are possible. The development of the approximate methods to solve the Schrödinger equation is main domain of quantum chemistry. These methods can be applied to predict the electronic structure of the molecules to complement various experimental studies or even predict the existence of new unknown compounds [5].

Quantum-chemical methods can be devised in several categories. The most accurate ones are often referred as *ab-initio* methods, as they aim to solve the Schrödinger equation for the optimal wavefunction without use of any empirical parameters. The price for the accuracy is paid, however, in high demand for the computational resources required to perform calculations. For this reasons, the application of *ab-initio* methods is limited to relatively small molecules. The cost of computations can be reduced by parametrizing the Schrödinger equation (semi-empirical methods) or by treating the chemical interactions between atoms and molecules in classical way, without constructing the wavefunction at all (for instance molecular mechanics).

Alternatively to the wavefunction, the properties of the system can be described by its electron density. This approach is justified on the basis of Hohenberg and Kohn [1] theorems, which provide the foundation of Density Functional Theory (DFT). The success of DFT based approaches can be attributed to Kohn and Sham [2], who transferred the original theory into the practical computational method. Nowadays, Kohn-Sham DFT is one of the most commonly used quantum-chemical tools to predict the electronic structure for many-electron molecules. Although it relies on approximation to relevant density functionals, which commonly involve empirical parameters, its accuracy compares favorably with the *ab-initio* methods. The electron density, however, is not as complicated object as wavefunction and the potentials involved in DFT calculations are local quantities. For this reasons, DFT can be successfully applied to compute the properties of the molecules of considerable size.

Influence of the environment on the physicochemical phenomena occurring in condensed matter is of the great importance in chemistry and biochemistry. The environment affects the reactivity of the enzymes, position and intensity of bands in various types of spectroscopies or the reaction rates in chemical reactions. Various quantum-chemical methods have been developed to describe solvent effects, commonly introducing embedding potential, which is acting on solvated molecule. Wesołowski and Warshel showed [4] that such potential is uniquely defined by the pair of two electron densities - that of investigated embedded system and that of its environment. Hence, this potential is referred as *orbital-free embedding potential*. The first of the two densities is expressed by means of quantum-chemical descriptors (Kohn-Sham orbitals, multideterminantal wavefunction, one-electron density matrix). The density of the environment can be taken from any level of description (quantum mechanics, molecular mechanics, experiment). For this reasons *orbital-free embedding potential* is ideally suited for multi-level modeling of interactions in condensed matter.

The *orbital-free embedding potential* is expressed by means of density functionals and it is an *exact* quantity in the exact limit of these functionals. The analytical form of some of these terms is, however, not know and it has to be approximated. The main motivation of this work originates from the interest in the properties of the *orbital-free embedding potential* and its possible applications to model chemical interactions. This involves extending its range of applicability by developing new approximations to the relevant density functionals and design of new algorithms and approaches for use of *orbital-free embedding potential* in large-scale multi-level modeling of condensed matter.

This thesis are organized as follows: Part I provides introduction to the used theoretical methods. In Chapter, 1 the Hohenberg-Kohn density functional theory and its extension proposed by Kohn and Sham are outlined. These fundamental concepts are used as a basis to describe the orbital-free embedding formalism, which is overviewed in Chapter 2. Chapter 3 introduces the time-dependent density functional theory. In Parts II to V the own research conducted during the Ph.D. thesis is described. The development of the new approximant to the non-additive kinetic energy potential - an important component of the *orbital-free embedding potential* in addressed in Part II. Chapters grouped together in Part III are related to the development of the new algorithms that facilitates application of embedding potential in multi-scale modeling of interactions in condensed matter. In Chapter 5, the linearization approximation to the embedding potential is introduced. To verify its adequacy, a special cycle of calculations, dubbed as *splitSCF*, is discussed. Chapter 6 presents the application of embedding formalism to find the equilibrium structures of noncovalently bound intermolecular complexes. In Part IV, the combination of orbital-free embedding formalism with method based on statistical mechanical theory of the liquids is introduced. This multi-level model is used to simulate solvatochromic shifts of selected chromophores in various solvents. This section also closes the main body of the research of this thesis. Part V concerns the auxiliary application of DFT based theoretical methods to provide explanation and conformation of experimental studies on borohydrates and provides a preparation stage for furtherer study on this complexes applying orbital-free embedding formalism. Finally, this work is summarized and final conclusions are drawn (Part VI).

Part I

Theory

Chapter 1

Density Functional Theory

Nowadays density functional theory (DFT) provides well established framework for the methods used to determine electronic structure of atoms and molecules which are recognized for the efficiency and reasonable accuracy. In foundations of DFT lay two theorems, formulated by Hohenberg and Kohn [1], proving that the electron density $\rho(\vec{r})$ defined as:

$$\rho(\vec{r_1}) = N \int ... \int |\Psi(\vec{r}_1, \vec{r}_2, ..., \vec{r}_N)|^2 d\vec{r}_2 ... d\vec{r}_N \tag{1.1}$$

can be used as an alternative to wavefunction $\Psi(\vec{r}_1, \vec{r}_2, ..., \vec{r}_N)$ to obtain ground state properties of many-electron systems. In eq. 1.1 $\vec{r}$ deontes spatial coordinates and N stands for number of electrons; closhed shell molecules will be considered in this chapter, therefore spin coordinate is ommited for simplicity. Density functional theory approach brings many practical advantages. Wavefunction is a complicated object depending on 3N spatial coordinates. The electron density is much simpler as it is function of only three spatial coordinates, therefore, many-body problem of obtaining ground-state energy of a given system using only its density is easier. In this chapter the foundations of density functional theory will be presented. The practical, widely used refolmulation of DFT by Kohn and Sham [2] will be also reviewed.

1.1 Hohenberg-Kohn theorems

The formal basis for the density functional theory was given in 1964 by Hohenberg and Kohn in the form of the two theorems.

First theorem says: *The density* $\rho_0(\vec{r})$ *corresponding to a non-degenerate ground state specifies the external potential* $v_{ext}(\vec{r})$ *up to a constant.*

This theorem states important fact: if we know the ground state density $\rho_0(\vec{r})$ we can *uniquely* define (up to an additive constant) the external potential $v_{ext}(\vec{r})$ corresponding to this density, and vice-versa.

Hohenberg-Kohn theorem has also furtherer going consequences. To show them, we will start from the number of electrons of the system, which is readily available from the electron density according to the equation:

$$N = \int \rho(\vec{r}) d\vec{r} \tag{1.2}$$

Additionally, the positions of the nuclei can be discovered by analyzing the cusps of the density, while the charge of the given nucleus can be learned from Kato cusp condition [6]. The availability of this information allows to determine the external potential $v_{ext}(\vec{r})$ corresponding the electron density of the given system. Note, that these observations also provide the proof for the correctness of first Hohenberg-Kohn theorem. Knowledge of the $v_{ext}(\vec{r})$ allows to write down the total Hamiltonian for the N-electron system and ground-state wavefunction can be calculated from Schrödinger equation. The ground state expectation value of any observable represented by any operator $\hat{O}$ can be, therefore, written as density functional:

$$O[\rho] = \langle \Psi_0[\rho] | \hat{O} | \Psi_0[\rho] \rangle \tag{1.3}$$

The second Hohenberg-Kohn theorem provides variational principle for the ground state density and states that *the total energy functional* $E[\rho]$ *reaches minimum at the non-degenerate ground state density* ρ_0.

The second Hohenberg-Kohn theorem can be showed by writing the expression for the total energy. Starting from eq. 1.3 it can be defined as:

$$E[\rho] = \langle \Psi_0[\rho] | \hat{T} + \hat{V} + \hat{V}_{ext} | \Psi_0[\rho] \rangle = F_{HK}[\rho] + \int v_{ext}(\vec{r}) \rho(\vec{r}) d\vec{r} \tag{1.4}$$

where $\hat{T}$, $\hat{V}$ and $\hat{V}_{ext}$ are kinetic energy, two particle interaction and external potential operators respectively. $F_{HK}[\rho]$ is universal Hohenberg-Kohn functional defined as (using eq. 1.3):

$$F_{HK}[\rho] = \langle \Psi_0[\rho] | \hat{T} + \hat{V} | \Psi_0[\rho] \rangle = \langle \Psi_0[\rho] | \hat{T} | \Psi_0[\rho] \rangle + \langle \Psi_0[\rho] | \hat{V} | \Psi_0[\rho] \rangle = T[\rho] + V[\rho] \tag{1.5}$$

The variational principle from the second Hohenberg-Kohn theorem is written as a minimum of the total energy functional (eq. 1.4) with respect to density:

$$E_0 = \min_{\rho} E[\rho] \tag{1.6}$$

It also means, that for any trial density, $\tilde{\rho}(\vec{r})$, following inequality can be written:

$$E[\tilde{\rho}(\vec{r})] \geq E[\rho_0(\vec{r})] = E_0 \tag{1.7}$$

where $\rho_0(\vec{r})$ and E_0 are ground-state density ground state and energy of multi-electron system respectively.

In order for the variational principle in eqs. 1.6 and 1.7 to be valid, the total energy functional has to be defined for any density ρ. The Hohenberg-Kohn functional $F_{HK}[\rho]$ is, however, defined only for a small set of v - representable densities, i.e. the densities which are associated with the antisymmetric ground-state wavefunction with some external potential $v_{ext}(\vec{r})$. The conditions for the density to be v-representable are unknown [7], therefore, in practice it is difficult to assure that any trial density $\tilde{\rho}(\vec{r})$ falls into this category. For this reasons it is desirable to extend the applicability of $F_{HK}[\rho]$ to a larger domain of densities. This task can be accomplished with constrained-search procedure introduced by Levy [3] (commonly referred as Levy constrained-search formulation).

According to definition of the density given in eq. 1.1, the ground-state density $\rho_0(\vec{r})$ can be obtained from ground state wavefunction Ψ_0. The same density $\rho_0(\vec{r})$ might be, however, obtained from other wavefunction,Ψ', which is not necessarily ground-state wavefunction. To distinguish between the two Levy used minimum energy principle:

$$\langle \Psi' | \hat{H} | \Psi' \rangle \geq \langle \Psi_0 | \hat{H} | \Psi_0 \rangle = E_0 \tag{1.8}$$

where $\hat{H}$ is Hamiltonian of N-electron system. The above equation can be rewritten (see eq. 1.4):

$$\langle \Psi' | \hat{T} + \hat{V} | \Psi' \rangle + \int v_{ext}(\vec{r})\rho(\vec{r})d\vec{r} \geq \langle \Psi_0 | \hat{T} + \hat{V} | \Psi_0 \rangle + \int v_{ext}(\vec{r})\rho(\vec{r})d\vec{r} \quad (1.9)$$

which gives:

$$\langle \Psi' | \hat{T} + \hat{V} | \Psi' \rangle \geq \langle \Psi_0 | \hat{T} + \hat{V} | \Psi_0 \rangle \quad (1.10)$$

Using above expression, similar as in eq. 1.5, a new functional $F_L[\rho]$ can be introduced and its minimum reads:

$$F_L[\rho] = \min_{\Psi \to \rho} \langle \Psi | \hat{T} + \hat{V} | \Psi \rangle \quad (1.11)$$

where search for the minimum is performed over all the wavefunctions giving the target density $\rho(\vec{r})$. Obviously, functional $F_L[\rho]$ will be equal to $F_{HK}[\rho]$ if the electron density $\rho(\vec{r})$ will be v-representable. It is worth to note, however, that density ρ in eq. 1.11 does not necessarily has to be v-representable. It is sufficient that it belongs to the wider class of the N-representable densities, i.e. the density that can be obtained from some antisymmetric wavefunction and obeys following relations [8]:

$$\rho(\vec{r}) \geq 0 \qquad \int \rho(\vec{r}) = N \qquad \int |\nabla \rho(\vec{r})^{\frac{1}{2}}|^2 < \infty \quad (1.12)$$

Using N-representable densities to find the minimum of the functional $F_L[\rho]$ eliminates the v-representability constraint put on the variational principle stated in second Hohenberg-Kohn theorem. The minimization of the total-ground state energy (eq. 1.4) can be now written as follows:

$$\begin{aligned} E_0 &= \min_{\Psi} \langle \Psi | \hat{T} + \hat{V} + \hat{V}_{ext} | \Psi \rangle \qquad (1.13) \\ &= \min_{\rho \to N} \left\{ \min_{\Psi \to \rho} \left[\langle \Psi | \hat{T} + \hat{V} | \Psi \rangle + \int v_{ext}(\vec{r})\rho(\vec{r})d\vec{r} \right] \right\} \\ &= \min_{\rho \to N} \left\{ F_L[\rho] + \int v_{ext}(\vec{r})\rho(\vec{r})d\vec{r} \right\} = \min_{\rho \to N} E[\rho] \end{aligned}$$

1.2 Kohn-Sham formulation of DFT

Although density functional theory formulated in Hohenberg and Kohn theorems is elegant, convenient and exact orbital-free approach to calculate properties of many-electron systems, in practice it is not widely used. The main reason for this lays in the unknown analytic form of the $F_{HK}[\rho]$ functional, especially its kinetic energy part, which provides the biggest contribution to the total energy in eq. 1.5. The first attempts to approximate $T[\rho]$ were made in pioneering works by Thomas and Fermi [9, 10] or von Weizsacker [11], however, up to date no universally applicable analytic form has been proposed. To overcome this difficulties Kohn and Sham [2] proposed alternative approach - to replace the system of N-interacting electrons by fictitious reference system of N-non-interacting electrons described by single Slater determinant Φ_{KS}. The Kohn-Sham method relies, therefore, on important *assumption*, that for every ground state density of interacting system there is non-interacting system which has the same density in its ground state. This assumption cannot be easily verified.

To describe non-interacting system, Kohn and Sham introduced orbitals. This allows to calculate kinetic energy of non-interacting electrons in an exact way, given the formula:

$$T_s[\rho] = \sum_{i=1}^{N} \langle \phi_i | -\frac{1}{2}\nabla^2 | \phi_i \rangle \tag{1.14}$$

In Kohn-Sham approach kinetic energy is no longer *explicit* functional of the electron density, but now it has *implicit* form dependent on orbitals. A considerable effort has been made to express $T_s[\rho]$ as an *explicit* density functional [12] (transforming KS-DFT in an orbital-free method), nevertheless the most promising approximations did not lead to satisfactory results usually giving errors in calculations in the range of 1% [12, 13].

The electron density in Kohn-Sham formalism is obtained as:

$$\rho(\vec{r}) = \sum_{i=1}^{N} |\phi_i(\vec{r})|^2 \tag{1.15}$$

On the basis of the second Hohenberg-Kohn theorem, the ground state electron density, $\rho_0(\vec{r})$, can be obtained by minimizing the total energy functional. In the case of Kohn-Sham approach the expression for the total energy reads:

$$E_{KS}[\rho] = T_s[\rho] + J[\rho] + E_{xc}[\rho] + \int v_{ext}(\vec{r})\rho(\vec{r})d\vec{r} \tag{1.16}$$

Term $J[\rho]$ represents the classical Coulomb energy:

$$J[\rho] = \frac{1}{2}\int\int \frac{\rho(\vec{r})\rho(\vec{r}')}{|\vec{r}-\vec{r}'|}d\vec{r}d\vec{r}' \tag{1.17}$$

whereas $E_{xc}[\rho]$ is called *exchange-correlation energy* and is defined as:

$$E_{xc} = T[\rho] - T_s[\rho] + V[\rho] - J[\rho] \tag{1.18}$$

All the terms in equation 1.16 are available in their exact analytic forms, only density functional of $E_{xc}[\rho]$ is not know, therefore, it needs to be approximated. Although it may seem that Kohn and Sham changed problem of approximating $F_{HK}[\rho]$ from original Hohenberg-Kohn theory into the problem of approximating $E_{xc}[\rho]$, however, compared to $F_{HK}[\rho]$ the values of $E_{xc}[\rho]$ are much smaller, therefore the errors arising from approximants to this term are also expected to be smaller.

The energy functional in eq. 1.16 can be rewritten in the terms of the orthonormal one-electron functions, $\{\phi_i\}$, commonly labelled as Kohn-Sham orbitals:

$$\begin{aligned} E_{KS}[\rho] &= T_s[\rho] + J[\rho] + E_{xc}[\rho] + \int v_{ext}(\vec{r})\rho(\vec{r})d\vec{r} \qquad (1.19)\\ &= \sum_{i=1}^{N}\langle\phi_i| - \frac{1}{2}\nabla^2|\phi_i\rangle + J[\rho] + E_{xc}[\rho] + \int v_{ext}(\vec{r})\rho(\vec{r})d\vec{r} \end{aligned}$$

Euler-Lagrange minimization of eq. 1.19 under the constrain for the orthonormality of the Kohn-Sham orbitals, $\int \phi_i^*(\vec{r})\phi_j(\vec{r}) = \delta_{ij}$, gives one-electron Kohn-Sham equations:

$$\left(-\frac{1}{2}\nabla^2 + v_{eff}^{KS}[\rho](\vec{r})\right)\phi_i(\vec{r}) = \varepsilon_i\phi_i(\vec{r}) \tag{1.20}$$

where

$$v_{eff}^{KS}[\rho](\vec{r}) = \sum_{i=1}^{N_{nuc}} -\frac{Z_i}{|\vec{r}-R_i|} + \int \frac{\rho(\vec{r})}{|\vec{r}-\vec{r}'|}d\vec{r} + \frac{\delta E_{xc}[\rho]}{\delta\rho} \tag{1.21}$$

is the Kohn-Sham effective potential. Note, that derivation of eqs. 1.20 and 1.21 is subject to additional *assumption*, that density functionals are differentiable (i.e. value of the functionals changes smoothly with the changes of the density).

The orbitals $\phi_i(\vec{r})$ from eq. 1.20 are used to obtain density according to eq. 1.15. It belongs to class of *non-interacting pure-state v-representable* electron densities, which means that it corresponds to the ground state in a system of non-interaction electrons in some external potential.

From practical point of view, introduction of the orbitals in Kohn-Sham formulation of density functional theory helps to overcome the difficulties in approximating $F_{HK}[\rho]$ found in original Hohenberg-Kohn theory for the price of making calculations computationally more demanding. Comparing to the wavefunction based *ab-initio* methods, however, it sill is an interesting alternative. The cost of solving of Kohn-Sham equations is smaller than cost of solving Hartree-Fock equations as Kohn-Sham effective potential is local potential and same for all the electrons. Moreover, unlike Hartree-Fock theory which is by definition an approximate method, Kohn-Sham DFT becomes exact theory in the limit of the exact functionals.

Turning to the exchange-correlation energy, development of the approximation to this quantity is subject of many studies [14]. Various strategies following different principles to construct approximate $E_{xc}[\rho]$ are proposed. Commonly, following Perdew [15], they are ordered on the "Jacob's ladder", where every rung of the ladder is another class of approximants giving results closer to the exact solution.

The first rung of the ladder is occupied by the local density approximation (LDA) to $E_{xc}[\rho]$, which depends only on the electron density. It is derived considering the model of the uniform electron gas and its generic form reads:

$$E_{xc}^{LDA}[\rho] = \int \epsilon_{xc}^{uniform}\left(\rho(\vec{r})\right)\rho(\vec{r})d\vec{r} \tag{1.22}$$

The expression for the exchange was given by Dirac [16], whereas for the correlation energy nowadays Vosko [17] parametrization of Ceperley-Alder [18] data for homogeneous uniform gas is used.

LDA is a starting point to derive furtherer approximants to $E_{xc}[\rho]$, nevertheless, it performs surprisingly good for many properties (for example geometries, harmonic frequencies). It is prone to failures if the density undergoes rapid changes, as it could be in the case of many molecules.

One step up on the ladder is made by adding the density gradient, $\nabla\rho$, dependence in the $E_{xc}[\rho]$ functional. This leads to group of semi-local functionals, commonly referred to as Generalized Gradient Approximation (GGA). Their general form reads:

$$E_{xc}^{GGA}[\rho] = E_{xc}^{LDA}[\rho] + \int \rho(\vec{r})^{\frac{4}{3}} F\left(\rho(\vec{r}), \nabla\rho(\vec{r})\right) d\vec{r} \tag{1.23}$$

where $F(\rho(\vec{r}), \nabla\rho(\vec{r}))$ is so called enhancement factor. The most popular GGA functionals include exchange functionals by Becke [19] (B88) and Perdew and Wang [20] (BP86), correlation energy functionals by Perdew [21] (P86) or Lee, Yang and Parr [22] (LYP) as well as approximations for exchange-correlation energy by Perdew and Wang [23, 24] (PW91) or Perdew, Burke and Ernzerhof [25] (PBE). The GGA strategy brought many improvements over the LDA and the success of Kohn-Sham formalism as a reliable method to calculate properties of many-electron systems is due to ”robustness” of these functionals.

Climbing up the ladder brings approaches going beyond GGA such as meta-GGA functionals which include dependence on laplasian of the density and kinetic energy density (with TPSS [26] approximant by Tao, Perdew, Staroverov and Scuseria as the most notable example) or the family of hybrid functionals which include dependence on orbitals. For a complete review of available functionals and their accuracy to obtain molecular properties see Ref. [27].

Chapter 2

One-electron equations for embedded orbitals

In Kohn-Sham formalism orbitals are introduced to minimize the errors arising from use of approximate density functionals in calculations of the properties of a system constituted of N-electrons. In this chapter a method allowing for reduction of the orbital space is introduced. Within this formalism, the total system of interest is divided in the two subsystems. The orbital description is retained only for one of them - denoted as the embedded system, whereas the remaining part, commonly denoted as the environment, is described in an *orbital-free* manner using only its electron density $\rho_B(\vec{r})$. As shown by Wesołowski and Warshel [4], for a pair of densities, $\rho_A(\vec{r})$ and $\rho_B(\vec{r})$, an exact *orbital-free embedding potential* can be derived and used in solution of Kohn-Sham-like one-electron equation (Eq. 20 in Ref. [4]) to find ground-state density of the embedded system constituing N_A electrons. For this reason, method described in this chapter is commonly referred as *orbital-free embedding*. Alternatively, in the literature this method is also known as Kohn-Sham Equations with Constrained Electron Density (abbreviated as KSCED). Additionally, in an iterative process where both subsystems switch their roles the Hohenberg-Kohn ground-state energy of the total system can be found. The latter is an equivalent algorithm to solve Cortona formulation of density functional theory [28].

2.1 Orbital-free embedding potential

Using the following elements of the Hohenberg-Kohn-Sham formulation of density functional theory: Hohenberg-Kohn theorems [1], a reference system of non-interacting electrons [2], and the corresponding density functional of the kinetic energy ($T_s[\rho]$) [3] in particular, leads to the embedding potential which is *exact* in the limit of exact functionals and *orbital-free* i.e. does not involve other descriptors of the environment than its electron density [4]. The pure-state non-interacting v-representable electron density ρ_A^{min}, such that added to some arbitrarily chosen density associated with the environment (ρ_B) minimizes the Hohenberg-Kohn energy functional for the whole system, can be obtained from the one-electron equations (Eqs. 20-21 in Ref. [4]) and equation 2.1 in this work:

$$\left[-\frac{1}{2}\nabla^2 + v_{eff}^{KSCED}\left[\rho_A, \rho_B; \vec{r}\right]\right]\phi_i^A = \epsilon_i^A \phi_i^A \qquad i = 1, N^A \tag{2.1}$$

where $\rho_A = 2\sum_i^{N_A} |\phi_i|^2$ and $v_{eff}^{KSCED}\left[\rho_A, \rho_B; \vec{r}\right]$ denotes a local potential which depends only on electron densities ρ_A and ρ_B. The label KSCED (Kohn-Sham Equations with Constrained Electron Density) is used here to indicate that the local potential differs from that in Kohn-Sham equations [2] for either the total system ($v^{KS}[\rho_A + \rho_B; \vec{r}]$)) or the isolated subsystem A ($v^{KS}[\rho_A; \vec{r}]$)). Also the one-electron functions ($\{\phi_i^A\}$) obtained from Eq. 2.1 are not optimal orbitals in neither Kohn-Sham systems. Atomic units are applied in all formulas which are given for spin-unpolarized case. The total effective potential in Eq. 2.1 is the sum of the conventional Kohn-Sham effective potential $v_{eff}^{KS}\left[\rho_A + \rho_B; \vec{r}\right]$ for the whole system evaluated for the electron density $\rho = \rho_A + \rho_B$ and another local potential ($v_t[\rho_A, \rho_B](\vec{r})$):

$$v_{eff}^{KSCED}\left[\rho_A, \rho_B; \vec{r}\right] = v_{eff}^{KS}\left[\rho_A + \rho_B; \vec{r}\right] + v_t[\rho_A, \rho_B](\vec{r}) \tag{2.2}$$

where $v_t[\rho_A, \rho_B](\vec{r})$ involves functional derivatives of the functional $T_s[\rho]$:

$$v_t(\vec{r}) = v_t[\rho_A, \rho_B](\vec{r}) = \left.\frac{\delta T_s[\rho]}{\delta \rho}\right|_{\rho=\rho_A+\rho_B} - \left.\frac{\delta T_s[\rho]}{\delta \rho}\right|_{\rho=\rho_A} \tag{2.3}$$

Note that no restriction is made concerning the overlap between ρ_A and ρ_B in real space.

The potential $v_t[\rho_A, \rho_B](\vec{r})$ can be alternatively expressed as:

$$v_t[\rho_A, \rho_B](\vec{r}) = \left.\frac{\delta T_s^{nad}[\rho, \rho_B]}{\delta \rho}\right|_{\rho=\rho_A} \tag{2.4}$$

where $T_s^{nad}[\rho_A, \rho_B]$ denotes the following difference:

$$T_s^{nad}[\rho_A, \rho_B] = T_s[\rho_A + \rho_B] - T_s[\rho_A] - T_s[\rho_B] \tag{2.5}$$

In view of subsequent discussions, it is convenient to split the total effective potential $v_{eff}^{KSCED}[\rho_A, \rho_B; \vec{r}]$ into two components: the Kohn-Sham effective potential for the isolated subsystem A ($v_{eff}^{KS}[\rho_A; \vec{r}]$), which is ρ_B independent, and the remaining part representing the environment:

$$v_{eff}^{KSCED}[\rho_A, \rho_B; \vec{r}] = v_{eff}^{KS}[\rho_A; \vec{r}] + v_{emb}^{KSCED}[\rho_A, \rho_B; \vec{r}] \tag{2.6}$$

where

$$\begin{aligned} v_{emb}^{KSCED}[\rho_A, \rho_B; \vec{r}] &= v_{ext}^B(\vec{r}) + \int \frac{\rho_B(\vec{r}')}{|\vec{r}' - \vec{r}|} d\vec{r}' \\ &+ \left.\frac{\delta E_{xc}[\rho]}{\delta \rho}\right|_{\rho=\rho_A+\rho_B} - \left.\frac{\delta E_{xc}[\rho]}{\delta \rho}\right|_{\rho=\rho_A} + v_t[\rho_A, \rho_B](\vec{r}) \end{aligned} \tag{2.7}$$

where $E_{xc}[\rho]$ denotes the Kohn-Sham functional of the exchange-correlation energy [2].

2.2 The subsystem formulation of density functional theory

In the subsystem formulation of density functional theory [28], several sets of one-electron functions are used to construct the electron density of each subsystem. Within each set, the one-electron functions are orthogonal.

In the particular case of two subsystems, considered here, a natural choice of the subsystems corresponds to individual molecules forming the complex. The key quantity in this formulation of DFT is the functional referred here as Ξ^S, which depends explicitly on two sets of one-electron functions ($\{\phi_i^A\}$, i=1, N^A, $\{\phi_i^B\}$, i=1, N^B) reads:

$$\begin{aligned}\Xi^S[\{\phi_i^A\},\{\phi_i^B\}] &= V[\rho_A+\rho_B]+J[\rho_A+\rho_B]+E_{xc}[\rho_A+\rho_B] \qquad (2.8)\\ +2\sum_{i=1}^{N^A}\left\langle\phi_i^A\left|-\frac{1}{2}\nabla^2\right|\phi_i^A\right\rangle &+ 2\sum_{i=1}^{N^B}\left\langle\phi_i^B\left|-\frac{1}{2}\nabla^2\right|\phi_i^B\right\rangle+T_s^{nad}[\rho_A,\rho_B]\end{aligned}$$

where

$$\rho_A=2\sum_{i=1}^{N^A}|\phi_i^A|^2 \qquad \rho_B=2\sum_{i=1}^{N^B}|\phi_i^B|^2 \qquad (2.9)$$

The density functionals $E_{xc}[\rho]$, $J[\rho]$, and $V[\rho]$, represent exchange-correlation energy, the Coulomb repulsion, and the energy of the interaction with external field (nuclei), respectively. These functionals are defined in the same way as in the Kohn-Sham formulation of DFT. The bi-functional $T_s^{nad}[\rho_A,\rho_B]$ defined in Eq. 2.5, is expressed by means of the density functional of the kinetic energy in the reference system of non-interacting electrons ($T_s[\rho]$) [3]. In practical calculations based on the Kohn-Sham formalism, the numerical value of $T_s[\rho]$ is available at the end of the self-consistent procedure without relying on any approximated functionals. In calculations based on the subsystem formulation of DFT, only the embedded orbitals are available. They are used to calculate the exact values of $T_s[\rho_A]$ and $T_s[\rho_B]$. The numerical value of of the total kinetic energy $T_s[\rho_A+\rho_B]$ is constructed using the exact results for $T_s[\rho_A]$ and $T_s[\rho_B]$ and the $T_s^{nad}[\rho_A,\rho_B]$ term, which is calculated by means of an approximated functional depending explicitly on two electron densities.

The functional $\Xi^S[\{\phi_i^A\},\{\phi_i^B\}]$ is related to the Hohenberg-Kohn energy functional $E^{HK}[\rho]$ [29]:

$$E^{HK}[\rho_A+\rho_B] = \min_{\{\phi_i^A\}\longrightarrow\rho_A}\ \min_{\{\phi_i^B\}\longrightarrow\rho_B}\ \Xi^S\left[\{\phi_i^A\},\{\phi_i^B\}\right] \qquad (2.10)$$

$$
\begin{aligned}
&= \min_{\{\phi_i^A\}\longrightarrow\rho_A} \Xi^E\left[\{\phi_i^A\},\rho_B\right] \\
&\leq \Xi^E\left[\{\phi_i^A\},\rho_B\right] \leq \Xi^S\left[\{\phi_i^A\},\{\phi_i^B\}\right]
\end{aligned}
$$

The equality is reached for the orbitals, obtained in the constrained search definition of $T_s[\rho]$ [3] provided the total electron density $\rho_A+\rho_B$ is v-representable i.e. the $E^{HK}[\rho_A+\rho_B]$ exists (for a complete discussion of the relation between the universal functionals in Hohenberg-Kohn theorem and their counterparts defined in constrained search see Ref. [30]).

Euler-Lagrange minimization of $\Xi^E[\{\phi_i^A\},\rho_B]$ with respect to $\{\phi_i^A\}$ leads to one-electron equations [4] given in Eq. 2.1.

Note that it is sufficient to know the electron density of the environment ρ_B to express the embedding potential given in Eq. 2.7. No information about the orbital structure of the environment is needed. For this reason, we refer to calculations using Eq. 2.7 as *orbital-free embedding.* In orbital-free embedding calculations [4], ρ_B is required to evaluate the effective embedding potential. The results of embedding calculations depend, therefore, on the choice made for ρ_B. If ρ_B and ρ_A are treated at the same footing, as in the original subsystem formulation of DFT by Cortona, the two sets of orbitals $\{\phi_i^A\}$ and $\{\phi_i^B\}$ minimizing $\Xi^S[\{\phi_i^A\},\{\phi_i^B\}]$ satisfy two sets of coupled equations:

$$\left[-\frac{1}{2}\nabla^2 + v_{eff}^{KSCED}\left[\rho_A,\rho_B\right]\right]\phi_i^A = \epsilon_i^A\phi_i^A \quad i=1,N^A \tag{2.11}$$

$$\left[-\frac{1}{2}\nabla^2 + v_{eff}^{KSCED}\left[\rho_B,\rho_A\right]\right]\phi_i^B = \epsilon_i^B\phi_i^B \quad i=1,N^B \tag{2.12}$$

where, the electron densities and orbitals are related via Eq. 2.9.

At a given external field (geometry of nuclei), minimization of the total energy with respect to ρ_A and ρ_B can be obtained in a self-consistent procedure (*freeze-and-thaw* [31]), in which Eqs. 2.11 and 2.12 are solved consecutively until convergence. In this way, the fully variational calculations based on the subsystem formulation of DFT are formulated as a self-consistent series of *orbital-free embedding* calculations.

The orbitals derived from Eqs. 2.11-2.12 ($\phi_{i(o)}^A$ and $\phi_{i(o)}^B$) yield the electron densities ρ_A^o and ρ_B^o. By construction, ρ_A^o and ρ_B^o are pure-state non-interacting v-representable. Therefore,

$$T_s[\rho_A^o] = 2\sum_{i=1}^{N^A}\left\langle \phi_{i(o)}^A \left| -\frac{1}{2}\nabla^2 \right| \phi_{(o)}^A \right\rangle \tag{2.13}$$

$$T_s[\rho_B^o] = 2\sum_{i=1}^{N^B}\left\langle \phi_{i(o)}^B \left| -\frac{1}{2}\nabla^2 \right| \phi_{(o)}^B \right\rangle \tag{2.14}$$

$$\tag{2.15}$$

In such a case, the right-hand-side of Eq. 2.8 evaluated for $\phi_{i(o)}^A$ and $\phi_{i(o)}^B$ equals exactly to $E^{HK}[\rho_A^o + \rho_B^o]$.

2.3 Orbital-free embedding potential beyond Kohn-Sham

Kohn-Sham density functional theory has been showed in many applications to be the method of choice to treat many chemical problems, providing good balance between cost and accuracy of calculations. There are, however, problems for which KS-DFT is unable to provide correct solution. Most commonly it is related to the flaws of approximations to exchange-correlation energy functional [32, 33, 34] or to the inadequacy of singledeterminantal description for a given system [35]. The orbital-free embedding formalism, which also uses Kohn-Sham non-interacting system of electrons as a reference, will naturally inherit all the difficulties and problems that Kohn-Sham theory meets. It is useful, therefore, to extend applicability of embedding potential to other approaches where this deficiencies does not exist.

2.3.1 Embedding multideterminantal wavefunction in orbital-free environment

Recently Wesołowski has proven [36] that for the multi-configurational *ab-inito* methods, where wavefunction is a basic variable, there can be defined an embedding operator $\hat{V}^{emb}$ accounting for the presence of the environment of a subsystem A:

$$\left[\hat{T}_{2N_A} + \hat{V}_{2N_A}^{ee} + \hat{V}_{ext}^{A} + \hat{V}^{emb}\right]\Psi_A^{MD} = E_A\Psi_A^{MD}, \tag{2.16}$$

and it is postulated to have the form of a local potential:

$$\hat{V}^{emb} = \sum_{i}^{2N_A} v_{loc}^{emb}(\vec{r}_i) \tag{2.17}$$

where the local potential $v_{loc}^{emb}(\vec{r})$ reads:

$$\begin{aligned} v_{loc}^{emb}(\vec{r}) &= v_{ext}^{B}(\vec{r}) + \int \frac{\rho_B(\vec{r}')}{|\vec{r}-\vec{r}'|} d\vec{r}' + \\ & \frac{\delta T_s^{nad}[\rho_A,\rho_B]}{\delta\rho_A(\vec{r})} + \frac{\delta E_{xc}^{nad}[\rho_A,\rho_B]}{\delta\rho_A(\vec{r})} + \frac{\delta\Delta F^{MD}[\rho_A]}{\delta\rho_A} + \text{const} \end{aligned} \tag{2.18}$$

The derivation of $v_{loc}^{emb}(\vec{r})$ is out of scope of this thesis, for complete review please see Ref. [36]. The density $\rho_A(\vec{r})$ is constructed from the multideterminantal functions Ψ_A^{MD} given in eq. 2.16 (the superscript MD is used to denote number of determinants used to express wavefunction, i.e. SD for a single deteterminant):

$$\rho_A(\vec{r}) = \left\langle \Psi_A^{MD} \middle| \sum_{i}^{2N_A} \delta\left(\vec{r}-\vec{r}_i\right) \middle| \Psi_A^{MD} \right\rangle \tag{2.19}$$

while ρ_B is an assumed density of environment.

Careful comparison between $v_{loc}^{emb}(\vec{r})$ given in eq. 2.18 and embedding potential $v_{emb}^{KSCED}[\rho_A,\rho_B;\vec{r}]$ in eq. 2.7 reveals that there is close similarity between both of them. The local potential $v_{loc}^{emb}(\vec{r})$ poses, however, additional term $\frac{\delta\Delta F^{MD}[\rho_{A_o}]}{\delta\rho_{A_o}}$ which arises from the fact the that the electron-electron interaction (V_{ee}) is treated exactly when the embedded system A is described by the multideterminantal wavefunction. The properties of $\Delta F^{MD}[\rho_A]$ will be briefly discussed later in this section.

Equation 2.16 together with eqs. 2.17 and 2.18 allows to find the ground-state wavefunction $\Psi_{A_o}^{MD}$ of the embedded system (and its density via eq. 2.19). The total energy of the whole system can be found by defining energy functional $\Xi^{EWF}[\Psi_A^{MD},\rho_B]$ that satisfy the following condition:

$$\Xi^{EWF}[\Psi_A^{MD},\rho_B] \geq \Xi^{EWF}[\Psi_{A_o}^{MD},\rho_B] = E^{HK}[\rho_{A_o}+\rho_B] \tag{2.20}$$

where $E^{HK}[\rho_{A_o} + \rho_B]$ is Hohenberg-Kohn energy functional. The functional $\Xi^{EWF}[\Psi_A^{MD}, \rho_B]$ reads:

$$\begin{aligned} \Xi^{EWF}[\Psi_A^{MD}, \rho_B] \quad = \quad & \left\langle \Psi_A^{MD} \left| \hat{T}_{2N_A} + \hat{V}_{2N_A}^{ee} \right| \Psi_A^{MD} \right\rangle + \\ & T_s[\rho_B] + J[\rho_B] + E_{xc}[\rho_B] + V_{ext}^B[\rho_B] + V_{ext}^A[\rho_B] + \\ & V_{ext}^A[\rho_A] + V_{ext}^B[\rho_A] + \int\int \frac{\rho_A(\vec{r})\rho_B(\vec{r'})}{|\vec{r} - \vec{r'}|} d\vec{r'} d\vec{r} + \\ & T_s^{nad}[\rho_A, \rho_B] + E_{xc}^{nad}[\rho_A, \rho_B] + \Delta F^{MD}[\rho_A], \end{aligned} \tag{2.21}$$

The functional $\Delta F^{MD}[\rho_A]$ is defined as a difference:

$$\Delta F^{MD}[\rho_A] = F^{HK}[\rho_A] - F^{MD}[\rho_A], \tag{2.22}$$

where $F^{HK}[\rho_A]$ is universal Hohenberg-Kohn functional (see eq. 1.5), while $F^{MD}[\rho_A]$ reads:

$$F^{MD}[\rho_A] = \min_{\Psi_A^{MD} \longrightarrow \rho_A} \left\langle \Psi_A^{MD} \left| \hat{T}_{2N_A} + \hat{V}_{2N_A}^{ee} \right| \Psi_A^{MD} \right\rangle \tag{2.23}$$

The domain of the wavefunctions used to search for the minimum of $F^{MD}[\rho_A]$ is limited to wavefunctions with fixed number of determinants which are used to construct Ψ_A^{MD}. This is not the case for $F^{HK}[\rho_A]$ where the minimum if performed over all the wavefunctions giving target density. If, however, Ψ_A^{MD} will be full CI wavefunction (MD equal to CI) the search domain for $F^{MD}[\rho_A]$ and $F^{HK}[\rho_A]$ will be identical, therefore $\Delta F^{MD}[\rho_A] = 0$. This provides the upper bound for $\Delta F^{MD}[\rho_A]$. The lower bound can be found when Ψ_A^{MD} is singledeterminantal wavefunction ($MD = SD$). In this case the difference $F^{HK}[\rho_A] - F^{MD}[\rho_A]$ will be the biggest and equal to $E_c[\rho]$.

The above equations provide general framework for methods where the system described by multideterminantal wavefunction is embedded in orbital-free environment. Applicability of such scheme requires, however, to rely on the approximations to $E_{xc}[\rho]$, $T_s^{nad}[\rho]$ and $\Delta F^{MD}[\rho_A]$. Several groups successfully used such combination of methods in practical applications. For a representative publications see Refs. [37] and [38].

2.3.2 Embedding one-electron reduced density matrix

In reduced density matrix functional theory (RDMFT) instead of electron density $\rho(\vec{r})$ as in DFT, the basic variable is one-electron reduced density matrix (commonly abbreviated as one-matrix), which for closed-shell 2N electron system is given by the equation:

$$\Gamma(\vec{r}, \vec{r}') \equiv 2N \int \Psi^*(\vec{r}', \ldots \vec{r}_N)\Psi(\vec{r}, \ldots \vec{r}_N) d\vec{r}_2 \ldots \vec{r}_N \tag{2.24}$$

The electron density of the system is given by the diagonal elements of the one-matrix:

$$\rho(\vec{r}) = \Gamma(\vec{r}, \vec{r})$$

The extended description of the density matrix functional theory formalism, involved approximations and practical calculations are out of scope of this thesis. Only brief description necessary to keep this section self-contained will be given. This section is based on Ref. [39]. For more detailed review of RDMFT formalism, see Ref. [8, 40, 41, 42, 43]

In RDMFT the energy is a functional of one-matrix:

$$E[\Gamma] = T[\Gamma] + \int \Gamma(\vec{r}, \vec{r})\ \upsilon_{ext}(\vec{r})\ d\vec{r} + E_{ee}[\Gamma] \tag{2.25}$$

where the kinetic energy functional $T[\Gamma]$ reads:

$$T[\Gamma] = -\frac{1}{2} \int \int \delta(\vec{r} - \vec{r}') \nabla^2_{\vec{r}} \Gamma(\vec{r}, \vec{r}')\ d\vec{r}\ d\vec{r}' \tag{2.26}$$

electron-electron interaction is defined by Levy constrained search procedure:

$$E_{ee}[\Gamma] = \min_{\Psi \to \Gamma} \langle \Psi | \hat{V}_{ee} | \Psi \rangle \tag{2.27}$$

and $\upsilon_{ext}(\vec{r})$ defines local external potential.

The ground state energy is found by minimizing the total energy functional, where the search for minimum is performed over all the N-representable one-matrices:

$$E_0 = \min_{N-\mathrm{rep}\ \Gamma} E[\Gamma] \tag{2.28}$$

It leads to the following Euler-Lagrange equation:

$$\delta(\vec{r}-\vec{r}')\left[-\frac{1}{2}\nabla^2_{\vec{r}}+\upsilon_{ext}(\vec{r})\right]+\upsilon_{ee}(\vec{r},\vec{r}')=\mu\ \delta(\vec{r}-\vec{r}') \tag{2.29}$$

Pernal and Wesołowski showed in Ref. [39] that in the case when subsystem A, described by one-matrix Γ_A, is embedded by subsystem B, described by density ρ_B (where total system density is given by $\rho=\rho_A+\rho_B$), the total energy functional in such a case can be written as:

$$\begin{aligned}\Xi^{\Gamma}[\Gamma_A,\rho_B] &= E[\Gamma_A]+\int\rho_A(\vec{r})\upsilon^B_{ext}(\vec{r})\,d\vec{r}\\ &+ 2J[\rho_A,\rho_B]+T^{nad}_s[\rho_A,\rho_B]+E^{nad}_{xc}[\rho_A,\rho_B]\\ &+ E^{HK}[\rho_B]+\int\rho_B(\vec{r})\upsilon^A_{ext}(\vec{r})\,d\vec{r}\end{aligned} \tag{2.30}$$

where $E[\Gamma_A]$ is defined in 2.25.

The ground state energy of such system is found by minimization of the functional given in eq. 2.30:

$$E^{emb}=\min_{\Gamma_A\to 2N-2N_B}\Xi^{\Gamma}[\Gamma_A,\rho_B] \tag{2.31}$$

where the search is performed over all N-representable embedded one-matrices Γ_A. This procedure leads to the following variational equation for $\Xi^{\Gamma}[\Gamma_A,\rho_B]$:

$$\frac{\delta\Xi^{\Gamma}[\Gamma_A,\rho_B]}{\delta\Gamma_A(\vec{r},\vec{r}')}=\frac{\delta E[\Gamma_A]}{\delta\Gamma_A(\vec{r},\vec{r}')}+\delta(\vec{r}-\vec{r}')\upsilon^{emb}_{loc}(\vec{r})=\mu\ \delta(\vec{r}-\vec{r}') \tag{2.32}$$

The variation $\frac{\delta E[\Gamma_A]}{\delta\Gamma_A(\vec{r},\vec{r}')}$ is given in eq. 2.29. The local embedding potential $\upsilon^{emb}_{loc}(\vec{r})$ defined as:

$$\upsilon^{emb}_{loc}(\vec{r})=\upsilon^B_{ext}(\vec{r})+\int\rho_B(\vec{r}')|\vec{r}-\vec{r}'|^{-1}\ d\vec{r}'+\frac{\delta T^{nad}_s[\rho_A,\rho_B]}{\delta\rho_A(\vec{r})}+\frac{\delta E^{nad}_{xc}[\rho_A,\rho_B]}{\delta\rho_A(\vec{r})} \tag{2.33}$$

It can be easily seen, that $\upsilon^{emb}_{loc}(\vec{r})$ in eq. 2.33 used to embed one-electron reduced density matrix has the same form as embedding potential defined for reference system of non-interacting electrons given in eq. 2.7. Density matrix functional theory has, however, several advantages over Kohn-Sham formulation of DFT. First of all, the wavefunction used to construct one-matrix (eq. 2.24) corresponds to the system of N-interacting electrons. This allows to extend the search domain of optimal densities minimizing the energy functional from pure-state non-interacting v-representable to much wider class of N-representable densities. Moreover, the kinetic energy functional is an explicit functional of one-matrix (eq. 2.26). The RDMFT is still subject of extensive study in various groups [42, 44], nevertheless the proof of existence of local embedding potential $v_{loc}^{emb}(\vec{r})$ in RMDFT case given by Pernal and Wesołowski makes it readily available for any practical calculations applying this formalism.

Chapter 3

Time Dependent Density Functional Theory

Reviewed in previous chapter Hohenberg-Kohn density functional theory and its reformulation given by Kohn and Sham allow to calculate the properties of many-electron systems from their electron density. The application of this approach is limited, however, only to the ground-state properties of molecules. Many chemical problems in the fields of spectroscopy, photochemistry or molecular optics for instance, where the response of the molecule to the non-static frequency dependent external electromagnetic field is subject of interest, are out of scope. In 1984 Runge and Gross [45] formulated theorems which allowed to extend DFT into time-dependent domain, where the response of the system to the perturbations in the external field can be described. In this chapter the most popular approach to treat the time-dependent external potentials will be presented. It is based on linear response equations obtained from perturbative expansion of the time-dependent Kohn-Sham equations. The extension of linear response DFT to the orbital-free embedding formalism will be also given.

3.1 Runge and Gross theorems

The foundations of the time dependent density functional theory are provided by theorems formulated by Runge and Gross [45]. It is stated that for every time-dependent external potential, $v_{ext}(\vec{r}, t)$, that can be expanded in Taylor series around time coordinate $t = t_0$, there is an invertible map (up to the spatially constant time dependent function) between this potential and time dependent charge density $\rho(\vec{r},t)$. Such a map is given by solving time-dependent Schrödinger equation with fixed initial state $\Psi(t_0) = \Psi_0$. This statement is an analog of the first Hohenberg-Kohn theorem from the time-independent theory.

The analog of the second Hohenberg-Kohn theorem for the time-dependent density functional theory is given by defining the action functional $A[\rho]$:

$$A[\rho] = B[\rho] - \int_{t_0}^{t_1} \int v_{ext}(\vec{r},t)\rho(\vec{r},t)d\vec{r}dt \tag{3.1}$$

where $B[\rho]$ is an universal (i.e. independent of the external potential) functional of the density. In this chapter we will limit ourselves to the practical extension of time dependent functional theory comparable to the Kohn-Sham formalism, therefore, rigorous development and definition of the functional $B[\rho]$ following Runge and Gross [45] will not be given. Instead we will consider Kohn-Sham system of non-interacting electrons for which this functional can be written as:

$$\begin{aligned} B[\rho] = \sum_i \int_{t_0}^{t_1} \langle \phi_i(\vec{r},t)| i\frac{\partial}{\partial t} - \frac{1}{2}\nabla^2 |\phi_i(\vec{r},t)\rangle dt \\ -\frac{1}{2}\int_{t_0}^{t_1} \int\int \frac{\rho(\vec{r},t)\rho(\vec{r}',t)}{|\vec{r}-\vec{r}'|} d\vec{r}d\vec{r}'dt - A_{xc}[\rho] \end{aligned} \tag{3.2}$$

where term $A_{xc}[\rho]$ is an analog of the exchange-correlation functional found in the time-independent case.

Functional $A[\rho]$ given in Eq. 3.1 has a stationary point at the exact time-dependent density:

$$\frac{\delta A}{\delta \rho(\vec{r},t)} = 0 \tag{3.3}$$

where the density $\rho(\vec{r},t)$ can be calculated as:

$$\rho(\vec{r},t) = \sum_i f_i |\phi_i(\vec{r},t)|^2 \tag{3.4}$$

where f_i is occupation number.

One-electron orbitals $\phi(\vec{r},t)$ can be obtained from solution of the time-dependent Kohn-Sham equation:

$$\left(-\frac{1}{2}\nabla^2 + v_{eff}(\vec{r},t)\right)\phi_i(\vec{r},t) = i\frac{\partial}{\partial t}\phi_i(\vec{r},t) \tag{3.5}$$

The time-dependent effective potential $v_{eff}(\vec{r},t)$ is defined as:

$$\begin{aligned} v_{eff}(\vec{r},t) &= v_{ext}(\vec{r},t) + \int \frac{\rho(\vec{r},t)}{|\vec{r}-\vec{r}'|}d\vec{r}' + v_{xc}[\rho](\vec{r},t) \\ &= v_{ext}(\vec{r},t) + v_{SCF}(\vec{r},t) \end{aligned} \tag{3.6}$$

with the exchange-correlation potential given as a functional derivative:

$$v_{xc}[\rho](\vec{r},t) = \frac{\delta A_{xc}[\rho]}{\delta\rho(\vec{r},t)} \tag{3.7}$$

In the spin-unrestricted case, eq. 3.5 is written as:

$$\left(-\frac{1}{2}\nabla^2 + v^{\sigma}_{eff}(\vec{r},t)\right)\phi_{i\sigma}(\vec{r},t) = i\frac{\partial}{\partial t}\phi_{i\sigma}(\vec{r},t) \tag{3.8}$$

where σ is the spin coordinate. The total density is equal to

$$\rho(\vec{r},t) = \rho_\alpha(\vec{r},t) + \rho_\beta(\vec{r},t) = \sum_\sigma\sum_i f_{i\sigma}|\phi_{i\sigma}(\vec{r},t)|^2 \tag{3.9}$$

and effective potential:

$$\begin{aligned} v^{\sigma}_{eff}(\vec{r},t) &= v_{ext}(\vec{r},t) + v^{\sigma}_{SCF}(\vec{r},t) \\ &= v_{ext}(\vec{r},t) + \int \frac{\rho(\vec{r},t)}{|\vec{r}-\vec{r}'|}d\vec{r}' + v_{xc}[\rho_\alpha,\rho_\beta](\vec{r},t) \end{aligned} \tag{3.10}$$

Similarly as in the Kohn-Sham theory the time-dependent exchange-correlation functional and its derivative needs to be approximated. This issue and involved simplifications will be discussed in more detail later in this chapter.

Time-dependent Kohn-Sham equation provides a starting point to consider the response of the system to the external time-dependent perturbation. According to the first theorem of Runge and Gross, the external potential is determined by charge density only, therefore, it is justified to separate the perturbations from electric and magnetic fields. In this chapter only the former will be discussed. The response of the system to the magnetic fields is domain of current density functional theory [46].

3.2 Linear response of the system

In popular spectroscopic techniques, like UV/Vis spectroscopy for instance, the response of the molecule to the time-dependent electric field is studied. If the applied electric field is small, the system responds in an linear fashion, i.e. the magnitude of the change of the given observable depends linearly on the strength of electric field. In such a case, the induced molecular dipole moment can be expanded in the Taylor series in different orders of the applied electric field [47]:

$$\mu_{ind} = \mu_0 + \alpha E + \frac{1}{2!}\beta E^2 + \frac{1}{3!}\gamma E^3 + ... \tag{3.11}$$

where E is the applied electric field, μ_0 is the permanent dipole moment, α, β and γ are linear polarizability, hyperpolarizability and second order hyperpolarizability tensors respectively. If the external electric field is non-static, i.e. $E = E_0 + E_\omega \cos(\omega t)$, α, β and γ depend on the frequency ω of the applied field.

Linear response of the system is determined by the linear polarizability tensor α. In present work we will limit ourselves only to such instances. If the perturbation comes from a stronger external field, for example due to laser beam used in transistent absorption spectroscopy experiments where such effects as second harmonics generation can be observed, higher order terms β and γ start to play important role.

The mean frequency dependent linear polarizability can be expressed as a trace of a tensor:

$$\bar{\alpha}(\omega) = \frac{1}{3}\mathrm{Tr}\alpha(\omega) = \sum_i \frac{f_i}{\omega_i^2 - \omega^2} \tag{3.12}$$

where ω_i are the excitation energies. Within linear response theory the excitation energies can be, therefore, obtained as a poles of averaged linear polarizability with oscillator strengths given by term f_i.

3.3 Linear response equations

The response of the system to the perturbation in the external field $\delta v_{ext}(\vec{r}, t)$ can be calculated on the basis of time dependent Kohn-Sham equation (Eq. 3.8). The perturbation to the total effective potential (Eq. 3.10) in the linear response formalism is written as:

$$\delta v^{\sigma}_{eff}(\vec{r}, t) = \delta v_{ext}(\vec{r}, t) + \delta v^{\sigma}_{SCF}(\vec{r}, t) \tag{3.13}$$

The self-consistent field $v^{\sigma}_{SCF}(\vec{r}, t)$ reacts instantaneously to the changes of the density. Its response depends on the linear response of the Kohn-Sham density matrix $\delta P_{ij\sigma}(\omega)$ (note: the Fourier transform from time domain t to frequency domain ω, i.e. $f(\omega) = \int_{-\infty}^{\infty} e^{-i\omega} f(t)dt$, has been used) [48]:

$$\delta v^{SCF}_{ia\sigma}(\omega) = \sum_{jb\tau} K_{ia\sigma,jb\tau}(\omega) \delta P_{jb\tau}(\omega) \tag{3.14}$$

The notation where letters i, j stands for occupied and a, b for unoccupied orbitals has been adopted in the above equation. Greek letters σ and τ denote spin coordinates.

The change in the density matrix is written (in the basis of the unperturbed molecular orbitals):

$$\delta P_{ia\sigma}(\omega) = \frac{f_{a\sigma} - f_{i\sigma}}{\omega - (\epsilon_{i\sigma} - \epsilon_{a\sigma})} \delta v^{eff}_{ia\sigma}(\omega) \tag{3.15}$$

where $\epsilon_{i\sigma}$ is orbital energy.

The term $K_{ia\sigma,jb\tau}(\omega)$ in equation 3.14 is so-called coupling matrix, which describes the linear response of self-consistent field. It can be divided into two terms [49] - the Coulomb part and exchange-correlation part:

$$K_{ij\sigma,kl\tau} = K^{Coul}_{ij\sigma,kl\tau} + K^{xc}_{ij\sigma,kl\tau} \tag{3.16}$$

with

$$K^{Coul}_{ij\sigma,kl\tau} = \int d\vec{r} \int d\vec{r}' \phi_{i\sigma}\phi_{j\sigma} \times \frac{1}{|\vec{r} - \vec{r}'|}\phi_{k\tau}\phi_{l\tau} \tag{3.17}$$

and

$$K^{xc}_{ij\sigma,kl\tau} = \int d\vec{r} \int d\vec{r}' \phi_{i\sigma}\phi_{j\sigma} \times f^{\sigma\tau}_{xc}(\vec{r}, \vec{r}', \omega)\phi_{k\tau}\phi_{l\tau} \tag{3.18}$$

Term $f_{xc}(\vec{r}, \vec{r'}, \omega)$ in Eq. 3.18 is exchange-correlation kernel, which is a derivative of exchange-correlation potential with respect to the time-dependent density:

$$f_{xc}(\vec{r}, \vec{r}', t, t') = \frac{\delta v_{xc}[\rho](\vec{r}, t)}{\delta \rho(\vec{r}', t')} \tag{3.19}$$

The exchange-correlation potential is defined in equation 3.7. The functional $A_{xc}[\rho]$ is an unknown quantity and it needs to be approximated. In practice the commonly used approach is the *adiabatic approximation* where it is assumed that for the external potential, which varies slowly in time, the following can be written [48]:

$$A_{xc} = \int_{t_0}^{t_1} E_{xc}[\rho_t]dt \tag{3.20}$$

where term E_{xc} is an exchange correlation functional known from static Kohn-Sham theory and $\rho_t(\vec{r}) = \rho(\vec{r}, t)$ is a density calculated at a given time t. Within the adiabatic approximation eq. 3.7 can be rewritten as:

$$v_{xc}[\rho](\vec{r}, t) \approx v_{xc}[\rho_t](\vec{r}) = \frac{\delta E_{xc}[\rho_t]}{\delta \rho_t(\vec{r})} \tag{3.21}$$

from which follows the expression for kernel f_{xc} (eq. 3.19):

$$f_{xc}(\vec{r}, \vec{r}', t, t') = \frac{\delta v_{xc}[\rho](\vec{r}, t)}{\delta \rho(\vec{r}', t')} \approx \delta(t - t')\frac{\delta v_{xc}[\rho_t](\vec{r})}{\delta \rho_t[\vec{r}']} \tag{3.22}$$

The simplest approximation to f_{xc} used in practical calculations is adiabatic local density approximation (dubbed ALDA) which is derived from the LDA expression for $v_{xc}[\rho]$.

Coming back to the calculations of excitation energies and corresponding oscillator strengths, it was shown in eq. 3.12 that they can be calculated from the dynamic polarizability α. Alternatively, the dynamic polarizability might be also expressed using the density matrix and written in general form as [48] (for the xz component):

$$\alpha_{xz}(\omega) = -2 \sum_{ij\sigma}^{f_{i\sigma}-f_{j\sigma}>0} x_{ji\sigma}(\mathrm{Re}\,\delta P_{ij\sigma})(\omega)/E_z(\omega) \tag{3.23}$$

where $E_z(\omega)$ is strength of the applied external non-static field ($\delta v_{ext}(\vec{r},t) = zE_z(t)$), $x_{ji\sigma}$ is the matrix element of $\hat{x}$ in the basis of the unperturbed orbitals and $(\mathrm{Re}\,\delta P_{ij\sigma})(\omega)$ is the real part of the density matrix.

Expression for the real part of the density matrix can be derived starting from equations 3.13, 3.14, 3.15 (for details see Ref. [48, 49, 50]). It reads:

$$\begin{aligned}\sum_{bj\tau}\left[\delta_{\sigma\tau}\delta_{ab}\delta_{ij}(\varepsilon_i-\varepsilon_a) - 2K_{ia\sigma,jb\tau} - \omega^2\frac{\delta_{\sigma\tau}\delta_{ab}\delta_{ij}}{(\varepsilon_i-\varepsilon_a)}\right](\mathrm{Re}\,\delta P_{jb\tau})(\omega) \\ = [\delta v_{ext}(\omega)]_{ia\sigma}\end{aligned} \tag{3.24}$$

The above equation can be further simplified to the eigenvalue equation [48, 49]:

$$\Omega\vec{\mathrm{F}}_i = \omega_i^2\vec{\mathrm{F}}_i \tag{3.25}$$

where four-index matrix Ω is given as:

$$\Omega_{ia\sigma,jb\tau} = \delta_{\sigma\tau}\sigma_{ij}\sigma_{ab}(\varepsilon_a-\varepsilon_i)^2 + 2\sqrt{(\varepsilon_a-\varepsilon_i)}K_{ia\sigma,jb\tau}\sqrt{(\varepsilon_b-\varepsilon_j)} \tag{3.26}$$

The eigenvalues ω in equation 3.25 are the excitation energies, whereas the corresponding oscillator strengths are obtained from eigenvectors $\vec{\mathrm{F}}_i$.

Equation 3.25 needs to be solved iteratively. In practice it is commonly treated in a post-Kohn-Sham calculations, where the unperturbed Kohn-Sham occupied and unoccupied orbitals are used to calculate required matrix elements and provide a reliable starting point.

3.4 Extension of orbital-free embedding formalism to include time-dependent density functional linear response

The generalization of the orbital-free embedding formalism to include TD-DFT linear response has been given by Casida and Wesolowski [51], where the embedded orbitals for subsystem A obtained from eq. 2.1 are used to construct the response of the system. This approach requires additional modification of the response kernel (see Eqs. 3.19 and 3.22) to include terms arising from the embedding potential. In the adiabatic approximation it reads (see supplementary information to Ref. [52]):

$$f^{total}(\vec{r},\vec{r}') = f^{free}(\vec{r},\vec{r}') + f^{emb}(\vec{r},\vec{r}') \tag{3.27}$$

where $f^{free}(\vec{r},\vec{r}')$ is the kernel for the isolated molecule (given in eq. 3.22) and $f^{emb}(\vec{r},\vec{r}')$ has the form:

$$f^{emb}(\vec{r},\vec{r}') = \frac{\delta^2 E_{xc}[\rho](\vec{r})}{\delta\rho(\vec{r})\delta\rho(\vec{r}')} - \frac{\delta^2 E_{xc}[\rho_A](\vec{r})}{\delta\rho_A(\vec{r})\delta\rho_A(\vec{r}')} + \frac{\delta^2 T_s^{nad}[\rho_A,\rho_B](\vec{r})}{\delta\rho_A(\vec{r})\delta\rho_A(\vec{r}')} \tag{3.28}$$

In the above equation it is assumed that the response of the subsystem B, i.e. the environment, is negligible. This approximation (furtherer abbreviated as NDRE - neglect of dynamic response of environment) will be valid only for the interfaces for which excitations are localized only on the embedded system and in the cases where the two subsystems does not absorb in the same spectral range.

Considering practical approximations to the kinetic part of the $f^{emb}(\vec{r},\vec{r}')$, it is derived from Thomas-Fermi [9, 10] formula for the kinetic energy of uniform electron gas.

The strategy to calculate the response of the density to the perturbation from the non-static external field of only the embedded system brings several advantages over the supermolecular treatment of the whole interface. It gives a direct access to the excitations localized only on chromophore. Moreover, as it was shown by Neugebauer [53] and others [54, 55, 56, 57], the quality of the results from TD-DFT strongly depends on the size of the system. With the increasing number of the atoms, non-physical low-lying charge transfer excitations start to appear, complicating extraction of the excitation of interest. Limiting the orbital space only to the confined system overcomes this problem. Additionally, it allows to easily identify the same absorption bands in different solvents. Also large environments can be included in calculations without sacrificing the consistency of the results.

The accuracy of the calculations of the time-dependent response of the embedded system hinges on several factors. First and the most obvious one is the adequacy of the NDRE approximation. It depends on the studied molecule, therefore, it needs to be examined case by case. The others are related to the functionals for the exchange-correlation energy and kinetic energy used to derive the respective terms in the embedding potential $v_{emb}^{KSCED}[\rho_A, \rho_B; \vec{r}]$ (eq. 2.7). Note that both $v_t[\rho_A, \rho_B](\vec{r})$ and exchange correlation components of $v_{emb}^{KSCED}[\rho_A, \rho_B; \vec{r}]$ are short-range terms and they disappear with the increasing distance from the embedded molecule. For this reason, the magnitude of errors arising from used approximate functionals can be expected to be the largest close to subsystem A, i.e. in the first solvatation shell, whereas on large distances $v_{emb}^{KSCED}[\rho_A, \rho_B; \vec{r}]$ becomes exact electrostatic embedding potential. Based on the benchmark study for the nucleic acids bases [52], the maximal deviation of the linear response orbital-free embedding calculations from the LR-TD-KSDFT is estimated to be 0.04 eV.

The beyond NDRE-approximation extension of linear response of subsystem density functional theory was developed by Neugebauer, where it is possible to include selected couplings between excitations on different subsystems [58]. As it was shown on the example of benzaldehyde dimer, where both chromophores absorb in the same spectral range, this approach leads to excellent results. For the most recent review on this method, please see Ref. [59].

Part II

Developments in Embedding Theory

Chapter 4

Orbital-free effective embedding potential at nuclear cusps

A strategy to construct approximants to the kinetic-energy-functional dependent component ($v_t[\rho_A, \rho_B](\vec{r})$) of the effective potential in one-electron equations for orbitals embedded in a frozen density environment (Eqs. 20-21 in [Wesołowski and Warshel, *J. Phys. Chem.* **97**, (1993) 8050]) is proposed. In order to improve the local behavior of the orbital-free effective embedding potential near nuclei in the environment, the exact behavior of $v_t[\rho_A, \rho_B](\vec{r})$ at $\rho_A \longrightarrow 0$ and $\int \rho_B d\vec{r} = 2$ is taken into account. As a result, the properties depending on the quality of this potential are invariably improved compared to the ones obtained using conventional approximants which violated the considered exact condition. The approximants obtained following the proposed strategy and especially the simplest one constructed in this work, are non-decomposable i.e. cannot be used to obtain the analytic expression for the functional of the total kinetic energy.

4.1 Motivation

In the quest for an orbital-free method to study many-electron systems initiated in the works of Thomas [9] and Fermi [10], an universally applicable approximation to the kinetic energy as density functional ($T[\rho]$) remains an unsolved issue. The functional $T_s[\rho]$ [3, 60] of the kinetic energy of non-interacting electrons is strongly related to this quest. For molecular or atomic systems, the differences between the two functionals are rather small [7]. Approximations to $T_s[\rho]$ are not needed in the Kohn-Sham formulation of density functional theory [2, 1] because of the availability of the Kohn-Sham orbitals. This successful formal framework, although not orbital-free, could be converted into the orbital-free method if both $T_s[\rho]$ and the exchange-correlation energy functional could be reasonably approximated by means of explicit density functionals. Currently, orbital-free methods are not robust enough to be nearly as universally applicable as the Kohn-Sham equations based ones [61, 62]. The exchange-correlation energy functional is a minor component of the total energy, whereas the kinetic energy is a major one. Whereas even the simplest approximations to the exchange-correlation energy functional applied within the Kohn-Sham framework lead to a reasonable description of many important properties of atoms, molecules, and solids, known approximations to $T_s[\rho]$ used to approximate $\frac{\delta T_s[\rho]}{\delta\rho}$ in variational orbital-free calculations lead usually to qualitatively wrong results [62, 63, 64, 65]. Although an acceptable approximation to $\frac{\delta T_s[\rho]}{\delta\rho}$ is the ultimate goal and an indispensable component of any successful orbital-free computational framework, related quantities are frequently subject of numerical comparisons instead. Most commonly, the kinetic energy itself, which is a global quantity, is subject of such analyses [66, 61, 62, 12, 67] or the density of the kinetic energy t_s ($T_s[\rho] = \int t_s(\vec{r})d\vec{r}$) [68, 69], which similarly to $\frac{\delta T_s[\rho]}{\delta\rho}$ is a local quantity. It has to be underlined, however, that t_s is not defined uniquely. For a recent concise review of strategies in approximating $T_s[\rho]$ and challenges involved see Ref. [70]. This makes the relation between accuracies of t_s^{appr} and $\frac{\delta T_s^{appr}[\rho]}{\delta\rho}$ derived from a common approximation to $T_s^{appr}[\rho]$ less straightforward. Since all quantities derived from $T_s[\rho]$ would be exact if the approximation to $T_s[\rho]$ were exact, it is tempting to assume that a good approximation to $T_s[\rho]$ would lead also to good approximations to the related quantities: energy differences such as the non-additive kinetic energy bi-functional ($T_s^{nad}[\rho_A, \rho_B] = T_s[\rho_A + \rho_B] - T_s[\rho_A] - T_s[\rho_B]$) and its functional derivatives ($\frac{\delta T_s^{nad}[\rho_A,\rho_B]}{\delta\rho_X}$, where $X = A$ or B).

Previous analyses concerning the applicability of several semi-local approximations to $T_s[\rho]$ in approximating the functional derivatives of $T_s^{nad}[\rho_A, \rho_B]$ [71, 72, 73], showed that there is no correlation between the errors in $T_s[\rho]$ and that of $\frac{\delta T_s^{nad}[\rho_A,\rho_B]}{\delta\rho_X}$ for the considered approximations. The one with the largest domain of applicability was selected. In the absence of exact reference data for $T_s^{nad}[\rho_A, \rho_B]$, the origin of the lack of correlation between the errors of these closely related quantities could not be investigated further. The work of Bernard *et al.* [13] provided the missing link. Owing to the availability of the exact numerical data for $T_s^{nad}[\rho_A, \rho_B]$, where ρ_A and ρ_B are obtained from a particular partitioning of a four-electron density, the accuracy of $T_s^{nad}[\rho_A, \rho_B]$ could be also analysed. Based on the study of commonly used kinetic energy functionals the lack of correlation along the series

$$T_s[\rho] \longrightarrow T_s^{nad}[\rho_A, \rho_B] \longrightarrow \frac{\delta T_s^{nad}[\rho_A, \rho_B]}{\delta \rho_X} \tag{4.1}$$

was shown. In particular, enforcing the exact properties on the functional $T_s[\rho]$, although it leads to the improvements in $T_s^{nad}[\rho_A, \rho_B]$, it does not improve $\frac{\delta T_s^{nad}[\rho_A,\rho_B]}{\delta \rho_X}$, as it was demonstrated for the Second Order Gradient Expansion (GEA2).

The lack of correlation between the elements of the series given in Eq. 4.1 suggest alternative, bottom-up approach of constructing the approximants to the non-additive kinetic energy, where the object of interest is not the parent expression for $T_s[\rho]$, but directly the functional derivative $\frac{\delta T_s^{nad}[\rho_A,\rho_B]}{\delta \rho_X}$. This strategy allows to build-in in a straightforward way an exact properties derived for $\frac{\delta T_s^{nad}[\rho_A,\rho_B]}{\delta \rho_X}$ (see Appendix A and B in this chapter), moreover, it also leads to the new class of the approximants for the non-additive kinetic energy and its derivative referred as *non-decomposable*, i.e. the analytic form of its two individual components $\left.\frac{\delta \tilde{T}_s[\rho]}{\delta \rho}\right|_{\rho=\rho_A+\rho_B}$ and $\left.\frac{\delta \tilde{T}_s[\rho]}{\delta \rho}\right|_{\rho=\rho_A}$ cannot be reconstructed.

Following this idea, in this chapter construction of the new *non-decomposable* approximant to $\frac{\delta T_s^{nad}[\rho_A,\rho_B]}{\delta \rho_X}$ obeying certain exact condition is presented and validated on several model systems.

4.2 Introduction

Numerical methods to study electronic structure in condensed matter use mainly techniques developed for periodic systems. In many cases, however, methods developed for finite systems are also used. They are especially adequate for ionic solids, liquids, molecular crystals, clusters of molecules, for instance, to study features of the electronic structure which are local in character. In such a case, the electronic structure is modeled only in some well-defined region in space of direct relevance. The effect of the atoms outside of this selected region (referred to as environment in this work) is taken into account by some *embedding potential*. Different strategies are applied in practice to represent the embedding potential. They differ in the choice of descriptor of the environment. The roughest simplification is to neglect the environment entirely. It is commonly made in studies of chemical bonding and reactivity in condensed phase if the solvent in which the reaction takes place is known to play a secondary role. Representing the environment (discrete or continuous, polarizable or not) by the electric field it generates, makes it possible to take into account the effect of the environment [74, 75]. Such classical treatment of the effect of the environment on the electronic structure is commonly used both in chemistry and in materials science (for review see Ref. [76, 77, 78]). The embedding potential in such methods is obviously orbital-free. It is, however, not exact because the quantum statistics nature of electrons is completely neglected. Taking into account the fermion nature of electrons might proceed by following a similar strategy as the one applied by Phillips and Kleinman in the construction of pseudopotentials in order to eliminate explicit treatment of core electrons [79]. For recent developments along these lines, see Ref. [80].

Following the work of Wesołowski and Warshel [4], the exact universal embedding potential can be introduced within the Hohenberg-Kohn-Sham formulation of density-functional theory. Its exact analytic from is given in equation 2.7. This orbital-free effective embedding potential and its $v_t[\rho_A, \rho_B](\vec{r})$ component in particular, are used in various types of multi-level numerical simulations (for a review, see Ref. [29] or Refs. [81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91] for representative recent reports). Such simulations deal with condensed matter systems, for which the electronic features of a selected subsystem (subsystem A) represent the primary target of investigations whereas ρ_B is subject to additional simplifications. Other formal frameworks use also $v_t[\rho_A, \rho_B](\vec{r})$ such as: Cortona's formulation of density functional theory [28], where ρ_B is not an assumed quantity but a result of fully variational calculations [28, 31, 92, 93, 94] or linear-response time-dependent density-functional-theory description of electronic excitations localized in embedded systems [51, 52]. Finally, the orbital-free effective embedding potential given in Eq. 2.7, and its $v_t[\rho_A, \rho_B](\vec{r})$ component in particular, are used in combination with traditional wave-function based methods by Carter and collaborators (see for instance Ref. [38]). For the formal analysis of applicability of such a combination, see Ref. [36] which shows that the exact embedding potential in such a case always comprises the $v_t[\rho_A, \rho_B](\vec{r})$ component.

The analytic form of $v_t[\rho_A, \rho_B](\vec{r})$ is not known. Therefore, approximants ($\tilde{v}_t[\rho_A, \rho_B](\vec{r})$) to this quantity are used in practical calculations. They are expressed explicitly by means of ρ_A, ρ_B, and/or its derivatives.

As a result of using an approximant to $v_t[\rho_A, \rho_B](\vec{r})$, all quantities derived in practical simulations are not exact and lead to $\tilde{v}_t$- *induced errors*. Obviously, other simplifications made in practice such as the use of approximants to the exchange-correlation functional, and the assumed ρ_B contribute as well to the overall error. In the present work, we analyze of the errors associated only with approximants to $v_t[\rho_A, \rho_B](\vec{r})$. It is worthwhile to point out here the recent work by Jacob *et al.* [95] in which an empirical procedure to construct overall errors in the whole orbital-free effective potential given in Eq. 2.7 in the cases where the redistribution of electrons between subsystem A and B at dissociation is known in advance. These authors, however, did not construct an approximant to $v_t[\rho_A, \rho_B](\vec{r})$ but used an explicit position-dependent special term for this purpose.

If an approximant to $\tilde{T}_s[\rho]$ is used in Eqs. 2.3-2.5 to obtain an analytic expression to $\tilde{v}_t[\rho_A, \rho_B](\vec{r})$, the resulting potential will be referred here as *decomposable* because the analytic form of all relevant approximants: $\tilde{T}_s[\rho]$, $\tilde{T}_s^{nad}[\rho_A, \rho_B]$, and $\tilde{v}_t[\rho_A, \rho_B](\vec{r})$, is also available. If the form of the used $\tilde{T}_s[\rho]$ comprises only low-level gradient-expansion [96] contributions, the corresponding decomposable $\tilde{v}_t[\rho_A, \rho_B](\vec{r})$ cannot recover the exact relation given in Eq. 4.8 (see Appendix A). Our interest in the local behavior of $v_t[\rho_A, \rho_B](\vec{r})$ at this particular limit is motivated by the fact that the corresponding conditions occur if ρ_B comprises two electrons tightly bound to a distant nucleus in the environment such as in the case of the helium atom, Li^+ cation, Be^{2+}, etc. These conditions do not apply rigorously for heavier nuclei but we expect that should be of relevance locally in a volume element centered on the nucleus where ρ_B is dominated by a doubly-occupied orbital.

The present work focuses on the investigation whether the exact relation given in Eq. 4.8 is of any practical relevance. To this end, we apply the following strategy: *i)* We use a model system (Appendix B), for which the conditions $\rho_A \longrightarrow 0$ and $\int \rho_B d\vec{r} = 2$ apply rigorously, to analyze the importance of enforcing the correct local behavior of $\tilde{v}_t[\rho_A, \rho_B](\vec{r})$. *ii)* We construct a simple approximant to $v_t[\rho_A, \rho_B](\vec{r})$ obeying the exact relation given in Eq. 4.8 in the vicinity of nuclei. This approximant is used to analyze the numerical significance of the considered exact property of $v_t[\rho_A, \rho_B](\vec{r})$

in real systems where the conditions $\rho_A \longrightarrow 0$ and $\int \rho_B d\vec{r} = 2$ do not apply rigorously.

Our ultimate goal is a new approximant to $v_t[\rho_A, \rho_B](\vec{r})$ which can be inexpensively evaluated in practice and obeys as much as possible of the relevant exact properties. It should pointed out in this context that the position-dependency of $v_t[\rho_A, \rho_B](\vec{r})$ is the result of non-homogeneity of ρ_A and/or ρ_B. Therefore, the symbol $v_t[\rho_A, \rho_B](\vec{r})$ (or $\tilde{v}_t[\rho_A, \rho_B](\vec{r})$ for approximants) is used throughout this work to indicate that this local quantity is a functional of ρ_A and ρ_B. Explicit position dependence is strongly undesired in density-functional-theory based methods because it is not straightforward to obtain *i*) such potential as a functional derivative of some density functional, and *ii*) functional derivatives of such potential needed in some formal frameworks [51, 52]. General symbols such v_t or $v_t(\vec{r})$ are used in some discussions where the issue of explicit position-dependence is not relevant.

Finally, we point out that the quality of the following quantities obtained from Eq. 2.1: orbitals, orbital energies, and the electron density, is directly determined by the used approximant to $v_t[\rho_A, \rho_B](\vec{r})$. The total energies, however, are affected by the errors in the approximant to the potential $v_t[\rho_A, \rho_B](\vec{r})$ but also by the errors in the approximant to $T_s^{nad}[\rho_A, \rho_B]$. Therefore, the quality of the energies cannot be uniquely attributed to that of the used approximant to $v_t[\rho_A, \rho_B](\vec{r})$.

4.3 Conventional (decomposable) approximants to $v_t[\rho_A, \rho_B](\vec{r})$

Before proceeding to the construction of the desired approximant to $v_t[\rho_A, \rho_B](\vec{r})$, which obeys the exact relation given in Eq. 4.8, we overview the conventional construction of approximants to $v_t[\rho_A, \rho_B](\vec{r})$ and the local behavior of the obtained potential near a nucleus. The conventional strategy, which is applied in our own works and the works by others so far, is to start from some explicit density functional $\tilde{T}_s[\rho]$ and to use its analytic form to derive the corresponding approximant to $T_s^{nad}[\rho_A, \rho_B]$:

$$T_s^{nad}[\rho_A, \rho_B] \approx \tilde{T}_s^{nad}[\rho_A, \rho_B] = \tilde{T}_s[\rho_A + \rho_B] - \tilde{T}_s[\rho_A] - \tilde{T}_s[\rho_B] \tag{4.2}$$

and to use the obtained analytic expression to obtain $\tilde{v}_t[\rho_A, \rho_B](\vec{r})$ by means of functional differentiation.

$$v_t[\rho_A, \rho_B](\vec{r}) \approx \tilde{v}_t[\rho_A, \rho_B](\vec{r}) = \left. \frac{\delta \tilde{T}_s^{nad}[\rho, \rho_B]}{\delta \rho} \right|_{\rho=\rho_A} \tag{4.3}$$

This strategy can be applied for any approximant to $\tilde{T}_s[\rho]$ provided its form makes it possible to obtain the analytic expression for $\tilde{v}_t[\rho_A, \rho_B](\vec{r})$. Simple functionals $\tilde{T}_s[\rho]$, which depend explicitly on densities and their gradients, are of particular practical interest. They lead to $\tilde{v}_t[\rho_A, \rho_B](\vec{r})$ which depends explicitly only on ρ_A, ρ_B and their first- and second derivatives. In the original work by Cortona [28], where the subsystem formulation of density functional theory was introduced, a decomposable $\tilde{v}_t[\rho_A, \rho_B](\vec{r})$ derived from the Thomas-Fermi [9, 10] kinetic energy functional was used to study ionic solids. In our own works, only decomposable $\tilde{v}_t[\rho_A, \rho_B](\vec{r})$ derived from gradient-dependent approximants to $T_s[\rho]$ were considered so far (see for instance the analyses of their accuracy in Ref. [72, 13] or their recent applications in multi-level computer simulations of condensed matter [86]).

Thomas-Fermi kinetic energy functional [9, 10], which is exact for the uniform electron gas, leads to the following approximant to $T_s^{nad}[\rho_A, \rho_B]$:

$$\tilde{T}_s^{nad(TF)}[\rho_A, \rho_B] = C_{TF} \int \left((\rho_A + \rho_B)^{5/3} - \rho_A^{5/3} - \rho_B^{5/3} \right) d\vec{r} \tag{4.4}$$

where $C_{TF} = \frac{3}{10}(3\pi^2)^{2/3}$.
The associated expression for $\tilde{v}_t^{TF}[\rho_A, \rho_B](\vec{r})$ reads:

$$\tilde{v}_t^{TF}[\rho_A, \rho_B](\vec{r}) = \frac{5}{3} C_{TF} \left((\rho_A + \rho_B)^{2/3} - \rho_A^{2/3} \right) \tag{4.5}$$

Using the approximant to $T_s[\rho]$ taken from the gradient expansion of the kinetic energy [96] truncated to the second order leads to the following approximant to $T_s^{nad}[\rho_A, \rho_B]$ [4]:

$$\tilde{T}_s^{nad(GEA2)}[\rho_A, \rho_B] = T_s^{nad(TF)}[\rho_A, \rho_B] - \frac{1}{72} \int \frac{|\rho_A \nabla \rho_B - \rho_B \nabla \rho_A|^2}{\rho_A \rho_B (\rho_A + \rho_B)} d\vec{r} \tag{4.6}$$

The associated expression for $\tilde{v}_t^{GEA2}[\rho_A, \rho_B](\vec{r})$ is given in Ref. [4].

For the the group of gradient-dependent approximants to $T_s[\rho]$ of the *generalized gradient approximation* form [72, 71] the analytic expression for $\tilde{T}_s^{nad}[\rho_A, \rho_B]$ reads:

$$\tilde{T}_s^{nad(GGA)}[\rho_A, \rho_B] = C_{TF} \int \left[(\rho_A + \rho_B)^{5/3} F(s_{AB}) - \rho_A^{5/3} F(s_A) - \rho_B^{5/3} F(s_B) \right] d\vec{r} \tag{4.7}$$

where $F(s)$ (*enhancement factor*) depends on a dimensionless quantity $s = \frac{|\nabla\rho|}{2(3\pi^2)^{1/3}\rho^{4/3}}$ (*reduced density gradient*). Various analytic forms of $F(s)$ were proposed in the literature [97, 98, 99, 67]. The associated analytic expression for $\tilde{v}_t^{GGA}[\rho_A, \rho_B](\vec{r})$ is given in Ref. [93]. The GGA form is flexible and includes $\tilde{T}^{TF}[\rho]$ and $\tilde{T}^{GEA2}[\rho]$ as special cases. Numerical values of s provide useful information about shell structure and the distance from the nucleus in atoms [100]. For an atom, s is known to be small near the nucleus, reach the values of about 3 in the valence region, and diverge exponentially to $+\infty$ at large distances. In molecules, it behaves similarly with a noticeable exception of stationary points of electron density (bond midpoints for instance) where $s = 0$. Each approximant given in Eqs. 4.4-4.7 comprises a dominant Thomas-Fermi component and satisfies two exact conditions:

- $T_s[\rho_A + \rho_B] - T_s[\rho_A] - T_s[\rho_B] = 0$ for non-overlapping ρ_A and ρ_B.
- For uniform ρ_A and ρ_B, they recover the exact analytical expression for $T_s[\rho_A + \rho_B] - T_s[\rho_A] - T_s[\rho_B]$.

The common feature of each among the above approximants to $T_s[\rho]$, which are decomposable and gradient-expansion based, is that none of them yields the exact analytic form of $v_t[\rho_A, \rho_B](\vec{r})$ at $\rho_A \longrightarrow 0$ and $\int \rho_B d\vec{r} = 2$ (see Appendix A):

$$v_t[\rho_A, \rho_B](\vec{r}) \longrightarrow v_t^{limit}[\rho_B](\vec{r}) = \frac{1}{8}\frac{|\nabla\rho_B|^2}{\rho_B^2} - \frac{1}{4}\frac{\nabla^2\rho_B}{\rho_B} \tag{4.8}$$

The expression given in Eq. 4.5, which provides the dominant contribution to gradient-expansion based approximants to $v_t[\rho_A, \rho_B](\vec{r})$ does not comprise the relevant term at all whereas the second-order term provides only 1/9 of the exact expression. Figure 4.1 shows $\tilde{v}_{emb}^{KSCED(LDA)}[\rho_A, \rho_B](\vec{r})$ for a spherically symmetric case: $\rho_B = \rho_{He}$, $v_{ext}^B(\vec{r}) = -2/r$, and $\rho_A \longrightarrow 0$,

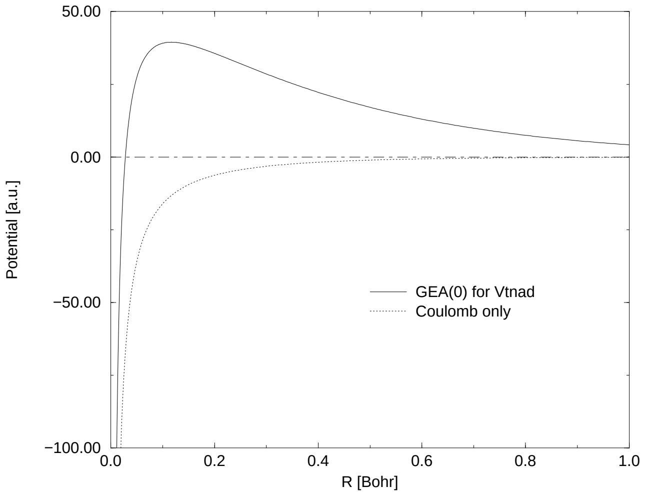


Figure 4.1: Effective potential (Eq. 2.7) derived using local density approximation for its exchange- and kinetic energy components for: $\rho_B = \rho_{He}$, $\rho_A \longrightarrow 0$, and $v^B_{ext}(\vec{r}) = -2/r$.

which represents a helium atom far from subsystem A. In the Figure as well as in the following discussion, r denotes the distance from the nucleus on which ρ_B is centered. The potential shows the features which are common also for heavier atoms if gradient expansion based approximants to $T_s[\rho]$ are used to derive $v_t[\rho_A, \rho_B](\vec{r})$: a very narrow and deep well (reaching $-\infty$) centered on the nucleus and surrounding it repulsive shell. The $v_t[\rho_A, \rho_B](\vec{r})$ component of the shown embedding potential is finite at the nucleus instead of behaving as $\frac{\zeta}{r}$. Note that the exact term has the same form as the potential due to Coulomb attraction by the nucleus of the charge Z. For $\zeta = Z$, this compensation is complete. The commonly used "effective" values of ζ are smaller than than of Z [101]. As a consequence, the compensation is only partial. For the particular case considered in the Figure, the missing $\frac{\zeta}{r}$ component does not lead to any bound states (Appendix B). In general, however, the wrong behavior of the singularity of $v_t(\vec{r})$ at the nucleus, can lead to an unphysical transfer of electron density from the investigated system to its environment (charge-leak [102]). This can occur if an artificially attractive (not sufficiently repulsive) approximant to the orbital-free effective embedding potential generates a bound state in the environment of the energy which is lower than the eigenvalue of the highest occupied embedded orbital associated with embedded subsystem A. Numerical cases confirming such scenario are known [95, 103]. Moreover, the numerical solution of the Schrödinger equation with $\tilde{v}_{emb}^{KSCED(LDA)}[\rho_A, \rho_B](\vec{r})$ for $\rho_B = \rho_{Li^+}$ and the external potential of the $-\frac{3}{r}$ form, shows a deeply lying node-less bound state of the energy -0.209665 hartree with the maximum of the radial electron density at $r^{max} = 2.912$ bohr (Appendix B). The above observations indicate that decomposable $\tilde{v}_t[\rho_A, \rho_B](\vec{r})$ obtained from low-order gradient-expansion based approximants to $T_s[\rho]$ might not be adequate for, at least, Li^+ cations in the environment. The fact that the bound state associated with an atom in the environment is too tightly bound to the nucleus indicates, however, that the problem might be also present in atoms comprising more electrons.

In a subsequent section, a simple approximant to $v_t[\rho_A, \rho_B](\vec{r})$ is constructed based on these observations. In principle, the exact limit should be applied in any volume element in which ρ_A vanishes and ρ_B is obtained from a doubly occupied orbital. In the proposed construction, the exact limit is imposed only at volume elements near heavier-than-hydrogen nuclei the exact relation given in Eq. 4.8 is most relevant there.

4.4 Building-in the exact local behavior of $v_t[\rho_A, \rho_B](\vec{r})$ at $\rho_A \longrightarrow 0$ for $\int \rho_B d\vec{r} = 2$.

The aforementioned flaws of decomposable strategy to construct gradient- and Laplacian dependent approximants to $v_t[\rho_A, \rho_B](\vec{r})$ suggest a bottom-up approach in which $v_t[\rho_A, \rho_B](\vec{r})$ is directly a target. A given approximant $\tilde{v}_t[\rho_A, \rho_B](\vec{r})$ will be referred to as *non-decomposable* if the analytic form of its two individual components $\left.\frac{\delta \tilde{T}_s[\rho]}{\delta \rho}\right|_{\rho=\rho_A+\rho_B}$ and $\left.\frac{\delta \tilde{T}_s[\rho]}{\delta \rho}\right|_{\rho=\rho_A}$ cannot be reconstructed. The non-decomposable strategy is motivated by the fact that there are exact properties of $v_t[\rho_A, \rho_B](\vec{r})$, which can be taken into account quite easily in $\tilde{v}_t[\rho_A, \rho_B](\vec{r})$, whereas building-in them into some approximant to $\tilde{T}_s[\rho]$ is less straightforward. Abandoning the decomposable strategy is motivated also by the results of our recent dedicated studies of the accuracy of various gradient-dependent approximant to $T_s^{nad}[\rho_A, \rho_B]$, which revealed that there is no correlation between the accuracy of $\tilde{T}_s^{nad}[\rho_A, \rho_B]$, $\tilde{v}_t[\rho_A, \rho_B](\vec{r})$ and the errors in the parent gradient-dependent approximant to $\tilde{T}_s[\rho]$ [13]. It should be also pointed out that, that the individual contributions $\tilde{T}_s^{nad}[\rho_A, \rho_B]$ are not needed in practice.

The non-decomposable approximant to $v_t[\rho_A, \rho_B](\vec{r})$ is constructed by enforcing the following exact conditions into its analytic form:

- The uniform electron gas limit:
  $\tilde{T}_s^{nad}[\rho_A, \rho_B] \longrightarrow \tilde{T}_s^{nad(TF)}[\rho_A, \rho_B]$ for uniform ρ_A and ρ_B.
- The zero-overlap limit:
  $\tilde{T}_s^{nad}[\rho_A, \rho_B] \longrightarrow 0$ for non-overlapping ρ_A and ρ_B.
- Convergence of the gradient-dependent terms to the limit given in Eq. 4.8:
  $\tilde{v}_t[\rho_A, \rho_B] \longrightarrow v_t^{limit}[\rho_B]$ at $\rho_A \longrightarrow 0$ and $\int \rho_B d\vec{r} = 2$.

The first two conditions are automatically satisfied by the decomposable gradient-expansion based approximants discussed in the previous section. Since such approximants proved to be sufficiently accurate for many systems the same conditions are retained in the new construction. The last condition is the key element of the present construction.

Before proceeding to the construction of the approximant obeying the considered exact condition, we note that $v_t[\rho_A, \rho_B](\vec{r})$ can be alternatively expressed as:

$$v_t[\rho_A, \rho_B] = \tilde{v}_t^{decomposable}[\rho_A, \rho_B] + f[\rho_A, \rho_B] \cdot v_t^{limit}[\rho_B] \tag{4.9}$$

All functionals in the above equation are determined locally and the argument $\vec{r}$ is not written explicitly for simplicity and the functionals $v_t[\rho_A, \rho_B]$ and $f[\rho_A, \rho_B]$ are simply related ($f[\rho_A, \rho_B] = \frac{v_t[\rho_A,\rho_B] - \tilde{v}_t^{decomposable}[\rho_A,\rho_B]}{v_t^{limit}[\rho_A,\rho_B]}$ if $v_t^{limit}[\rho_A, \rho_B]$ is non-zero). The above form of $v_t[\rho_A, \rho_B]$ provides a convenient starting point for construction of approximants to $v_t[\rho_A, \rho_B]$. For any decomposable $\tilde{v}_t[\rho_A, \rho_B]$ which violates the exact relation given in Eq. 4.8, the functional $f[\rho_A, \rho_B]$ has a clear physical meaning as a *switching factor* determining whether it is needed to add locally the missing component of the embedding potential.

The first approximation made here is replacing the *switching factor* defined in Eq. 4.9 by a *switching function*:

$$f[\rho_A, \rho_B](\vec{r}) \approx \tilde{f}(\rho_A, \rho_B, \nabla\rho_B, \nabla^2\rho_B) \tag{4.10}$$

The above considerations lead to the following general form of the approximant to $v_t[\rho_A, \rho_B](\vec{r})$:

$$\tilde{v}_t[\rho_A, \rho_B] = \tilde{v}_t^{decomposable}[\rho_A, \rho_B] + \tilde{f}(\rho_A, \rho_B, \nabla\rho_B, \nabla^2\rho_B) \cdot v_t^{limit}[\rho_B] \tag{4.11}$$

This form of $\tilde{v}_t[\rho_A, \rho_B]$ provides a clear interpretation for the switching function, which can be used as guideline in construction of approximants - it "detects" such volume elements for which the conditions $\rho_A \longrightarrow 0$ and $\int \rho_B d\vec{r} = 2$ are most relevant and turns on the exact terms if they are not accounted for properly in the first term. It is worthwhile to underline, that the approximant of the form given in Eq. 4.11 does not satisfy the exact relation given in Eq. 4.8 rigorously if the decomposable component comprises the contribution derived from zeroth order in gradient expansion. For appropriately chosen switching function, however, the gradient-dependence of such approximants satisfies the exact limit even in such cases.

Eq. 4.11 represents a general form of a non-decomposable approximant to $v_t[\rho_A, \rho_B]$. Various alternative approximants of this form can be constructed depending on the choice made for its decomposable part and the switching function. As far as the choice for the decomposable component, both the gradient-free potential given in Eq. 4.5 and the decomposable potential derived from the Lembarki-Chermette [98] approximants to $T_s[\rho]$ were shown in dedicated studies [72, 13] to be reasonably accurate if ρ_A and ρ_B do not overlap strongly. Both these approximants comprise the zeroth order contribution. If $\tilde{v}_t^{TF}[\rho_A, \rho_B]$ is used as decomposable component in Eq. 4.9, the terms $v_t^{limit}[\rho_B]$ and $\tilde{v}_t^{decomposable}[\rho_A, \rho_B]$ are local functions depending explicitly on ρ_A, ρ_B, $\nabla\rho_B$, and $\nabla^2\rho_B$. Approximating $f[\rho_A, \rho_B]$ by a local function depending explicitly on these quantities leads to an approximant to $v_t[\rho_A, \rho_B]$ involving a similar computational effort as conventional low-order gradient-expansion based decomposable approximants. The subsequent construction concerns the switching function if the Thomas-Fermi functional is used as the decomposable component.

4.4.1 The switching function $\tilde{f}$ for environments comprising one-nucleus and two-electrons

In constructing $\tilde{f}(\rho_A, \rho_B, \nabla\rho_B, \nabla^2\rho_B)$ the following additional requirements (simplifications) are made:

- $\tilde{f}(\rho_A, \rho_B, \nabla\rho_B, \nabla^2\rho_B)$ is one in the vicinity of a nucleus to account fully for the missing $\frac{\zeta}{r}$ component.
- The criterion for determining the range at which $v_t^{limit}[\rho_B]$ is nuclear number independent.

- $\tilde{f}(\rho_A, \rho_B, \nabla\rho_B, \nabla^2\rho_B)$ does not depend on ρ_A (to obtain the analytic form of $\tilde{T}_s^{nad}[\rho_A, \rho_B]$: $(f[\rho_A, \rho_B] \approx \tilde{f}[\rho_B])$

The above criteria are very restrictive and leave us with not many choices. The last one leads to the following form of the switching factor:

$$f[\rho_A, \rho_B] \approx \tilde{f}(\rho_B, \nabla\rho_B, \nabla^2\rho_B) \tag{4.12}$$

Approximating $f[\rho_A, \rho_B]$ by some function $f(\rho_B)$ is one of possible further simplifications. It is, however, very unlikely that a ρ_B-based switching function could be universal. The electron density near the nucleus depends on the effective nuclear charge ζ [101] and varies strongly from atom to atom. It is possible, however, to design an ζ-independent criterion. To this end, we consider the reduced density gradient $(s_B(\vec{r}))$ defined as:

$$s_B = \frac{|\nabla\rho_B|}{2(3\pi^2)^{1/3}\rho_B^{4/3}} \tag{4.13}$$

For ρ_B obtained from hydrogenic orbital $1s$ defined by some effective nuclear charge ζ, the following ζ-independent observations can be made: *i)* For $\rho_B^{1s} = 2|1s|^2$, $s_{r=0} = (6\pi)^{-1/3} = 0.376$, and *ii)* $v_t^{limit}[\rho_B^{1s}]$ changes sign from positive to negative at $s_{r=2/\zeta} = \exp(4/3) \cdot s_{r=0} = 1.426$. These observations suggest that the switching function can take a very simple and ζ-independent form

$$\tilde{f}(\rho_B, \nabla\rho_B) = \tilde{f}(s_B) = \Theta(s_B - s_B^{min}) \times \Theta(s_B^{max} - s_B) \tag{4.14}$$

where $\Theta(x) = 1$ for $x \geq 0$ and $\Theta(x) = 0$ for $x < 0$ and $s_B^{min} = 0.376$ and $s_B^{max} = 1.426$.

Since Eq. 2.1 can be used also to obtain forces [94], it is preferable to use a smooth switching from 0 to 1 instead of Θ in the above definition (see Appendix C).

The obtained switching factor $\tilde{f}$ allows us to investigate the importance of enforcing the exact relation given in Eq. 4.8 in cases where the environment comprises one nucleus and two electrons at large separation from subsystem A. Not only the condition $\int \rho_B d\vec{r} = 2$ applies rigorously, but the considerations leading to the ζ-independent values of s_B^{max} and s_B^{min} apply as well. Such model systems as the complexes Li^+-H_2O and Be^{2+}-H_2O with the water molecule playing the role of subsystem A fall into this category. At equilibrium geometry, adding the $\tilde{f} \cdot v_t^{limit}[\rho_B]$ term results in a desired effect on calculated properties for these complexes. The lowest unoccupied embedded orbital associated with subsystem A is indeed localized on the cation and its energy is shifted by 0.262 eV in the case of Li^+ and by 0.830 eV in the case of Be^{2+}. Addition of $\tilde{f} \cdot v_t^{limit}[\rho_B]$ reduces also the dipole moment of a water molecule in the vicinity of the cation by 0.071 Debye and 0.409 Debye for Li^+ and Be^{2+}, respectively. Such noticeable numerical effects obtained for systems, for which the condition $\int \rho_B d\vec{r} = 2$ applies rigorously, indicate clearly that the exact relation given in Eq. 4.8 is relevant for practical calculations where the environment is larger.

4.4.2 Fine-tuning the thresholds in the switching function $\tilde{f}$

Using only universal parameters s_B^{min} and s_B^{max} is very appealing. The considerations of the previous section leading to their numerical values of these thresholds do not apply in real systems. The construction of the switching function described in the previous section and choice of the thresholds s_B^{min} and s_B^{max} in particular can be expected not to be adequate for other systems for the following reasons:

- The criterion $0.376 \leq s_B \leq 1.426$ applies at hydrogen nucleus where addition of $v_t^{limit}[\rho_B]$ is not expected to be needed. We recall here that it is the danger of a collapse of electron density on a doubly occupied hydrogenic 1s orbital provides the physical motivation for introducing the $v_t^{limit}[\rho_B]$. Moreover, numerical studies on molecular electron densities indicate clearly that enforcing the local behavior of the density of the kinetic energy near nucleus corresponding to the von Weizsäcker expression leads indeed to significant improvements of the approximant to $T_s[\rho]$ for all nuclei except that of hydrogen [104].

- For heavier atoms, electron density at the nucleus comprises contributions from other orbitals than the hydrogenic 1s. This can lead to the possibility that $v_t^{limit}[\rho_B]$ is negative although $s_B \leq 1.426$.

- The criterion $0.376 \leq s_B \leq 1.426$ might also be satisfied near stationary points of the electron density such as bond midpoints. It is very unlikely that the condition $\int \rho_B d\vec{r} = 2$ can be relevant to any volume element centered on a stationary points. Therefore addition of $v_t^{limit}[\rho_B]$ lacks formal justification there. Although the condition

> $0.376 \leq s_B$ assures that $v_t^{limit}[\rho_B]$ is not added at the stationary point, where $|\nabla\rho_B| = 0$, or in the close proximity to it, the numerical value of this threshold requires verification in real systems. It should be added at this point also that, if linear combination of atomic orbitals is used to construct embedded orbitals, the quality of description of the density at the nuclear cusp depends on the used basis set. A weaker criterion should be used in practice to assure that the $v_t^{limit}[\rho_B]$ is indeed added in the vicinity of a nucleus.

We start with the choice made for s_B^{max}. Numerical analyses in the systems discussed in the next section show that $v_t^{limit}[\rho_B]$ is negative locally even if s_B is smaller than 1.426 for heavier nuclei. This suggest that this threshold should be reduced. The smaller is the value for this threshold the less probably is inclusion of negative $v_t^{limit}[\rho_B]$, which is desired from the point of view of universality of this threshold, but it comes at the expense of loosing a part of the desired effect by reducing the range at which addition of $v_t^{limit}[\rho_B]$ applies. We use the model system considered in Appendix B to estimate the effect associated with the reduction of this range. The underlying assumption leading to the value of 1.426 value is rigorously true in the model system. The desired effect of reducing the charge distribution on top of the nucleus is achieved mainly by adding $v_t^{limit}[\rho_B](\vec{r})$ very close to the nucleus i.e. where $s_B < 0.6$. Increasing further the value of the s_B threshold leads to smaller effect. Moreover, adding locally $v_t^{limit}[\rho_B]$ to the potential near the nucleus leads to negligible effect on orbital energies in the system considered in Appendix B (it reaches a peak of about 10^{-4} eV at $s_B = 1.426$). These results for the model system indicate that any choice for $0.6 < s_B^{max} < 1.426$ is acceptable. In real systems discussed in the next section, lowering the threshold from 1.426 to to 0.9 assures that $v_t^{limit}[\rho_B]$ is added only if it is positive.

The fine-tuning of s_B^{min} follows other considerations. We note that, in the model system considered in Appendix B, s_B^{min} can be reduced even to zero without affecting the results because lower values of s_B than 0.376 do not occur near the nucleus. To make sure that no nuclei is overlooked even if the chosen atomic basis set in practical calculations is such that the exact relation $s_B = 0.376$ for $r \longrightarrow 0$ cannot be rigorously satisfied, s_B^{min} is reduced from 0.376 to 0.3. This change leads to negligible numerical effects if ρ_B corresponds to atomic electron densities. For molecular ρ_B, retaining the criterion based on the s_B^{min} is necessary to avoid unjustified additions of

$v_t^{limit}[\rho_B]$ near stationary points.

The criteria based only on s_B^{min} and s_B^{max} are not sufficient if the environment comprises hydrogen atoms because they are satisfied also at hydrogen nucleus. To avoid adding the $v_t^{limit}[\rho_B]$ near hydrogens, the proposed switching function includes additionally the criterion based on smallness of ρ_B. It is required that ρ_B is larger than the square of the 1s wave function of the hydrogen atom (Z=1) at $r = 0$ which equals to $1/\pi = 0.318$. Concerning ρ_B^{min}, increasing the idealized value of 0.318 to even 1 does not affect the results for hydrogen-free systems because the density on top of any nucleus, which is heavier than hydrogen, is at least one order of magnitude larger. Increasing the value ρ_B^{min} is desired for the same reasons as the ones motivating the decrease of s_B^{min}. The value of $\rho_B^{min} = 0.7$ was arbitrary chosen for practical calculations.

The final form of the "fine-tuned" switching function of more general applicability and used in the subsequent section for studying the importance of imposing the exact relation given in Eq. 4.8 in real systems takes the following form:

$$\begin{aligned}\tilde{f} &= \Big(\exp(\lambda(-s_B + s_B^{min})) + 1\Big)^{-1} \times \Big(1 - (\exp(\lambda(-s_B + s_B^{max})) + 1)^{-1}\Big) \\ &\times \Big(\exp(\lambda(-\rho_B + \rho_B^{min})) + 1\Big)^{-1} \qquad (4.15)\end{aligned}$$

where $s_B^{min} = 0.3$, $s_B^{max} = 0.9$, $\rho_B^{min} = 0.7$.

The final form of the approximant to $v_t[\rho_A, \rho_B]$, which will be denoted as $\tilde{v}_t^{NDSD}[\rho_A, \rho_B]$, where NDSD stands for Non-Decomposable approximant using first- and Second Derivatives of ρ, reads:

$$\tilde{v}_t^{NDSD}[\rho_A, \rho_B] = \tilde{v}_t^{TF}[\rho_A, \rho_B] + \tilde{f}(\rho_B, \nabla\rho_B) \cdot v_t^{limit}[\rho_B] \qquad (4.16)$$

where $\tilde{f}(\rho_B, \nabla\rho_B)$ is given in Eq. 4.15 The notion of non-decomposability is brought up here because the the second term in Eq. 4.11 does not have the form of a difference between two functional derivatives of some common explicit density functional $\tilde{T}_s[\rho]$.

The analytic expression for $\tilde{T}_s^{nad(NDSD)}[\rho_A, \rho_B]$, which yields $\tilde{v}_t^{NDSD}[\rho_A, \rho_B]$ after functional differentiation with respect to ρ_A can be easily constructed because its decomposable component is given in Eq. 4.4 whereas the non-decomposable $v_t^{limit}[\rho_B]$ component is ρ_A independent. It reads:

$$\begin{aligned}\tilde{T}_s^{nad(NDSD)}[\rho_A,\rho_B] &= C_{TF}\int\left((\rho_A+\rho_B)^{5/3}-\rho_A^{5/3}-\rho_B^{5/3}\right)d\vec{r} \\ &+ \int\tilde{f}(\rho_B,\nabla\rho_B)\cdot\rho_A(\vec{r})\tilde{v}_t^{limit}[\rho_B](\vec{r})d\vec{r}+C[\rho_B]\end{aligned} \tag{4.17}$$

where $C[\rho_B]$ is ρ_A-independent. To assure the proper dissociation limit $C[\rho_B]$ must vanish.

Finally, we underline that the constructed approximant to $v_t[\rho_A,\rho_B]$ comprises always the zero-th order term which does not depend on density gradients explicitly. We recall here that this component is finite at the nucleus and we have chosen to retain it. A more refined treatment of this gradient-free term, which would disappear at the considered limit if the approximant were exact is possible. Works on such aproximants are in progress in our lab. The present work, however, concerns the importance of the missing gradient-dependent terms. Switching off the zero-order term would require introduction of another switching function which would make drawing conclusions concerning the main issue of this work less straightforward.

It is worthwhile, to recall here that the issue of the amount of the von Weizsäcker term which should be added to the Thomas-Fermi expression is not new in the context of approximants to $T_s[\rho]$[12]. It was also investigated by Yang[105] in view of approximating density of the kinetic energy or more recently by Kluner et al.[38] who proposed switching off the von Weizsäcker term completely at large density gradients. Despite apparent similarity, these developments are not directly related to the present analysis of approximants to $v_t[\rho_A,\rho_B]$ which are non-decomposable i.e. their analytic forms do not allow for the construction of a "parent" approximant to $T_s[\rho]$.

4.5 Numerical validations

Procedure to analyze $\tilde{v}_t$-generated errors

In practical applications of Eq. 2.1, the results depend on ρ_B as well as on the approximant to $v_t[\rho_A,\rho_B](\vec{r})$. As far as assessment of the quality of the used approximant to $v_t[\rho_A,\rho_B](\vec{r})$ is concerned, a general procedure was proposed in one of our earlier works [13, 71]. Its principal element is the comparison between numerical values of the calculated property (energy components, dipole moments, total electron density, etc.) obtained from two fully variational formal frameworks: that of Cortona [28] and that of Kohn and Sham [2]. Results obtained from both frameworks are not exact but the difference between them can be attributed only to the approximant to $v_t[\rho_A, \rho_B](\vec{r})$ if all technical parameters (approximant to the exchange-correlation functional, basis sets for expanding orbitals, algorithms to calculate used matrix elements) are the same. We point out here, however, that direct comparisons between the total electron densities derived from Kohn-Sham- and Cortona's calculations are cumbersome because these quantities are local. In practice, it is more convenient to use global quantities (norm of the difference between these densities, or selected observables) in such analyses (see the next section).

To obtain the pair of electron densities ρ_A and ρ_B which minimizes the total energy in Cortona's type of calculations, the self-consistent super cycle of embedding calculations (*freeze-and-thaw* cycle) [31] is performed. At each iteration, one-electron equations for embedded orbitals (Eq. 2.1) are solved. In the subsequent iteration, ρ_A and ρ_B exchange their role. The *freeze-and-thaw* iterations continue until self-consistency. In the end, a pair of electron densities (ρ_A^0 and ρ_B^0) and the corresponding two sets of embedded orbitals are obtained. Obviously, the notion of *embedded system* and its *environment* becomes meaningless because both subsystems are treated on the equal footing. *Freeze-and-thaw* calculations are conducted in practice for small model systems to validate the used ρ_B in large scale multi-level numerical simulations [86] or, as it is made in the present work, to asses the used approximant to $v_t[\rho_A, \rho_B](\vec{r})$. All numerical results obtained with the $\tilde{v}_t^{NDSD}[\rho_A, \rho_B](\vec{r})$ approximant with the smoothing parameter $\lambda = 500$. Such results can be practically considered as equivalent to the use of stepfunction Θ (see Appendix C). Using a smaller value, such as $\lambda = 50$ recommendable for practical calculations, affects the discussed shifts in dipole moments or orbital-energies by less than 1%.

The effect of imposing the limit for $v_t[\rho_A, \rho_B]$ at $\rho_A \longrightarrow 0$ and $\int \rho_B d\vec{r} = 2$ on the complexation induced dipole moments.

$\tilde{v}_t$-generated errors in complexation induced dipole moments can be expected to be strongly affected by the local behavior of the used $\tilde{v}_t[\rho_A, \rho_B](\vec{r})$ near nuclei. Lack of sufficient repulsion near the nucleus might lead to an artificial transfer of electron density between subsystems reflected in the numerical values of the dipole moment. Therefore, this quantity was chosen for the analysis of the $\tilde{v}_t$-generated errors in a representative set of intermolecular complexes including charged, polar and non-polar ones at their equilibrium geometries. The key details of the numerical implementation of the relevant equations are as follows *i)* LDA approximation for exchange-correlation functional: exchange [16] and Vosko *al.* parametrization [17] of the Ceperley-Alder data [18] *ii)* TZP atomic basis sets of the STO type [106], *iii)* supermolecular expansion of electron density of each subsystem (KSCED(s) label in Ref. [73], *iv)* equilibrium geometries obtained from Kohn-Sham LDA calculations with TZP basis sets (available from the authors upon request), *v)* numercial implementation of the $\tilde{v}_t^{NDSD}$ in ADF code ([106]) version ADF2007 downloaded from svn repository on 2007-05-23.

Tables 4.1 and 4.2 collect the complexation induced dipole moments in neutral or charged complexes, respectively. First of all, imposing the exact relation given in Eq. 4.8 decreases the $\tilde{v}_t$-generated errors in each of the considered cases. The effect on the errors is negligible if subsystems are neutral. This indicates that the origin of the errors lies not in the violation the condition considered in this work. For systems comprising charged components, the effect of imposing the exact relation given in Eq. 4.8 is evident. The errors are invariably reduced. The reduction of the relative errors depends on the system from such a case as Li^+-F^- (from 9.8% to 8%) to a strong reduction in such systems as Na^+-Br^- and (from 1.4% to 0.8%) and Na^+-H_2O (from 0.23% to 0.05%) for instance. Similarly as in the group of complexes formed by neutral molecules, the origin for the remaining errors lies somewhere else. The above numerical examples lead to the following principal conclusions:

- Violation of the exact relation given in Eq. 4.8 contributes to the overall error in the calculated quantities but this contribution varies from one system to another. It is rather negligible for complexes formed by neutral components. It is numerically significant for complexes comprising charged components.

- Our simple strategy to impose the exact relation given in Eq. 4.8 in the vicinity of nuclear cusps leads invariably to reduction of errors. Therefore it can be used generally as correction to any approximant violating the exact relation.

- The construction of the approximant obeying the exact relation given

Table 4.1: Total dipole moments (μ in Debye) obtained from the *freeze-and-thaw* calculations using two approximations $v_t[\rho_A, \rho_B](\vec{r})$. The target Kohn-Sham results are given for comparison. Only dipole moment components along the principal axis connecting the subsystems are given. Relative percentage errors of the dipole moments derived from subsystem based calculations ($\frac{\Delta\mu^{subsystem} - \Delta\mu^{Kohn-Sham}}{\Delta\mu^{Kohn-Sham}} 100\%$) are given in parentheses.

		freeze-and-thaw		Kohn-Sham
		Eq. 4.5	Eq. 4.16	
A	B	μ		
Li^+	F^-	5.429 (9.8)	5.540 (8.0)	6.020
Li^+	Cl^-	5.768 (15.5)	5.930 (13.1)	6.828
Li^+	Br^-	5.706 (18.9)	5.931 (15.7)	7.033
Na^+	F^-	7.576 (1.6)	7.606 (1.2)	7.697
Na^+	Cl^-	8.509 (1.2)	8.561 (0.5)	8.608
Na^+	Br^-	8.637 (1.4)	8.686 (0.8)	8.758
Be^{2+}	O^{2-}	4.225 (31.0)	4.341 (29.1)	6.120
Mg^{2+}	O^{2-}	6.741 (3.5)	6.843 (2.1)	6.988
H_2O	H_2O	2.669 (3.9)	2.669 (3.9)	2.779
HF	HF	3.000 (2.8)	3.000 (2.8)	3.090
He	CO_2	0.014 (-8.6)	0.014 (-8.4)	0.013
Ne	CO_2	0.027 (-7.6)	0.027 (-7.6)	0.025

Table 4.2: Total dipole moment (μ in Debye) obtained from the *freeze-and-thaw* calculations using two approximations $v_t[\rho_A, \rho_B](\vec{r})$. The target Kohn-Sham results are given for comparison. Only dipole moment components along the principal axis connecting the subsystems are given. Relative percentage errors of the dipole moments derived from subsystem based calculations ($\frac{\Delta\mu^{subsystem} - \Delta\mu^{Kohn-Sham}}{\Delta\mu^{Kohn-Sham}} 100\%$) are given in parentheses.

		freeze-and-thaw		Kohn-Sham
		Eq. 4.5	Eq. 4.16	
A	*B*	μ		
Li^+	H_2O	5.298	5.351	5.513
		(3.91)	(2.94)	
Li^+	F_2	7.783	7.804	7.807
		(0.31)	(0.04)	
Li^+	CO_2	9.387	9.401	9.396
		(0.10)	(-0.05)	
Na^+	H_2O	6.512	6.524	6.527
		(0.23)	(0.05)	
H_3O^+	Ar	2.681	2.688	2.707
		(0.96)	(0.70)	
NH_4^+	Ar	1.661	1.661	1.816
		(8.52)	(8.52)	
Be^2+	He	12.475	12.565	12.986
		(3.94)	(3.24)	
Be^2+	H_2O	12.066	12.298	13.569
		(11.08)	(9.37)	
Mg^2+	He	17.302	17.316	17.338
		(0.20)	(0.13)	
Mg^2+	H_2O	19.441	19.531	19.751
		(1.57)	(1.11)	

[a] for charged systems the dipole moment is calculated for the cation at the origin.

in Eq. 4.8 and the used switching criteria in particular, correspond to a real case where a distant nucleus is surrounded by a frozen-density shell comprising two electrons (He, Li^+, Be^{2+}, etc.). The fact that the errors are reduced also for systems, where these idealized conditions do not apply, indicates that the exact relation given in Eq. 4.8 is important in practice and should be taken into account in construction of approximants to $v_t[\rho_A, \rho_B](\vec{r})$.

- Reflecting the exact relation given in Eq. 4.8 in $\tilde{v}_t[\rho_A, \rho_B](\vec{r})$ does not eliminate $\tilde{v}_t$-generated errors. For some systems, it even does not lead to noticeable improvements. Therefore, their origin must be looked for somewhere else.

The effect of imposing the limit for $v_t[\rho_A, \rho_B]$ at $\rho_A \longrightarrow 0$ and $\int \rho_B d\vec{r} = 2$ on orbital energies

In this section, we analyze the complexation induced shifts of orbital energies derived using the approximant $\tilde{v}_t^{NDSD}[\rho_A, \rho_B](\vec{r})$ considered in the previous section. Opposite to the dipole moments discussed previously, direct comparisons between the calculated shifts and the corresponding reference data are less straightforward. However, we investigate the numerical effect associated with imposing the exact relation given in Eq. 4.8 in view of the numerical practice which indicates that shifting the levels of unoccupied orbitals localized in the environment would be desirable. Conventional decomposable approximants to $v_t[\rho_A, \rho_B](\vec{r})$ lead to artificially low levels of unoccupied orbitals in the environment which might cause unphysical effects such as charge-transfer between subsystems [103] or erroneous other observables [95].

The numerical results for a model system considered in Appendix B show that inclusion of the $v_t^{limit}[\rho_B]$ term everywhere where $0.6 \leq s_B \leq 0.9$ into the effective embedding potential leads to a positive shift of the energy level of the unoccupied orbital. In real intermolecular systems, the conditions considered in Appendix B ($\rho_A \longrightarrow 0$ and $\rho_B = 2|1s|^2$) do not apply rigorously for the following reasons. If the subsystems are in finite separation and the atom-centered basis set used to construct embedded orbitals comprises atoms in both subsystems (supermolecular expansion labeled as KSCED(s) in Ref. [73]), ρ_A can be significant at a nucleus associated with subsystem B. As a consequence, the condition $\rho_A \longrightarrow 0$ may not apply. Moreover, if the atom in the environment comprises also occupied $2s$ shell the condition

Table 4.3: The effect of adding non-decomposable contribution to the embedding potential on the energy on the lowest unoccupied orbital localized in the environment (LUEO). For each complex, the used electron density of the environment (ρ_B) is the ground-state Kohn-Sham (LDA) electron density of the isolated subsystem B.

subsystem B	subsystem A	ϵ_{LUEO} $\tilde{v}_s = \tilde{v}^{TF}$	ϵ_{LUEO} $\tilde{v}_s = \tilde{v}^{NDSD}$	$\Delta\epsilon_{LUEO}$
Li^+	F^-	-2.249	-2.162	0.087
Li^+	Cl^-	-2.669	-2.570	0.099
Li^+	Br^-	-2.741	-2.638	0.103
Na^+	F^-	-1.908	-1.799	0.108
Na^+	Cl^-	-2.202	-2.172	0.030
Na^+	Br^-	-2.253	-2.221	0.032
Be^{2+}	O^{2-}	-4.922	-4.919	0.003
Mg^{2+}	O^{2-}	-4.634	-4.608	0.026

$\int \rho_B d\vec{r} = 2$ does not apply as well. The aim of the numerical analysis of this section is the verification whether the desired effect of shifting unoccupied orbital levels occurs in such cases.

Tables 4.3 and 4.4 collect the values of energy levels corresponding to the lowest lying orbital localized mainly in the environment for the previously considered complexes. The calculations are made not in the end of the *freeze-and-thaw* cycle but for the same ρ_B obtained from Kohn-Sham calculations for the isolated subsystem B. Including the non-decomposable ($v_t^{limit}[\rho_B]$) in the approximant to $v_t[\rho_A, \rho_B](\vec{r})$ leads to the shifts of the energy levels of unoccupied orbitals of the magnitude which significantly larger than that in the model system. It is a very desired effect of the new approximant.

In the Li^+-H_2O case discussed in Ref. [103, 95], the energy of the lowest unoccupied embedded orbital localized on Li^+ crosses with that of the highest occupied embedded orbital localized on H_2O at the intermolecular distance of 13 Å if Eq. 4.5 is used for $v_t[\rho_A, \rho_B](\vec{r})$. At larger separations, the self-consistent procedure to solve Eqs. 2.1 does not converge due to localization of the highest occupied embedded orbital which jumps between subsystems in subsequent iterations. Imposing the exact relation given in Eq. 4.8 leads to shifts of the energy of the unoccupied embedded orbital localized at Li^+ and does not affect the levels of occupied orbitals at all (see Figure 4.2, and

Table 4.4: The effect of adding non-decomposable contribution to the embedding potential on the energy on the lowest unoccupied orbital localized in the environment (LUEO). For each complex, the used electron density of the environment (ρ_B) is the ground-state Kohn-Sham (LDA) electron density of the isolated subsystem B.

subsystem B	subsystem A	ϵ_{LUEO} $\tilde{v}_s = \tilde{v}^{TF}$	ϵ_{LUEO} $\tilde{v}_s = \tilde{v}^{NDSD}$	$\Delta\epsilon_{LUEO}$
Li^+	H_2O	-7.533	-7.335	0.198
Li^+	F_2	-8.501	-8.228	0.273
Li^+	CO_2	-9.069	-8.838	0.231
Na^+	H_2O	-6.295	-6.227	0.068
Be^2+	He	-26.681	-25.915	0.766
Mg^2+	He	-18.376	-17.990	0.386
Be^2+	H_2O	-21.737	-21.253	0.484
Mg^2+	H_2O	-16.570	-16.253	0.317

Tables 4.5 and 4.6). Note that the levels still cross but it occurs at larger separations (about 16 Å) than in the $\tilde{v}_t^{TF}[\rho_A, \rho_B]$ case (about 13 Å). Moreover, the shifts increase with intermolecular separation. Crossing the levels in this case, is a qualitatively wrong because it would result in wrong redistribution of charges upon dissociation. The origin of crossing might be attributed to several factors, *i*) the used approximant to the exchange-correlation potential (wrong relative positions of frontier orbitals in the dissociated fragments), *ii*) errors in $\tilde{v}_t^{NDSD}[\rho_A, \rho_B](\vec{r})$ (not sufficiently repulsive), *iii*) the particular choice for ρ_B (note the strong dependence of $v_t^{limit}[\rho_B]$ on ζ for hydrogenic ρ_B discussed earlier). The first among these factors is not important because the relative order of frontier orbitals in isolated H_2O and Li^+ is correct. Therefore, the two remaining factors lie at the origin of the observed crossing.

We turn back to the approximant $\tilde{v}_t^{NDSD}[\rho_A, \rho_B](\vec{r})$. The observed significant shift of the level-crossing point provides thus the clear evidence of a great relevance of the exact relation given in Eq. 4.8 in practice. In the aforementioned work by Jacob *et al.* [95], the levels do not cross at all owing to a position dependent term assuring that the whole embedding potential disappears at dissociation. In practice, if the dissociation limit is known in advance and the $\tilde{v}_t^{NDSD}[\rho_A, \rho_B](\vec{r})$ together with some approximant to the exchange-correlation potential describe it incorrectly, the empirical correc-

Table 4.5: The effect of adding non-decomposable contribution to the embedding potential on the energy on the highest occupied embedded orbital (HOEO). For each complex, the used electron density of the environment (ρ_B) is the ground-state Kohn-Sham (LDA) electron density of the isolated subsystem B.

subsystem B	subsystem A	ϵ_{HOEO} $\tilde{v}_s = \tilde{v}^{TF}$	ϵ_{HOEO} $\tilde{v}_s = \tilde{v}^{NDSD}$	$\Delta\epsilon_{HOEO}$
Li^+	F^-	-6.716	-6.637	0.079
Li^+	Cl^-	-6.399	-6.340	0.059
Li^+	Br^-	-6.122	-6.069	0.053
Na^+	F^-	-5.223	-5.202	0.21
Na^+	Cl^-	-5.397	-5.383	0.014
Na^+	Br^-	-5.277	-5.264	0.013
Be^{2+}	O^{2-}	-6.724	-6.697	0.027
Mg^{2+}	O^{2-}	-5.135	-5.103	0.032
H_2O^a	H_2O	-6.799	-6.799	0.000
HF^a	HF	-9.450	-9.450	0.000
H_2O^b	H_2O	-8.010	-8.011	-0.001
HF^b	HF	-10.896	-10.896	0.000

[a] acceptor of hydrogen bond

[b] donor of hydrogen bond

Table 4.6: The effect of adding non-decomposable contribution to the embedding potential on the energy on the highest occupied embedded orbital (HOEO). For each complex, the used electron density of the environment (ρ_B) is the ground-state Kohn-Sham (LDA) electron density of the isolated subsystem B.

subsystem B	subsystem A	ϵ_{HOEO} $\tilde{v}_s = \tilde{v}^{TF}$	ϵ_{HOEO} $\tilde{v}_s = \tilde{v}^{NDSD}$	$\Delta\epsilon_{HOEO}$
Li^+	H_2O	-14.530	-14.493	0.037
Li^+	F_2	-16.507	-16.493	0.014
Li^+	CO_2	-14.749	-14.743	0.006
Na^+	H_2O	-13.284	-13.276	0.008
Be^2+	He	-37.558	-37.339	0.219
Mg^2+	He	-31.410	-31.379	0.031
Be^2+	H_2O	-25.208	-25.029	0.179
Mg^2+	H_2O	-21.178	-21.117	0.061

tion introduced in Ref. [95] can be always combined with $\tilde{v}_t^{NDSD}[\rho_A,\rho_B](\vec{r})$ as a pragmatic solution.

4.6 Alternative non-decomposable approximants to $v_t[\rho_A,\rho_B]$

In principle, any decomposable approximant can be used as the first term of Eq. 4.16 instead of the term derived from Thomas-Fermi functional considered in the previous section. In particular, this term can be derived from second-order gradient expansion or generalized gradient expansion type of approximants to $T_s[\rho]$ overviewed in the introductory sections. In this section, we analyze the effect of imposing the relation given in Eq. 4.8 on gradient-dependent decomposable approximants to $v_t[\rho_A,\rho_B](\vec{r})$. It is worthwhile, however, to point out that the gradient-free decomposable component of $\tilde{v}_t^{NDSD}[\rho_A,\rho_B]$ cannot recover the exact relation of Eq. 4.8 at all whereas the gradient-dependent decomposable approximants do it only partially. Therefore, a proper care needs to be taken to avoid double-counting of gradient-dependent terms in the approximant to $v_t[\rho_A,\rho_B]$ near nuclei. For instance, $\tilde{v}_t^{GEA2}[\rho_A,\rho_B]$ comprises already 1/9 of $v_t^{limit}[\rho_B]$. Therefore, the plateau of

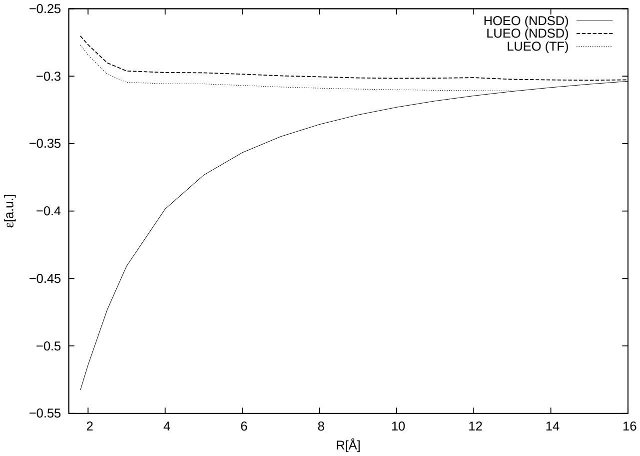


Figure 4.2: The effect of imposing the exact relation given in Eq. 4.8 on the orbital levels for the Li^+-H_2O complex at various intermolecular separations. Only the energy of the highest embedded occupied orbital (HOEO) and lowest unoccupied embedded orbital (LUEO) are shown. The TF and NDSD energies of HOEO are not distinguishable on the figure and only one set of data is shown. The NDSD curve is obtained using $\lambda = 50$ (see the discussion in Appendix C).

the switching function reaches not 1 but 8/9 in such a case. For GGA-type of approximants to $v_t[\rho_A, \rho_B]$, the value at the plateau must reflect the analytic dependence of the approximant on reduced density gradient s at small gradients.

In this section, we construct two alternative approximants of the Eq. 4.16 form which take into account exact relation given in Eq. 4.8. One in which $\tilde{v}_t^{TF}[\rho_A, \rho_B]$ in Eq. 4.16 is replaced by either $\tilde{v}_t^{GEA2}[\rho_A, \rho_B]$ [4] or $\tilde{v}_t^{GGA97}[\rho_A, \rho_B]$ [72]. In both cases, enforcing the exact relation given in Eq. 4.8 on the constructed approximant brings improvements (see Table 4.7). This results indicates that the exact relation given in Eq. 4.8 is indeed important in practice as imposing it improves the decomposable approximants which violate it.

Compared to the results obtained using $\tilde{v}_s^{NDSD}[\rho_A, \rho_B](\vec{r})$ discussed earlier in this work, the two alternative non-decomposable approximants to $v_s[\rho_A, \rho_B](\vec{r})$ are not better.

We recall here the main reasons for singling out $\tilde{v}_t^{GGA97}[\rho_A, \rho_B])$ among other decomposable ones which depend explicitly on densities ρ_A and ρ_B as well as their first- and second derivatives.

Table 4.7: Total dipole moment (μ in Debye) obtained from the *freeze-and-thaw* calculations using different approximations $v_t[\rho_A,\rho_B](\vec{r})$. The target Kohn-Sham results are given in Tables 4.1 and 4.2. Only dipole moment components along the principal axis connecting the subsystems are given. Relative percentage errors of the dipole moments derived from subsystem based calculations ($\frac{\Delta\mu^{subsystem}-\Delta\mu^{Kohn-Sham}}{\Delta\mu^{Kohn-Sham}}100\%$) are given in parentheses.

system		v_t^{GEA2}	$v_t^{GEA2}+f\cdot v_t^{limit}$	v_t^{GGA97}	$v_t^{GGA97}+f\cdot v_t^{limit}$
Li^+	Cl^-	5.114	5.256	5.674	5.807
		(25.1)	(23.0)	(16.9)	(14.9)
Li^+	H_2O^a	5.024	5.078	5.256	5.293
		(8.9)	(7.9)	(4.8)	(4.0)
Na^+	Cl^-	7.612	7.654	8.321	8.360
		(11.6)	(11.1)	(3.3)	(2.9)
Na^+	H_2O^a	6.286	6.300	6.446	6.457
		(3.7)	(3.5)	(1.24)	(1.1)
Be^{2+}	O^{2-}	3.807	3.888	4.125	4.212
		(37.8)	(36.5)	(32.6)	(31.8)
HF	HF	2.991	2.991	2.992	2.993
		(3.2)	(3.2)	(3.2)	(3.1)

[a] for charged systems the dipole moment is calculated for the cation at the origin.

- $\tilde{v}_t^{GEA2}[\rho_A,\rho_B]$ obtained from the second-order gradient-expansion of $T_s[\rho]$ leads typically to worse results than that obtained from zeroth order [13, 73] indicating that the contribution to $v_t[\rho_A,\rho_B]$ due to the second term in Eq. 4.6 is erroneous. The deterioration of the results is pronounced the most at very small overlaps between ρ_A and ρ_B.

- $\tilde{v}_t^{GGA97}[\rho_A,\rho_B]$ was introduced as a pragmatic solution replacing $\tilde{v}_t^{GEA2}[\rho_A,\rho_B]$. Due to its analytic form, the gradient-dependent contribution disappears at small overlaps between ρ_A and ρ_B.

- The functional $\tilde{T}^{LC94}[\rho]$ generating the decomposable approximant $\tilde{v}_s^{GGA97}[\rho_A,\rho_B]$ is known to be also a very good approximant to $T_s[\rho]$.

In $\tilde{v}_t^{NDSD}[\rho_A,\rho_B]$, the problematic second-order term lying at the origin of flaws of $\tilde{v}_t^{GEA2}[\rho_A,\rho_B]$ is either present, wherever the exact relation for $v_t[\rho_A,\rho_B](\vec{r})$ given in Eq. 4.8 is expected to be relevant (i.e. in the vicinity of nuclei) or it is absent. Numerical results collected in Tables 4.1, 4.2, and 4.7 support the above formal reasons to consider $\tilde{v}_t^{NDSD}[\rho_A,\rho_B]$ as the successor of $\tilde{v}_t^{GGA97}[\rho_A,\rho_B]$.

4.7 Conclusions

Approximants to $T_s[\rho]$ are object of continuous efforts due to their possible use in the orbital-free types of models dating back to the works of Thomas and Fermi [9, 10]. For a representative current developments see the review by Wang and Carter [61]. For any approximant $\tilde{T}_s[\rho]$, it is straightforward to construct the associated *decomposable* approximant to $v_t[\rho_A,\rho_B](\vec{r})$. In methods using orbtial-free effective embedding potential [4], the approximants to the total kinetic energy are not directly used. The key quantity is the approximant to $v_t[\rho_A,\rho_B](\vec{r})$. Decomposable approximants constructed starting from gradient expansion approximantion violate the relation given in Eq. 4.8 which involves directly $v_t[\rho_A,\rho_B](\vec{r})$.

Numerical results presented in this work indicate clearly that taking into account the exact relation given in Eq. 4.8 leads to improved approximants. Errors in the complexation induced dipole moments are reduced by more than 50% in some cases. For neutral systems, reduction of error takes also place but its magnitude is typically negligible. The bulk of numerical results presented in this work concerns the effect of imposing the exact relation given in Eq. 4.8 on the simplest decomposable approximant to $v_t[\rho_A,\rho_B](\vec{r})$ obtained from the Thomas-Fermi approximant to $T_s[\rho]$. Violation of the exact relation given in Eq. 4.8 by decomposable approximants to $v_t[\rho_A,\rho_B](\vec{r})$ is numerically relevant not only in this case but also if other simple gradient-dependent approximants to $T_s[\rho]$ are used as indicated by supplementary numerical results given in this work.

This indicates that one of important sources of inaccuracies in the conventional (i.e. decomposable and gradient-expansion based) approximant to $v_t[\rho_A,\rho_B](\vec{r})$ was identified. The origin of the remaining contributions $\tilde{v}_t$-generated errors lies probably somewhere else.

As the result of the present work, a new approximant to ($\tilde{v}_t^{NDSD}[\rho_A,\rho_B](\vec{r})$ given in Eq. 4.16) is proposed as an alternative to the ones obtained previ-

ously using conventional decomposable strategy especially if it uses low-order terms in the gradient expansion of $T_s[\rho]$ [96]. Advantages of $\tilde{v}_t^{NDSD}[\rho_A, \rho_B](\vec{r})$ are the following:
i) **Formal:** We believe that a proper strategy to improve approximants to functionals in density functional theory should proceed by imposing the most relevant exact conditions and $\tilde{v}_t^{NDSD}[\rho_A, \rho_B](\vec{r})$ was constructed in this way.
ii) **Practical:** The numerical results reported in this work show indeed that imposing this condition improves the obtained electron density and that the improvement varies from negligible to significant depending on the system.
iii) **Numerical:** Evaluating $\tilde{v}_t^{NDSD}[\rho_A, \rho_B](\vec{r})$ involves the same quantities as evaluating its counterparts derived from gradient-expansion- (up to second order) and so called generalized gradient approximants to $T_s[\rho]$.

Concerning the area of applicability of $\tilde{v}_t^{NDSD}[\rho_A, \rho_B](\vec{r})$, it should be underlined that some arbitrary choices were made concerning the criteria for "detecting" the vicinity of a nucleus based only on electron density ρ_B and its derivatives in the proposed construction. The criteria to switch on the $v_t^{limit}[\rho_B]$ term were designed based on the analysis of model system of relevance for elements of the first-, second- and the third period. Finally, we underline that the applied non-decomposable constriction can be combined with any decomposable approximant to $\tilde{v}_t^{NDSD}[\rho_A, \rho_B](\vec{r})$.

The local potential $\tilde{v}_t^{NDSD}[\rho_A, \rho_B](\vec{r})$ introduced in this work is intrinsically non-decomposable. Although the analytic form of the functional $\tilde{T}_s^{nad(NDSD)}[\rho_A, \rho_B]$ and the potential $\tilde{v}_t^{NDSD}[\rho_A, \rho_B](\vec{r})$ are given in this work, the analytic form of neither $\tilde{T}_s^{NDSD}[\rho]$ nor $\frac{\delta \tilde{T}_s^{NDSD}[\rho]}{\delta \rho}$ is available. Therefore, the introduced here non-decomposable strategy can be seen as the first attempt to decouple the search for $\tilde{T}_s[\rho]$ and its functional derivative, which are needed in orbital-free calculations, from the search for an adequate approximant to the kinetic-kinetic-energy dependent component of the effective orbital-free embedding potential. Opposite to orbital-free strategy [9, 10], neither $\tilde{T}_s[\rho]$ nor its functional derivative are needed in methods applying orbital-free effective embedding potential given in Eq. 2.7.

Finally, imposing the exact relation for $v_t[\rho_A, \rho_B](\vec{r})$ given in Eq. 4.8 leads to shifts of the level of the unoccupied orbitals localized in the environment. Such shift is strongly desired in view of the earlier reports on possible practical inconveniences resulted from artificially low position of such levels when approximants to $v_t[\rho_A, \rho_B](\vec{r})$, which do not take into account the cusp condition, are used [103, 95].

Acknowledgment: This work was supported by Swiss National Science Foundation.

Appendix A

Orbital-free embedding potential for $\rho_A \longrightarrow 0$ and $\rho_B = 2|1s|^2$

For small $\delta\rho$ such that $\delta\rho \longrightarrow 0$,

$$\begin{aligned}
\delta T_s^{nad}[\rho_A, \rho_B] &= T_s^{nad}[\rho_A + \delta\rho, \rho_B] - T_s^{nad}[\rho_A, \rho_B] \\
&= T_s[\rho_A + \delta\rho + \rho_B] - T_s[\rho_A + \rho_B] - T_s[\rho_A + \delta\rho] + T_s[\rho_A] \\
&= \int \frac{\delta T_s[\rho]}{\delta\rho}(\vec{r})\bigg|_{\rho=\rho_A+\rho_B} \delta\rho(\vec{r})d\vec{r} - \int \frac{\delta T_s[\rho]}{\delta\rho}(\vec{r})\bigg|_{\rho=\rho_A} \delta\rho(\vec{r})d\vec{r} + O(\delta^2\rho)
\end{aligned} \tag{4.18}$$

If also ρ_A is small i.e. $\rho_A \longrightarrow 0$,

$$T_s^{nad}[\rho_A + \delta\rho, \rho_B] - T_s^{nad}[\rho_A, \rho_B] = \int \frac{\delta T_s[\rho]}{\delta\rho}(\vec{r})\bigg|_{\rho=\rho_B} \delta\rho(\vec{r})d\vec{r} + O(\delta^2\rho) \tag{4.19}$$

Therefore,

$$\frac{\delta T_s^{nad}[\rho, \rho_B]}{\delta\rho}(\vec{r})\bigg|_{\rho\longrightarrow 0} \approx \frac{\delta T_s[\rho]}{\delta\rho}(\vec{r})\bigg|_{\rho\longrightarrow\rho_B} \tag{4.20}$$

The above result that the kinetic-energy component of v_{emb}^{KSCED} is just the functional derivative of the $T_s[\rho]$ calculated for $\rho = \rho_B$ makes it possible to express it analytically for any ρ_B which comprises just two electrons. For one-electron and two-electron- spin-compensated systems the exact expression reads [11]:

$$T_s[\rho] = T_s^W[\rho] = \int \frac{1}{8}\frac{|\nabla\rho|^2}{\rho}d\vec{r} \;\text{ for }\; \int \rho d\vec{r} = 2 \tag{4.21}$$

Therefore,

$$\frac{\delta T_s^W[\rho]}{\delta\rho}\bigg|_{\rho=\rho_B} = \frac{1}{8}\frac{|\nabla\rho_B|^2}{\rho_B^2} - \frac{1}{4}\frac{\nabla^2\rho_B}{\rho_B} \;\text{ if }\; \int \rho_B d\vec{r} = 2 \tag{4.22}$$

Using Eq. 4.22 in Eq. 4.20 leads to the asymptotic form of the kinetic energy component of v_{emb}^{KSCED} in the case where ρ_A and ρ_B do not overlap significantly and ρ_B is a two-electron spin-less electron density reads:

$$\left.\frac{\delta T_s^{nad}[\rho,\rho_B]}{\delta\rho}\right|_{\rho=\rho_A\longrightarrow 0,\int\rho_B d\vec{r}=2} = \frac{1}{8}\frac{|\nabla\rho_B|^2}{\rho_B^2} - \frac{1}{4}\frac{\nabla^2\rho_B}{\rho_B} \tag{4.23}$$

ρ_B representing a doubly occupied hydrogenic $1s$ function ($1s = \sqrt{\zeta^3/\pi}\cdot\exp(-\zeta r)$) reads:

$$\rho_B^{1s}(\vec{r}) = \rho_B^{1s}(r) = 2\cdot\zeta^3/\pi\cdot\exp\left(-2\zeta r\right), \tag{4.24}$$

For ρ_B^{1s}, Eq. 4.23 leads to the following potential:

$$\left.\frac{\delta T_s^{nad}[\rho,\rho_B]}{\delta\rho}(\vec{r})\right|_{\rho=\rho_A\longrightarrow 0,\rho_B=\rho_B^{1s}} = \frac{\zeta}{r} - \frac{\zeta^2}{2} \tag{4.25}$$

The potential given in Eq. 4.25 is repulsive for $r < \frac{2}{\zeta}$. For hydrogenic densities ρ_B^{1s}, the reduced density gradient equals to a ζ-independent value of 1.426. Near the nuclear cusp, therefore, a pair of electrons on the $1s$ shell provides a local repulsive potential which compensates the Coulomb attraction due to the nuclear charge. Note that the effective nuclear charge ζ for the most tightly bound orbital ($1s$) in a multi-electron atom is smaller than the charge of the corresponding nucleus (Z) [101]. As a consequence, the compensation is perfect only for one-electron hydrogenic systems.

Appendix B

Dependence of the energies of bound states localized in the environment far from subsystem A on approximants to $v_t[\rho_A, \rho_B](r)$

We consider numerical solutions of the Schrödinger equation for one electron in the spherically symmetric potential which takes the general from given in Eqs. 2.6-2.7. The analyzed potential corresponds to $v_{eff}^{KS}[\rho_A; \vec{r}] = 0$, $v_{ext}^{B}(\vec{r}) = -\frac{Z}{r}$, and $v_{emb}^{KSCED}[\rho_A, \rho_B; \vec{r}]$ defined in Eq. 2.7 for $\rho_A \longrightarrow 0$ and $\rho_B = \rho_B^{1s}$. Such a case represents a local potential around a nucleus with the charge Z localized in the environment far from from the investigated subsystem A and ρ_B comprising entire contribution from doubly occupied $1s$ shell centered on this nucleus. Dirac's exchange energy expression [16] is used to derive the exchange-correlation component of $v_{emb}^{KSCED}[\rho_A, \rho_B; \vec{r}]$. The solutions of one-electron Schrödinger equation for such potential is obtained by radial quadrature described in Refs. [107, 108] implemented numerically using MATLAB2006 environment [109]. Finite difference approximation and the matrix representation with 127 radial points is used. The above model system is used to investigate the dependence of the lowest energy level on the used approximant to $v_t[\rho_A, \rho_B](r)$.

Such approximant to $\tilde{v}_t[\rho_A, \rho_B](r)$, which is finite at nuclear cusp, might lead to appearance of artificially stabilized bound states due to improper balance between the nuclear attraction, classical electron-electron repulsion, which are both descried exactly and improperly behaving $\tilde{v}_t[\rho_A, \rho_B](r)$. This imbalance does not cause qualitative problems in the Z=2 (helium atom) case. No bound states occur if $v_t[\rho_A, \rho_B](r)$ is approximated by:

$$\begin{aligned} \tilde{v}_t^{model0} &= \left.\frac{\delta T_s^{nad(TF)}[\rho, \rho_B]}{\delta \rho}\right|_{\rho=\rho_A \longrightarrow 0} \\ &= \frac{5}{3} C_{TF} \rho_B^{2/3} \end{aligned} \tag{4.26}$$

derived from the uniform-electron gas expression for v_t given in Eq. 4.5.

For $Z = 3$ (i.e. Li$^+$), however, $\tilde{v}_t^{model0}$ leads to a bound state of the energy -0.2096654 hartree. This value will be used as a reference for analysis of other approximants to $v_t[\rho_A, \rho_B](\vec{r})$.

The expression derived from the second-order gradient approximation to $\frac{\delta T_s^{nad}[\rho_A, \rho_B]}{\delta \rho_A}$ reads:

$$\begin{aligned}\tilde{v}_t^{model1} &= \left.\frac{\delta T_s^{nad(GEA2)}[\rho,\rho_B]}{\delta\rho}\right|_{\rho=\rho_A\longrightarrow 0} \\ &= \frac{5}{3}C_{TF}\rho_B^{2/3} - \frac{1}{72}\frac{\nabla^2\rho_B}{\rho_B} + \frac{1}{144}\frac{|\nabla\rho_B|^2}{\rho_B^2}\end{aligned} \tag{4.27}$$

leads to a further lowering of the energy to -0.611935 hartree.

In the following part, other potentials will be considered, for which a local non-decomposable contribution is added to that given in Eq. 4.26 near the nucleus.

$$\begin{aligned}\tilde{v}_t^{model2}(r) &= \frac{5}{3}C_{TF}\rho_B^{2/3} \\ &+ \begin{cases} \frac{1}{8}\frac{|\nabla\rho_B|^2}{\rho_B} - \frac{1}{4}\frac{\nabla^2\rho_B}{\rho_B^2} = \frac{\zeta}{r} - \frac{\zeta^2}{2} \text{ for } r \le r^T \\ \\ 0 \text{ for } r > r^T \end{cases}\end{aligned} \tag{4.28}$$

Figure 4.3 shows that, compared to the decomposable result, the addition of a non-decomposable component destabilizes the energy but only for such small values of r at which $s_B \le 1.6$. The maximal destabilization occurs at $s_B = 1.426$ in line with the change of sign of the added term. A similar picture emerges from the analysis of the electron density at $r = 0$. Comparisons of Figures 4.3 and 4.4 reveals that addition of the non-decomposable term to the effective potential affects the orbital energies and electron density in a different manner. Even very close to the nucleus this addition reduced electron density without affecting the orbital energy noticeably. From the point of view of choosing s_B^{max} determining the range of this additional potential it is worthwhile to notice that the limit $s = 1.426$ should not be exceeded. The energy level starts a rapid descend and the on-top density starts to rise again at larger values of s_B^T. As far as the lower limit for s_B is concerned, it should not be smaller than about 0.5 because points at which $s_B < 0.5$ influence significantly the charge density at the nucleus.

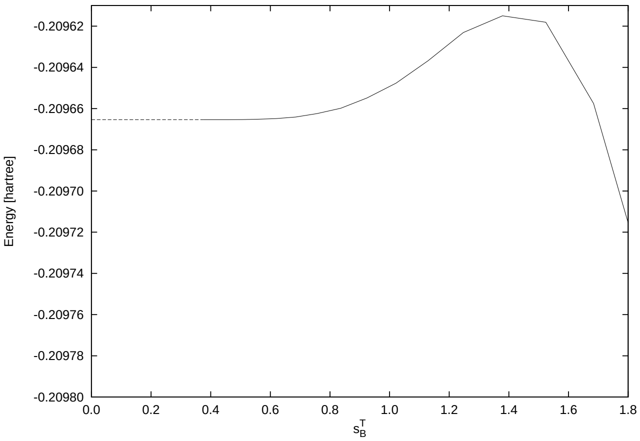


Figure 4.3: The ground-state energy level in the model system for different values of the threshold s_B^T corresponding to the range r^T of the $\frac{1}{8}\frac{|\nabla\rho_B|^2}{\rho_B} - \frac{1}{4}\frac{\nabla^2\rho_B}{\rho_B^2}$ term in Eq. 4.28. For $s_B^T < 0.376$, the corresponding r^T does not exists in the model system and the results obtained without this additional term are shown.

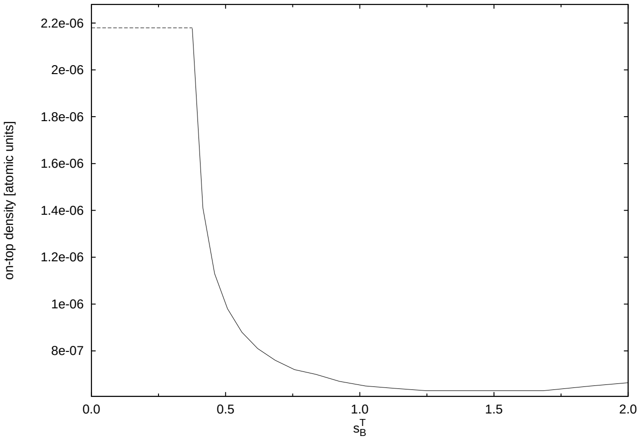


Figure 4.4: The electron density at $r = 0$ in the model system for different values of the threshold s_B^T corresponding to the range r^T of the $\frac{1}{8}\frac{|\nabla\rho_B|^2}{\rho_B} - \frac{1}{4}\frac{\nabla^2\rho_B}{\rho_B^2}$ term in Eq. 4.28. For $s_B^T < 0.376$, the corresponding r^T does not exists in the model system and the results obtained without this additional term are shown.

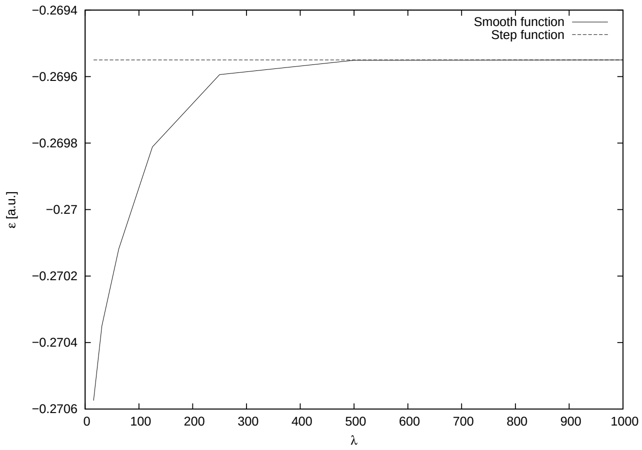


Figure 4.5: The energy of the lowest unoccupied orbital (LUEO) of Li^+-H_2O plotted as a function of the smoothing parameter λ in Eq. 4.29. The horizontal line represents the result obtained the sharp step function θ.

Appendix C

Smoothing the thresholds

In practical calculations, the use of the step-function Θ is not desirable. We consider a smooth switching of the Fermi-Dirac statistics form:

$$\tilde{f} \;=\; \Big(\exp(\lambda(-s_B + s_B^{min})) + 1\Big)^{-1} \tag{4.29}$$

where the parameter λ determines the smoothness of the switch. The value of $\lambda = 500$ in Eq. 4.29 leads to equivalent results to that obtained with the step function Θ (differences in dipole in the range of 10^{-6} Debye and orbital energies in the range of 10^{-10} hartree). See also Figure 4.5. In the present work the value of $\lambda = 500$ is used in all discussions concerning the importance of the $v_t^{limit}[\rho_B]$ term (induced dipole moments for instance).

It is worthwhile to point out here that the switching criterion is based on the electron density of the tightest bound electrons in the subsystem B. If the geometry of the system changes, the switching place moves together with the corresponding nucleus. Therefore, even a rather sharp switching can be expected to be acceptable in practice. In particular cases, however, smoother switching would be desirable. Figure 4.6 shows the energy of the lowest unoccupied embedded orbital localized on Li^+ for the Li^+-H_2O complex at various intermolecular separations for three values of λ. Indeed, reducing the value of λ from 500 to 50 smooths out some irregularities on the curve without changing its overall shape. Note that the variation of the orbital energy is one order of magnitude smaller than the effect of the presence of the $v_t^{limit}[\rho_B]$ in the NDSD approximant. We can conclude, therefore, that the numerical value of λ should not be considered physically meaningful if it exceeds 50. It should be considered as a technical parameter in this range. In practical calculations, the value of $\lambda = 50$ should be rather used.

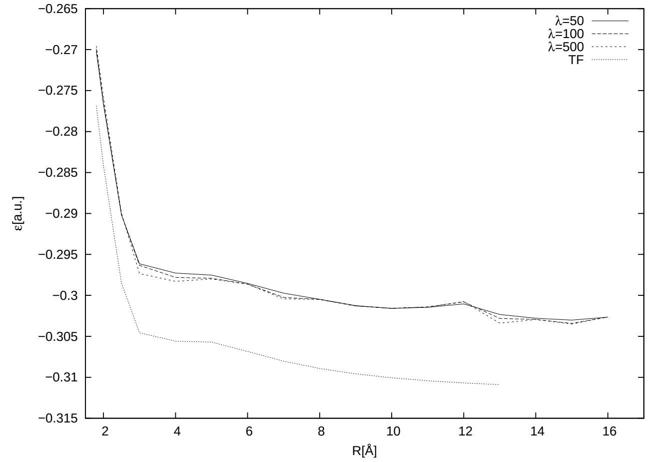


Figure 4.6: The energy of the lowest unoccupied embedded orbitals (LUEO) of Li^+-H_2O plotted as function distance for different values of λ.

Part III

Algorithms

Chapter 5

Linearized Orbital-Free Embedding Potential in Self-Consistent Calculations

Conventionally, solving one-electron equations for embedded orbitals (Eqs. 20-21 in [Wesołowski and Warshel, *J. Phys. Chem.* **97**, (1993) 8050]) proceeds by a self-consistent procedure in which the whole effective potential, including its embedding component, is updated in each iteration. We propose an alternative scheme (*splitSCF*), which uses the linearized embedding potential in the inner iterative loop and the outer-loop is used to account for its deviations from linearity. The convergence of the proposed scheme is investigated for a set of weakly bound intermolecular complexes representing typical interactions with the environment. The outer loop is shown to converge very fast. No more than 3-4 iterations are needed. Errors due to skipping the outer loop completely and using the electron density obtained in the absence of the environment in the linearized embedding potential are investigated in detail. It is shown that this computationally attractive simplification, used already in numerical simulations by others, is adequate not only for van der Waals and hydrogen-bonded complexes but even if the complex comprises charged components i.e. where strong electronic polarization takes place. In charge-transfer type of complexes, larger changes of electron of density upon complex formation occur and the above simplification is not recommended.

5.1 Introduction

Embedding strategies are commonly used to study large chemical systems and condensed matter as the alternative to methods taking into account all components of the investigated system at the same quantum mechanical level of description. The common element in embedding methods is restricting the use of explicit quantum mechanical level to a selected subsystem and taking into account the rest by some potential (*embedding potential*). Density functional theory [1, 2] provides the concepts and definitions such as a reference system of non-interacting electrons, the exchange-correlation functional ($E_{xc}[\rho]$), and kinetic energy of non-interacting systems [3] ($T_s[\rho]$) which allow one to express the embedding potential in a system-independent way [4]. To minimize the total energy of a whole system comprising two components: the one described at the orbital level, i.e. using the orbitals $\{\phi_i^A\}$ which yield the density $\rho_A = 2\sum_i^{N_A}\left|\phi_i^A\right|^2$, and the other one (*environment*) described only by some assumed electron density (ρ_B) by solving the one-electron equation given in Eq. 2.1.

The density functionals $T_s[\rho]$ and $E_{xc}[\rho]$ are not linear. Therefore, the left-hand sides of the following equations:

$$T_s^{nad}[\rho_A, \rho_B] = T_s[\rho_A + \rho_B] - T_s[\rho_A] - T_s[\rho_B] \tag{5.1}$$

$$E_{xc}^{nad}[\rho_A, \rho_B] = E_{xc}[\rho_A + \rho_B] - E_{xc}[\rho_A] - E_{xc}[\rho_B] \tag{5.2}$$

do not disappear and each of them defines a bi-functional depending on two electron densities. The functional derivatives of these bi-functionals calculated with respect to ρ_A provide the last two terms in the orbital-free embedding potential:

$$v_t[\rho_A, \rho_B](\vec{r}) = \left.\frac{\delta T_s^{nad}[\rho, \rho_B]}{\delta\rho}\right|_{\rho=\rho_A} = \left.\frac{\delta T_s[\rho]}{\delta\rho}\right|_{\rho=\rho_A+\rho_B} - \left.\frac{\delta T_s[\rho]}{\delta\rho}\right|_{\rho=\rho_A} \tag{5.3}$$

and

$$\left.\frac{\delta E_{xc}^{nad}\left[\rho, \rho_B\right]}{\delta\rho}\right|_{\rho=\rho_A} = \left.\frac{\delta E_{xc}\left[\rho\right]}{\delta\rho}\right|_{\rho=\rho_A+\rho_B} - \left.\frac{\delta E_{xc}\left[\rho\right]}{\delta\rho}\right|_{\rho=\rho_A} \tag{5.4}$$

Since the bi-functionals $T_s^{nad}[\rho_A, \rho_B]$ and $E_{xc}^{nad}[\rho_A, \rho_B]$ and their derivatives are defined implicitly and their universally applicable analytic forms are not known, practical applications of Eq. 2.1 in numerical simulations of condensed matter and/or molecular (see for instance [110, 111, 84, 112, 52, 81, 91, 85, 82, 83, 87, 88]) use approximants to $\left.\frac{\delta T_s^{nad}[\rho,\rho_B]}{\delta\rho}\right|_{\rho=\rho_A}$ and $\left.\frac{\delta E_{xc}^{nad}[\rho,\rho_B]}{\delta\rho}\right|_{\rho=\rho_A}$.

In our own numerical implementations of Eq. 2.1 [113, 114], solving it proceeds by a similar self-consistent procedure as that most commonly applied for Kohn-Sham equations [2]. In both cases, the effective potential is a local potential which is a functional of electron density. Therefore, in the self-consistent procedure to solve either equations the effective potential is updated at each iteration. This is reflected in conventional numerical implementations of these two types of equations. In the present work, we consider an alternative pathway to reach self-consistency in the case of Eq. 2.1. The dominant component of the embedding potential (the first two terms in Eq. 2.7) is constant and does not require reevaluation during the self-consistent procedure. However, its exchange-correlation- and kinetic components need reevaluation each time ρ_A changes because each of them is the functional derivative of a density functional which is not linear in ρ_A (for detailed discussion of non-linearity of the exact functional $E_{xc}[\rho]$, see Ref. [115], and Ref. [116] for $T_s[\rho]$). The above considerations on non-linearity hold also for common approximants to $T_s[\rho]$ and $E_{xc}[\rho]$ used in Eq. 2.7. Concerning non-linearity of the kinetic energy component, linearizing it by expanding around reference density ρ_A^0 chosen to be the density in the absence of environment was found to be acceptable simplification [117]. For van der Waals contacts, hydrogen bonds, and complexes involving cations the the corresponding errors were found to be numerically insignificant.

This work builds on the above observations and the fact that, owing to similar analytic forms of the kinetic- and the exchange-correlation components of the orbital-free effective embedding potential, the latter can be expected to be treated in a similar way as that in the linearization scheme used previously for $T_s^{nad}[\rho_A, \rho_B]$. Linearizing not only $T_s^{nad}[\rho_A, \rho_B]$ but also $E_{xc}^{nad}[\rho_A, \rho_B]$ leads to the embedding potential, which does need to be recalculated each time ρ_A changes. Note that the electrostatic component of the embedding potential does not depend on ρ_A at all. Therefore, an alternative numerical procedures to solve Eq. 2.1 can be envisaged which take advantage of the ρ_A-independence of the total embedding potential after linearization of its non-linearized components. In the present work, such an alternative

- *splitSCF* - is proposed. Compared to the conventional self-consistent calculations, *splitSCF* involves two self-consistent loops (Figure 5.1. The inner one uses linearized embedding potential whereas the outer one assures that the linearized- and exact embedding potentials are the same in the end.

The *splitSCF* calculations provide, therefore, an alternative pathway to achieve self-consistency compared to conventional calculations. Splitting the iterative cycles opens also the possibility to introduce additional simplifications leading to reduction in time of calculations.

A key element in the *splitSCF* calculations is the *linearized orbital-free embedding potential.* For any set of embedded orbitals, the associated electron density is $\rho_A = 2\sum_i^{N_A}\left|\phi_i^A\right|^2$. Expanding the embedding potential $v_{emb}^{KSCED}[\rho_A,\rho_B;\vec{r}]$ in a Taylor series around some well-chosen reference density (ρ_A^0) and truncating the series after the first non-disappearing term leads to the potential referred to in this work as *linearized orbital-free effective embedding potential*:

$$\begin{aligned} v_{emb}^{KSCED}[\rho_A,\rho_B;\vec{r}] \;\approx\; & v_{emb}^{KSCED}[\rho_A^0,\rho_B;\vec{r}] = v_{ext}^{B}(\vec{r}) + \int \frac{\rho_B(\vec{r}')}{|\vec{r}'-\vec{r}|}d\vec{r}' + \quad (5.5)\\ & \left.\frac{\delta E_{xc}^{nad}\left[\rho,\rho_B\right]}{\delta\rho}\right|_{\rho=\rho_A^0} + v_t[\rho_A^0,\rho_B](\vec{r}) \end{aligned}$$

The above potential is to be used in the inner loop in the *splitSCF* calculations (Figure 5.1). The *splitSCF* scheme to solve Eq. 2.1 brings another advantages. We recall here that the last two terms in the potential given in Eq. 2.7 are density functionals. Therefore, if the densities ρ_A^0 and ρ_A are the same, the corresponding embedding potentials are the same as well. If, however, they differ only slightly, skipping the outer loop might be an acceptable approximation (*linearization approximation*). If the linearization approximation is applied in particular large-scale computer simulations, the *splitSCF* provides an easy to use tool to estimate the adequacy of such simplification for any type of contacts in the interface between the system described at orbital-level and its environment. The differences between the results derived from Eq. 2.1 using either Eq. 2.7 or Eq. 5.5 form of the embedding potential will be referred as *linearization errors* in this work. From our numerical experience [117], the good choice for ρ_A^0 is the isolated density of the embedded subsystem.

The present work concerns the linearization of only the $v_{emb}^{KSCED}[\rho_A,\rho_B;\vec{r}]$ component of the whole effective potential in Eq. 2.1. For several reasons,

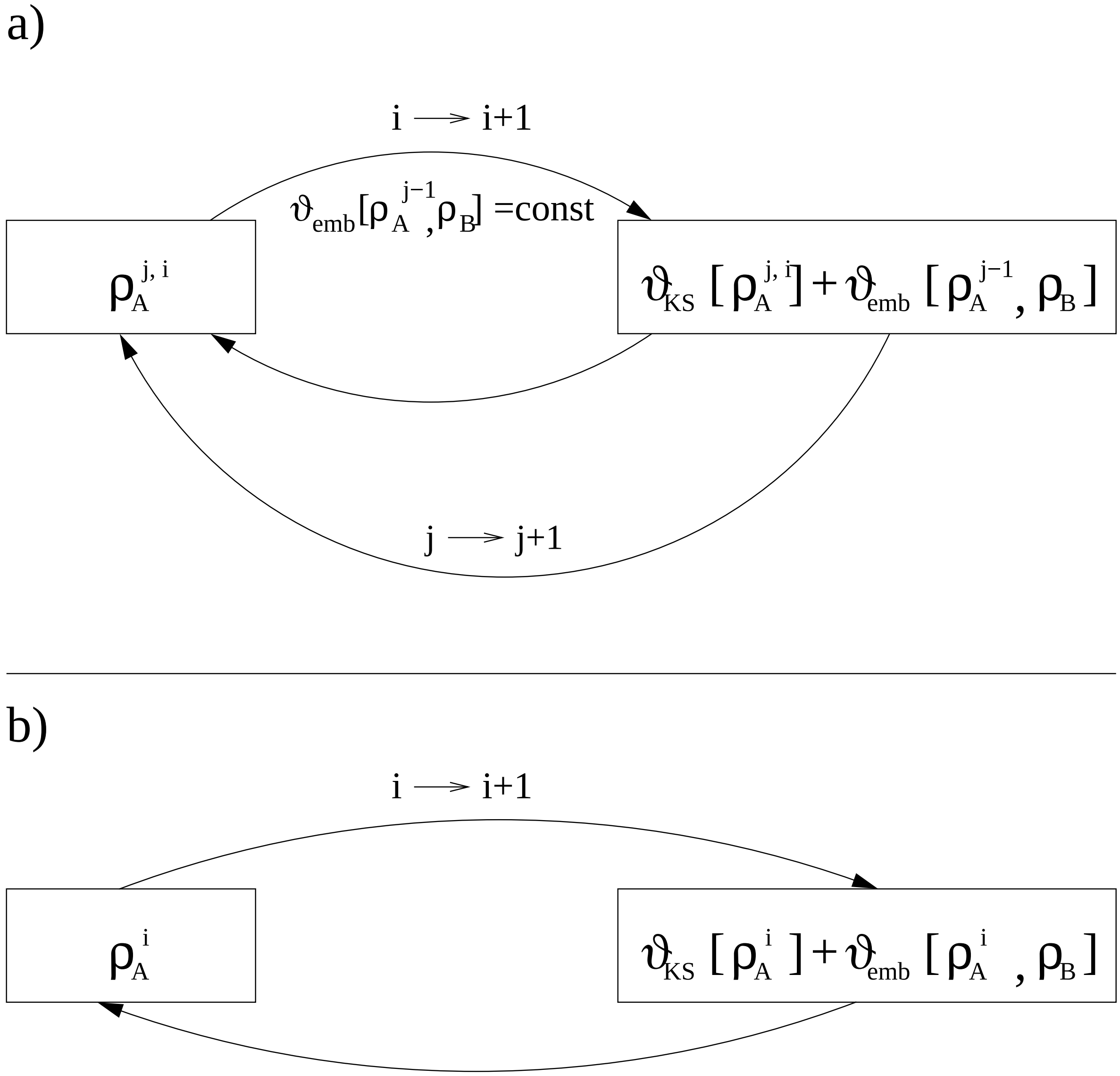


Figure 5.1: a) The *splitSCF* scheme: In the inner loop (i-index), the embedding potential $v_{emb}[\rho_A, \rho_B]$ is evaluated for ρ_A taken from the previous iteration in the outer loop (j-index) and remains constant, whereas the $v_{KS}[\rho_A]$ component is recalculated as ρ_A changes. b) The conventional SCF scheme: Both $v_{KS}[\rho_A]$ and $v_{emb}[\rho_A, \rho_B]$ are recalculated as ρ_A changes.

linearization of the whole effective potential is of lesser practical interest. The $E_{xc}^{nad}[\rho_A, \rho_B]$ and $T_s^{nad}[\rho_A, \rho_B]$ terms represent the differences between the energies of the embedded system and its environment in their complexed form and their isolated forms. Compared to the electrostatic component of $v_{emb}^{KSCED}[\rho_A, \rho_B; \vec{r}]$ (which doeas not depend on ρ_A at all), or to the exchange-correlation term in $v_{eff}^{KS}[\rho_A; \vec{r}]$, they involve small contribution to the total energy, therefore, errors arising from their linearization can be expected to be less significant. Retaining the exchange-correlation component of $v_{eff}^{KS}[\rho_A; \vec{r}]$ in its non-linearized original form makes it also possible to perform *splitSCF* in such schemes which use different approximants for the exchange-correlation potential in $v_{eff}^{KS}[\rho_A; \vec{r}]$ and in the embedding potential. This concerns cases where the use of more involved Kohn-Sham-equation-based computational schemes are needed for $v_{eff}^{KS}[\rho_A; \vec{r}]$ (for example such where orbital-dependent forms of the exchange-correlation energy are used [118]). In this context, it is worthwhile to mention computational schemes which use the orbital-free effective embedding potential of Eq. 2.7 but outside of the formal framework given in Eq. 2.1 such as the use of this embedding potential in combination of wave-function based treatment of embedded subsystem [36]. Practical calculations following such a generalized scheme have been already reported in the literature where the linearized [37] or partially linearized [38] potential given in Eq. 2.7 was used. The present work provides the numerical analysis of the adequacy of this simplification made in such calculations. To this end, we use the *splitSCF* scheme to investigate the magnitude of the errors introduced by using the linearized functionals in the orbital-free effective embedding potential given in Eq. 5.5 for such observables as molecular density, embedded orbitals, interaction energies, or complexation induced dipole moments.

5.2 Computational details

5.2.1 Approximants to $T_s^{nad}[\rho_A, \rho_B]$ and $E_{xc}^{nad}[\rho_A, \rho_B]$

In principle, the convergence of the *splitSCF* procedure and the adequacy of the linearization approximation should be subject of dedicated studies for any approximant to $T_s^{nad}[\rho_A, \rho_B]$ and $E_{xc}^{nad}[\rho_A, \rho_B]$. In our applications of Eq. 2.1, only two types of approximants to the orbital-free effective embedding potential are used so far - one which is gradient-free and another one which depends on up-to second derivatives of the electron density. For this reason, the present analysis is concerns only these two recommended approximants to the potential given in Eq. 2.7: the gradient-free (denoted *LDA*) and gradient-dependent (denoted with *GGA*). In each case, the approximants to $T_s^{nad}[\rho_A, \rho_B]$ and $E_{xc}^{nad}[\rho_A, \rho_B]$ are matched by their formal origin and types of variable on they depend (either gradient-free or gradient-dependent).

As far as $T_s^{nad}[\rho_A, \rho_B]$ is concerned, the analytic form of its gradient-free approximant is obtained using Eq. 5.1 and the Thomas-Fermi expression for the kinetic energy functional [9, 10]. In the gradient-dependent case, the Lembarki-Chermette [98] approximant is used for this purpose. The gradient-free approximant, was used already by Cortona [28] in calculations following his formulation of density functional theory which also hinge on the approximant to $T_s^{nad}[\rho_A, \rho_B]$ and in our original work introducing orbital-free embedding [4]. This approximant is adequate in most cases although suffers from some systematic flaws (see Refs. [13, 119]). The gradient-dependent approximant to $T_s^{nad}[\rho_A, \rho_B]$ was introduced in Ref. [72] to overcome the problems of the quantitative wrong behavior of the gradient-dependent approximant derived from regular gradient expansion approximantion [96] reported in Ref. [73]. The analytic forms of these two approximants to the kinetic energy component of the orbital-free embedding potential can be found in Refs. [4, 67].

As far as $E_{xc}^{nad}[\rho_A, \rho_B]$ is concerned, its analytic form is obtained using Eq. 5.2 and either the local density approximation [16, 18, 17] or the Perdew and Wang [120, 121] form of the exchange-correlation energy.

5.2.2 Test sets of intermolecular complexes

The current generation of approximants to $T_s^{nad}[\rho_A, \rho_B]$, make the orbital-free embedding calculations adequate only for such cases where the overlap between the densities ρ_A and ρ_B is small. As a rule of thumb, weak intermolecular complexes near and beyond equilibrium geometry fall into this category. A cheaply calculable approximant to $T_s^{nad}[\rho_A, \rho_B]$ applicable for cases where ρ_A and ρ_B correspond to covalently linked subsystems is probably not in view [13].

The set of the systems used to test the adequacy of the *linearization approximation* follows into category of small ρ_A-ρ_B overlap systems. The considered intermolecular complexes can be ordered according to increasing magnitude of deformation of electron density upon the intermolecular interaction:

weakly bonded van der Waals complexes (Ne-Ne, CH_4-CH_4, C_2H_2-C_2H_2), hydrogen bonded complexes (HF-HF, H_2O-H_2O, NH_3-H_2O), one charged complex (Li^+-H_2O), and one charge-transfer complex (NH_3-ClF). The same set of intermolecular complexes was used in our previous study on concerning linearization of $T_s^{nad}[\rho_A, \rho_B]$. The charge-transfer complex (NH_3-ClF) is included in the set although linearization can not be expected to be applicable in this a case. This system provides a useful reference for discussion of the convergence of the *splitSCF* procedure.

5.2.3 Numerical implementation

The aug-cc-pVTZ [122, 123] (cc-pVTZ [124, 123] for Li) atom centered basis sets is used in all centers of the system (supermolecular expansion of the basis sets refereed to as KSCED(s) calculations in Ref. [72]). The terms representing electron-electron interactions are calculated using the set of fitting functions referred to as GEN-A4* in Ref. [125]. Numerical integrations were performed using the pruned grid with 90 radial and 590 angular points. All the calculations are performed using our code - deMon2K-KSCED [113] - which solves Eq. 2.1 and is based on the deMon2K [126] code.

5.3 Results

Performing the full double-loop of *splitSCF* leads to results which are numerically indistinguishable from that obtained in conventional calculations (within the tolerance of convergence criterion set to 10^{-8}) in all investigated systems.

Figure 5.2 shows the convergence of the outer loop for the H_2O molecule in the H_2O-H_2O complex. In this case, one water molecule corresponds to ρ_A and the other to ρ_B according to the convention used in all equations in this work. Both partners in this complex are polar and the density derived from Eq. 2.1 for the water molecule in the dimer can be expected to differ from that of the isolated water molecule due to electronic polarization. It is also worthwhile to recall that atom centered basis is used to construct embedded orbitals and that all the centers in the dimer are included. The complexation induced changes of electron density can involve, therefore, smearing of electron density among both partners in the dimer. For all quantities shown in Figure 5.2, the result of the conventional calculations are reached practically in the third outer loop iteration. Also in the charge-transfer complex NH_3-ClF, for which the largest complexation induced changes of electron density occurs among the systems investigated here, *splitSCF* the outer loop converges rapidly (the results obtained in four iterations and the convergent ones are indistinguishable in the figure). The above result suggests that skipping the outer-loop can be considered as an additional useful simplifications in many possible applications.

The principal body of results presented in this work concerns a more detailed analysis of the numerical consequences of skipping the outer loop i.e. making the *linearization approximation* for a number of molecular properties. For each system, the considered quantities include: *i*) the norm of the difference between the density and the Kohn-Sham density of the whole system, *ii*) the orbital energies, *iii*) the dipole moment, *iv*) the interaction energy. The first quantity is a well defined mathematical object whereas the other ones are of more direct interest to chemical applications. The linearization errors in orbital energies and dipole moments are determined only by the difference between the exact and linearized embedding potential. In each case, the difference between the results obtained with- and without linearization approximation is analyzed.

5.3.1 Electron densities

We start the analysis of the effect of the linearization approximation on electron density. To this end, the geometrical distance (denoted by M) between the electron density derived from either the Kohn-Sham calculations or Eq. 2.1 is analyzed. M is defined as:

$$M = \frac{1}{N}\sqrt{\int \left(\rho^{KS}(\vec{r}) - (\rho_A^{KSCED}(\vec{r}) + \rho_B(\vec{r}))\right)^2 d\vec{r}} \tag{5.6}$$

where N is the number of electrons in both subsystems, and $\rho_A^{KSCED}(\vec{r})$ is the density obtained either with- or without linearization of the orbital-free embedding potential.

Linearization errors in M are evaluated as differences between M obtained with- and without linearization approximation. M might be non-zero and positive even if the linearization approximation is not applied for two principal reasons. First of all, the difference $\rho^{KS}(\vec{r}) - \rho_B(\vec{r})$ might be not pure-state non-interacting v-representable. For instance, if the assumed electron density ρ_B is larger than $\rho^{KS}(\vec{r})$ for some $\vec{r}$. Secondly, even if $\rho^{KS}(\vec{r}) - \rho_B(\vec{r})$

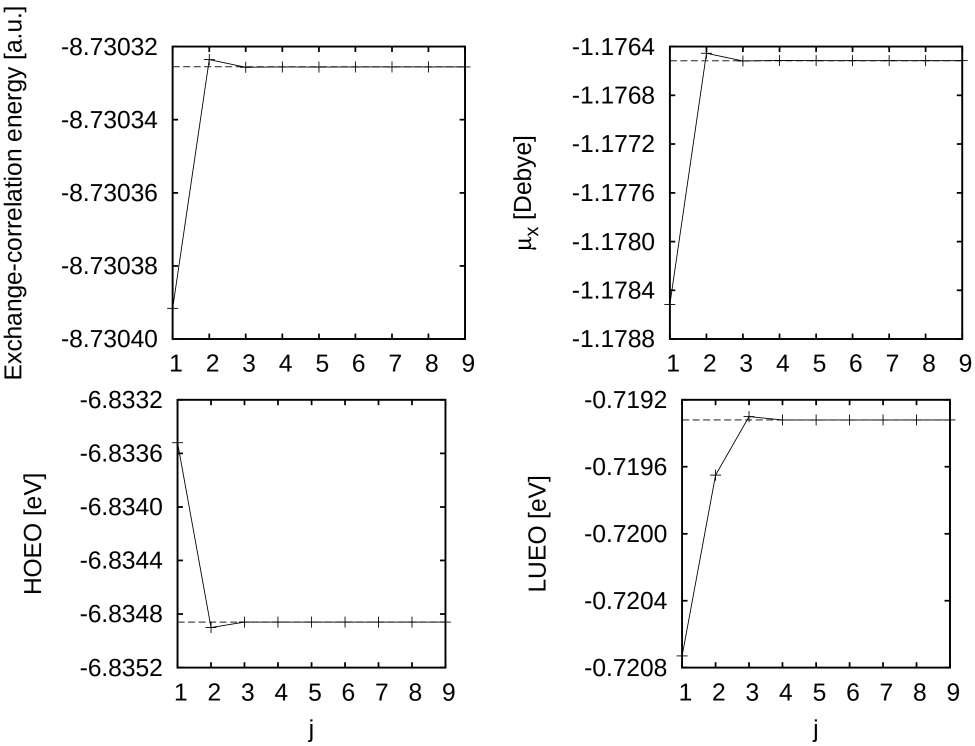


Figure 5.2: The convergence of the outer-loop of the *splitSCF* procedure for various properties of the H_2O molecule in the H_2O-H_2O complex. The results of conventional self-consistent calculations are indicated by dashed lines.

is pure-state non-interacting v-representable, replacing the exact functional $T_s^{nad}[\rho_A, \rho_B]$ by an approximant in Eq. 2.1 leads to an error in electron density. As a consequence, the obtained density ρ_A might differ from $\rho^{KS}(\vec{r}) - \rho_B(\vec{r})$.

Since electron density is a local quantity comparisons between two densities are not straightforward. The use of a global quantity, M makes the discussion of the effects of linearization errors on density significantly simpler. The linearization errors in M together with the reference values of this quantity are collected in tables 5.1 and 5.2.

Table 5.1: The effect of linearization approximation on electron density. For definition of the norm M, see Eq. 5.6. Data for gradient-free approximant (see text) to the orbital-free embedding potential.

Subsystem A	Subsystem B	M_{lin}	M	Error (ppm)
Ne	Ne	0.66679989	0.66679989	0.00
CH_4	CH_4	0.28714628	0.28714629	0.03
C_2H_2	C_2H_2	0.28925929	0.28925929	0.00
H_2O	H_2O^a	0.46032338	0.46032348	0.22
H_2O^a	H_2O	0.46034112	0.46034127	0.33
NH_3^a	H_2O	0.41736040	0.41736046	0.14
H_2O	NH_3^a	0.41733219	0.41733230	0.26
HF^a	HF	0.55946510	0.55946525	0.27
HF	HF^a	0.55945578	0.55945582	0.07
Li^+	H_2O	0.55220921	0.55220917	-0.07
H_2O	Li^+	0.55239902	0.55239716	-3.37
NH_3	ClF	0.68951231	0.68951448	3.15
ClF	NH_3	0.68945022	0.68945044	0.32

[a] hydrogen-bond acceptor.

Table 5.2: The effect of linearization approximation on electron density. For definition of the norm M, see Eq. 5.6. Data for gradient-dependent approximant (see text) to the orbital-free embedding potential.

Subsystem A	Subsystem B	M_{lin}	M	Error (ppm)
Ne	Ne	0.66960937	0.66960938	0.01
CH_4	CH_4	0.28994978	0.28994979	0.03
C_2H_2	C_2H_2	0.29195543	0.29195543	0.00
H_2O	H_2O^a	0.46325912	0.46325905	-0.15
H_2O^a	H_2O	0.46326234	0.46326231	-0.06
NH_3^a	H_2O	0.42026551	0.42026560	0.21
H_2O	NH_3^a	0.42025334	0.42025329	-0.12
HF^a	HF	0.56237171	0.56237161	-0.18
HF	HF^a	0.56237513	0.56237508	-0.09
Li^+	H_2O	0.55588946	0.55588985	0.70
H_2O	Li^+	0.55609369	0.55609600	4.15
NH_3	ClF	0.69043602	0.69043492	-1.59
ClF	NH_3	0.69037676	0.69037674	-0.03

[a] hydrogen-bond acceptor

The quality of the densities is very good, both for LDA and GGA approximants to the embedding potential. Comparison to the non-linearized case reveals that most of the differences are very small, typically relative errors are in the range of 0.3 ppm. In the case of H_2O-Li^+, strong polarization of the water molecule due to the interaction with charged Li^+ atom results in the slightly larger error than in other molecules. Similarly, for the NH_3-ClF molecule, the deformation of the density because of the charge transfer character of the interaction leads to the relative linearization error reaching 3.15 ppm in the case of the gradient-free approximant.

Small linearization errors in M are not easily related to any quantities used in chemically relevant discussions. Appropriate quantities will be analyzed in the subsequent sections. Here, we note that the absolute value of M arises from the choice made for ρ_B and the choice made for the approximants to the kinetic energy component of the orbital-free embedding potential. The fact that the linearization errors are significantly smaller than M itself indicates that the former choices affect the electron density derived from Eq. 2.1 more significantly than the linearization approximation.

5.3.2 Orbital energies

The linearization errors in orbitals energies together with the complexation induced shifts of these quantities are collected in Tables 5.3 and 5.4. Only the energies of the highest occupied embedded orbital (HOEO) and lowest unoccupied embedded orbital (LUEO) are reported here.

The linearization errors for hydrogen-bonded- and van der Waals complexes are very small. Typically, the errors are two orders of magnitude smaller than complexation induced shifts of the orbital energy. In the LDA case, strong polarization of H_2O molecule induced by Li^+ cation in the Li^+-H_2O complex results in the linearization error reaching up to 8.0 meV and 117.3 meV, for HOEO and LUEO, respectively. Although such magnitude of the error is the largest in this group of complexes which does not include the charge-transfer ones, the linearization approximation is still acceptable. In the LUEO case, it amounts to only 1.8% of the total complexation induced orbital energy shift. As expected, the strong deformation of the electron density due to

Table 5.3: Linearization errors in energies (in meV) of the highest occupied embedded orbital and the lowest unocupied embedded orbital. Data for gradient-free approximant (see text) to the orbital-free embedding potential. The complexation induced shifts of orbital energies ($\Delta\epsilon^{HOEO}$ and $\Delta\epsilon^{LUEO}$) are given for comparison.

Subsystem *A*	Subsystem *B*	$\delta^{lin}(\Delta\epsilon^{HOEO})$	$\delta^{lin}(\Delta\epsilon^{LUEO})$	$\Delta\epsilon^{HOEO}$	$\Delta\epsilon^{LUEO}$
Ne	Ne	0.0	0.0	-2.9	82.2
CH_4	CH_4	0.2	8.6	16.6	-252.0
C_2H_2	C_2H_2	0.7	14.6	-39.4	-143.7
H_2O	H_2O^a	-1.3	1.4	551.4	216.4
H_2O^a	H_2O	0.8	0.0	-586.1	-220.0
NH_3^a	H_2O	0.4	-0.7	-669.8	-196.6
H_2O	NH_3^a	-1.4	4.5	710.6	134.0
HF^a	HF	1.4	0.7	-993.4	-468.8
HF	HF^a	-1.2	2.2	410.2	212.5
Li^+	H_2O	-0.1	1.3	1535.4	398.8
H_2O	Li^+	8.0	117.3	-7077.2	-6568.9
NH_3	ClF	26.5	151.2	-979.7	-603.7
ClF	NH_3	0.1	0.8	884.3	1333.5

[a] hydrogen-bond acceptor

Table 5.4: Linearization errors in energies (in meV) of the highest occupied embedded orbital and the lowest unocupied embedded orbital. Data for gradient-dependent approximant (see text) to the orbital-free embedding potential. The complexation induced shifts of orbital energies ($\Delta\epsilon^{HOEO}$ and $\Delta\epsilon^{LUEO}$) are given for comparison.

Subsystem A	Subsystem B	$\delta^{lin}(\Delta\epsilon^{HOEO})$	$\delta^{lin}(\Delta\epsilon^{LUEO})$	$\Delta\epsilon^{HOEO}$	$\Delta\epsilon^{LUEO}$
Ne	Ne	0.1	1.8	-10.9	-42.8
CH_4	CH_4	0.2	9.8	-2.1	-293.8
C_2H_2	C_2H_2	0.3	25.2	-58.1	-169.9
H_2O	H_2O^a	1.0	9.1	508.6	168.4
H_2O^a	H_2O	-0.4	-3.6	-644.9	-252.9
NH_3^a	H_2O	1.2	-2.5	-746.4	-223.0
H_2O	NH_3^a	1.1	10.8	756.5	174.1
HF^a	HF	-1.9	-6.9	-1037.7	-506.6
HF	HF^a	0.9	10.9	361.5	141.5
Li^+	H_2O	0.9	4.5	1449.2	73.6
H_2O	Li^+	3.9	-65.5	-7127.2	-6547.8
NH_3	ClF	-6.9	-2.9	-1253.7	-1129.2
ClF	NH_3	1.2	5.2	817.1	1207.9

[a] hydrogen-bond acceptor

the charge-transfer character of NH_3-ClF is a source of a large linearization error in orbital energies. The error in the energy of LUEO reaches even 25% of the complexation induced shift. These large relative errors confirm that linearization approximation is not acceptable for the complexes of this type.

5.3.3 Interaction energies

The results collected in table 5.5 reveal very small linearization errors for both gradient-free and gradient-dependent approximants to the embedding potential. As expected, due to the significant deformation of the density in H_2O-Li^+, the linearization error is more pronounced for this complex. Its magnitude is, however, five orders of magnitude smaller than the absolute value of the binding energy, which is of no importance in discussions relevant for chemical accuracy. As expected, the linearization errors reaching 0.01712 kcal/mol (we provide four significant numbers for reference purposes) are the largest in the NH_3-ClF case. The linearization induced errors in the interaction energies tend to be smaller in the case of gradient-free (LDA) than gradient-dependent (GGA) approximant to the embedding potential.

5.3.4 Dipole moments

Tables 5.6 and 5.7 collect linearization errors in complexation induced dipole moments.

The errors are rather small in most cases. Except for C_2H_2-C_2H_2, CH_4-CH_4, and NH_3-ClF, they do not exceed 1% of complexation induced shifts of the dipole moment. Surprisingly, the largest error is found for C_2H_2-C_2H_2, reaching 14% of and 4% of $\Delta\mu_y^A$ for LDA- and GGA approximants respectively. These large relative errors exceed even the ones in the NH_3-ClF complex, for which the relative errors equal to 7% and 1.4% respectively. Moreover, in the C_2H_2-C_2H_2 complex the linearization errors in $T_s^{nad}[\rho_A, \rho_B]$ and in $E_{xc}^{nad}[\rho_A, \rho_B]$ do not cancel (for comparison, see our previously reported results [117]). This result indicates that adequacy of the linearization

Table 5.5: Linearization errors in the energy of interaction (in kcal/mol) between embedded subsystem and its environment. LDA and GGA denote gradient-free and gradient-dependent approximants to the embedding potential, respectively (see text).

		LDA		GGA	
Subsystem A	Subsystem B	E_{int}	$\delta^{lin}(E_{int})$	E_{int}	$\delta^{lin}(E_{int})$
Ne	Ne	-0.08	0.00000	-0.40	0.00000
CH_4	CH_4	-0.43	0.00000	-0.84	0.00003
C_2H_2	C_2H_2	-1.53	0.00008	-1.88	0.00015
H_2O	H_2O^a	-4.07	0.00017	-4.86	0.00064
H_2O^a	H_2O	-3.77	0.00004	-4.68	0.00023
NH_3^a	H_2O	-5.05	0.00005	-6.15	0.00063
H_2O	NH_3^a	-5.44	0.00019	-6.34	0.00064
HF^a	HF	-3.21	0.00009	-4.09	0.00033
HF	HF^a	-2.86	0.00010	-3.96	0.00050
Li^+	H_2O	-25.99	0.00000	-27.34	0.00001
H_2O	Li^+	-40.87	0.00220	-43.21	0.00190
NH_3	ClF	-0.88	0.01712	-7.34	0.00766
ClF	NH_3	-0.63	0.00023	-5.70	0.00159

[a] hydrogen-bond acceptor

Table 5.6: Linearization errors in dipole moments (in mDebye). Data for gradient-free approximant (see text) to the orbital-free embedding potential. Interaction induced dipole $\Delta\mu_q^A$ moments are given for comparison.

Subsystem A	Subsystem B	$\delta^{lin}(\Delta\mu_x^A)$	$\delta^{lin}(\Delta\mu_y^A)$	$\delta^{lin}(\Delta\mu_z^A)$	$\Delta\mu_x^A$	$\Delta\mu_y^A$	$\Delta\mu_z^A$
Ne	Ne	0.0	0.0	0.0	0.0	0.0	0.7
CH_4	CH_4	0.0	0.0	0.3	-0.0	0.0	-6.4
C_2H_2	C_2H_2	-1.0	1.6	0.0	-117.2	10.9	0.0
H_2O	H_2O^a	-2.0	-0.0	0.4	-262.7	-3.3	43.9
H_2O^a	H_2O	-1.0	0.0	-0.3	-206.4	4.6	65.4
NH_3^a	H_2O	-0.4	0.6	0.0	-264.9	72.3	0.3
H_2O	NH_3^a	-2.2	-0.3	-0.0	-344.1	-21.6	-0.1
HF	HF^a	1.4	0.2	-0.0	238.1	58.1	-0.0
HF^a	HF	1.2	-0.2	0.0	145.9	-38.0	0.0
Li^+	H_2O	0.0	0.0	0.0	0.0	0.0	-19.7
H_2O	Li^+	0.0	0.0	-10.7	-0.0	-0.0	-1596.0
NH_3	ClF	0.0	-0.0	-47.3	0.0	0.0	-675.8
ClF	NH_3	0.0	0.0	0.7	-0.0	0.0	-795.8

[a] hydrogen-bond acceptor

Table 5.7: Linearization errors in dipole moments (in mDebye). Data for gradient-dependent approximant (see text) to the orbital-free embedding potential. Interaction induced dipole $\Delta\mu_q^A$ moments are given for comparison.

Subsystem A	Subsystem B	$\delta^{lin}(\Delta\mu_x^A)$	$\delta^{lin}(\Delta\mu_y^A)$	$\delta^{lin}(\Delta\mu_z^A)$	$\Delta\mu_x^A$	$\Delta\mu_y^A$	$\Delta\mu_z^A$
Ne	Ne	0.0	0.0	0.0	0.0	0.0	2.2
CH_4	CH_4	-0.0	0.0	0.4	0.0	-0.0	-7.1
C_2H_2	C_2H_2	-0.4	0.5	0.0	-114.5	10.9	0.0
H_2O	H_2O^a	1.0	-0.0	0.2	-247.1	-3.2	41.6
H_2O^a	H_2O	0.4	0.0	-0.2	-209.7	4.4	-62.3
NH_3^a	H_2O	-1.7	0.4	0.0	-274.5	68.8	0.3
H_2O	NH_3^a	1.1	-0.0	-0.0	-328.5	-21.8	-0.1
HF^a	HF	-1.8	-0.1	0.0	238.9	57.4	-0.0
HF	HF^a	-0.7	-0.2	0.0	133.9	-36.5	0.0
Li^+	H_2O	0.0	0.0	0.1	-0.0	-0.0	-17.5
H_2O	Li^+	0.0	0.0	-3.1	-0.0	-0.0	-1648.0
NH_3	ClF	0.0	-0.0	11.6	0.0	-0.0	-858.1
ClF	NH_3	0.0	0.0	3.1	0.0	-0.0	-725.6

[a] hydrogen-bond acceptor

approximation should be subject to verifications on the case by case basis. The proposed *splitSCF* scheme provides a numerical tool for performing such verifications.

5.4 Discussion and Conclusions

The conventional strategy to obtain self-consistent solution of Eq. 2.1 and *splitSCF* lead to equivalent numerical results for all systems considered. All *splitSCF* calculations reported here started with ρ_A^0 chosen to be that ground state electron density of the embedded molecule in absence of environment. Since the potential given in Eq. 2.7 is an explicit functional of electron density, the linearization errors disappear if ρ_A^0 happens to be equal to ρ_A. The numerical studies presented in this work show that the density of the isolated embedded molecule is an optimal choice for ρ_A^0 used to start the *splitSCF* calculations. The outer loop converges rapidly. For the H_2O-H_2O complex, the convergent solution is practically reached in one outer-loop calculation. Even for charge-transfer complexes, where linearization approximation is not adequate, the *splitSCF* calculations are applicable because of the fast convergence of the outer loop.

We do not, however, propose using *splitSCF* in large-scale computations as an alternative to conventional algorithms. This work focuses rather on using the simplified scheme *linearization approximation*, in which the outer loop iterations are not performed at all, and the actual electron density ρ_A is replaced by the density of the isolated subsystem in the analytic expression for the embedding potential given in Eq. 2.7. For each particular type of interactions with the environment, we propose to use the *splitSCF* procedure (Fig. 5.1) to verify the adequacy of such linearization approximation.

It was found that the linearization errors in such quantities as: energies of embedded orbitals, energy of interaction with the environment, and dipole moment, are not significant for a large class of embedded systems including such where the interactions with environment involve: van der Waals bonded contacts, hydrogen bonds, and even charged species. In charge-transfer complexes, however, the molecular densities are strongly deformed upon complexation and the *linearization approximation* is not acceptable. Indeed, the numerical results for the NH_3-ClF complex show that the error in the orbital energies reach up to 25% of magnitude of the whole complexation induced shift of the energy level.

Turning back to systems for which linearization is adequate, it is worthwhile to notice that they also comprise charged complexes in which the electron density of a neutral partner can be expected to differ significantly from that in the absence of interactions (electronic polarization). Compared to van der Waals or hydrogen-bonded complexes, the deformation of the density upon complexation leads to noticeably larger linearization errors in the Li^+-H_2O complex. These errors are, however, rather small and the linearization approximation remains an acceptable simplifications even in such a case. It is important to underline that the linearization aoproximation is not directly related to neglect of electronic polarization which is accounted for by the electrostatic component of the orbital-free embedding potential in its either exact or linearized form. The exchange-correlation- and non-additive kinetic energy terms include electron repulsion, which is much smaller compared to the electrostatic components, linearizing these functionals in ρ_A by expanding them around ρ_A^0 is still applicable even for charged complexes.

If applicable, the linearization approximation can be used to reduce the time of calculations. In our current numerical implementation of the *splitSCF* scheme, which does not take all of its possible advantages in this respect, the time needed to perform one iteration is reduced from 25 s/iteration (conventional calculations) to 17 s/iteration (inner loop in the *splitSCF* calculations) even for such small systems as the HF-HF dimer. For larger systems, avoiding recalculation of the embedding potential within the self-consistent loop leads to larger savings. For C_2H_2-C_2H_2, the corresponding times are 859.8

s/iteration and 145.5 s/iteration. For systems, where the environment is significantly larger than the embedded system, and if linearization approximation is applicable, the expected computational benefits of this approximation are the largest.

Besides the possibility of full control of the linearization errors, the *splitSCF* scheme offers other advantages. It can be relatively easy implemented into existing quantum chemistry codes, because the embedding potential given in Eq. 5.5 is a conventional ρ_A-independent external potential and can be simply implemented as an additional constant potential in the procedure evaluating two-center integrals. Moreover, since a relatively small number of outer-loop cycles is needed (three-four) to converge to the same solution as the conventional scheme (Fig. 5.2), *splitSCF* offers possibility of additional computational savings due to the constancy of the embedding potential in the inner loop.

Essentially, the same conclusions concerning applicability of the linearization approximation can be drown for gradient-free and gradient-dependent approximates to the orbital-free effective embedding potential given in Eq. 2.7. Typically, however, the linearization errors are smaller in the latter case. The present analysis concerns only two approximant to the orbital-free embedding potential. Recently, a new approximant to $T_s^{nad}[\rho_A, \rho_B]$ was constructed (Eq. 23 in [119]) by enforcing the proper behavior of its derivative at small ρ_A. The same conclusions concerning convergence of *splitSCF* and adequacy of the linearization approximation can be expected for this new approximant because it comprises terms already linear in ρ_A and the LDA approximant considered in this work.

Finally, it is worthwhile to point out our previously reported study [117] on linearization of only $T_s^{nad}[\rho_A, \rho_B]$ for the same set of intermolecular complexes. The previous and the current analyses lead to essentially the same conclusions concerning the adequacy of using the isolated molecule density as ρ_0^A in Eq. 5.5. Comparing the linearization errors discussed in these two works indicates that the higher order terms in bi-functionals $T_s^{nad}[\rho_A, \rho_B]$ and $E_{xc}^{nad}[\rho_A, \rho_B]$ expanded around ρ_A^0 cancel each other to some extend.

Chapter 6

Equilibrium geometries of noncovalently bound intermolecular complexes derived from subsystem formulation of density functional theory

In this chapter the subsystem formulation of density functional theory is used to obtain equilibrium geometries and interaction energies for a representative set of non-covalently bound intermolecular complexes. The results are compared with literature benchmark data. The range of applicability of two considered approximations to the exchange-correlation- and non-additive kinetic energy components of the total energy is determined. Local density approximation, which does not involve any empirical parameters, leads to excellent intermolecular equilibrium distances for hydrogen bonded complexes (maximal error 0.13 Å for NH_3-NH_3). It is a method of choice for a wide class of weak intermolecular complexes including also dipole-bound and the ones formed by rare gas atoms or saturated hydrocarbons. The range of applicability of the chosen generalized gradient approximation, which was shown in our previous works to lead to good interaction energies in π-stacked systems, remains limited to this group because it improves neither binding energies nor equilibrium geometries in the wide class of complexes for which local density approximation is adequate.

The subsystem formulation of DFT allows to perform complete (or partial) minimization of the interface energy with respect to the density and geometrical parameters of both (or one of) subsystems. Several algorithms are possible to achieve this task. All of them are outlined in this chapter. We discuss in detail an efficient energy minimization procedure in which optimization of the geometry and the electron density of each subsystem is made simultaneously. This approach leads to the most accurate results. The representative application this strategy can be found in work by Fradelos *et al.* [127] where the cooperative effect of hydrogen-bonded chains in the environment of *cis*-7-hydroxyquinoline was studied. Algorithms involving partial minimization of only one subsystem are also of practical importance. They become handy in problmes where the geometry of only one of interacting parnters is subject of interest, for instance equilibrium structure of small chromophore embedded in large environment. Moreover, the time required to perform optimization of only one subsystem is reduced when compared to the schmes involving full optimization of the whole interface.

Additional savings in the time of calculations in the case of geometry optimizations with subsystem formulation of DFT might be also gained by using described in previous chapter linearized orbital-free embedding potential and *splitSCF* technique. In this case also geometrical degrees of freedom are optimized in addition to the two iterative loops shown in Figure 5.1. It has to be noted, however, that in such calculations convergence of the outer loop is indispensable to obtain forces. The *splitSCF* technique can be considered of a great value in the studies where the equilibrium geometry of a large systems needs to be find. In this chapter on the example of selected benchmark molecules the first application of the linearized orbital-free embedding potential in the geometry optimizations is presented together with the discussion on the accuracy and benefits arising from this approach.

6.1 Introduction

The principal motivation for this work originates in our interest in the *orbital-free embedding* formalism [4] to study environment-induced changes of the electronic structure of an embedded species: localized electronic excitations [52, 86], hyperfine tensor [111], dipole moments [110], f-levels [84], the gap between the high- and low spin potential energy surfaces [112], for instance. In the orbital-free embedding calculations, all the information about the environment is confined in its electron density and only the selected subsystem is described at the orbital level.

The quality of such properties of the total system as electron density distribution, total energy, response properties, etc., derived from the orbital-free embedding calculations is determined by two factors: the use of approximate density functionals for exchange-correlation- and non-additive kinetic energy instead of the corresponding exact quantities (see the methods section below), and the choice of the electron density corresponding to the environment, which is derived from some other methods involving lower computational costs. Whereas the accuracy of the used functionals cannot be controlled because their exact forms are known only for some systems, the effect of the choice of the electron density of the environment can be easily verified in practice. The electron density assigned to the environment can be also subject of optimization. The process of minimization of the total energy with respect to both components of the total electron density can proceed as a series of partial minimizations, in which both subsystems exchange their roles until minimum is reached. [31] Of course, both subsystems are treated in such calculations on equal footing and the notion of *environment* and *embedded subsystem* looses its meaning at the end of the minimization procedure. Fully variational calculations represent numerical implementation of the subsystem formulation of density functional theory (DFT) introduced by Cortona [28].

In the multi-level computer simulations applying orbital-free embedding formalism, fully variational calculations can be applied as a complementary tool to asses the adequacy of the electron density chosen to represent the environment. For instance, the effect of relaxation of the electron density of the environment in model systems was reported in several previous publications [52, 86, 111, 84].

This work concerns the source of errors in orbital-free embedding calculations arising from the use of approximate density functionals for exchange-correlation and non-additive kinetic energies. To this end, the subsystem formulation of DFT is used to minimize the total energy with respect to electron densities of both subsystems in a representative sample of weakly interacting intermolecular complexes. Compared to investigations of the adequacy of the applied density functional reported previously, we focus the analysis not on interaction energies only but on equilibrium geometries.

The effect of the environment on the electronic structure of the embedded subsystem can be seen as the result of two effects: the environment induced changes of the geometry and the direct electronic effects (for a recent representative analysis, see Ref. [90]). In many cases, the geometry of the investigated system is known form either experiment or computational studies applying other methods. It would be, however, desirable to apply the *orbital-free embedding* type of calculations also to optimize the geometry of the embedded subsystem without relying on structural data obtained from other methods.

Studying the applicability of the subsystem formulation of density functional theory to derive equilibrium geometries, is made here not only for the outlined pragmatic reasons. Opposite to the errors in the total energy, which originate from the errors in the functionals and their derivatives, the errors in the equilibrium geometry originate only from the fact that the functional derivatives (effective potentials) of the relevant density functionals are not exact. We note that the errors in electron density and all one-electron properties also depend only on the quality of the effective potentials.

Opposite to the Kohn-Sham formulation of DFT, not a single reference system of non-interacting electrons but several reference systems of non-interacting electrons are used in the subsystem formulation of DFT [28]. As a consequence, different components of the total energy are approximated by means of explicit density functionals than in calculations based on the Kohn-Sham framework. In the subsystem formulation of DFT, the approximated components include exchange-correlation energy and small part of the kinetic energy (non-additive kinetic energy). Both local density approximation (LDA) and generalized gradient approximation (GGA) types of functionals for the kinetic energy component have been used/tested [72]. Using LDA for all relevant energy contributions functionals in subsystem formulation of DFT results in a computational method which is entirely parameter-free. In previous computational studies of weakly intermolecular complexes, which focused mainly on interaction energies, this approximation proved to be very good for hydrogen-bonded complexes [128] as well as a number of other complexes formed by atoms or non-polar molecules Ne-Ne, F_2-Ne, N_2-N_2, N_2-Ar, Ar-Ar, CH_4-CH_4, for instance [93]. For a large class of weak intermolecular complexes, however, such as diatomic molecules interacting with benzene [129], benzene dimer [130], C_3H_6-Ar, C_6H_6-Ar, C_6H_6-CH_4, C_6H_6-C_2H_6, C_3H_8-C_3H_8, C_6H_6-C_2H_4, C_6H_6-C_2H_2 [93], LDA leads to unsatisfactory results. As a rule of thumb, LDA fails in obtaining interaction energies if π-systems are involved in the intermolecular interaction [131]. For such system, a particular combination of gradient dependent functionals of the GGA type proposed and tested for the first time in Ref. [129] improves the interaction energies qualitatively. Unfortunately, this approximation worsens the interaction energies in the case of systems for which LDA is adequate. We underline that opposite to the LDA case, the GGA functionals are not defined uniquely. In our choice for GGA functionals, motivated by their properties, the non-additive kinetic energy is approximated using such a GGA functional, which leads to the best associated functional derivative in the case of weakly overlapping pairs of electron densities [72]. As far as the exchange-correlation component is concerned, the chosen approximation is the functional of Perdew and Wang [120, 121], which has the most similar analytic form to the one for the kinetic energy part and satisfies the Lieb-Oxford condition [132].

It is worthwhile to recall that in the original applications of the subsystem formulation of density functional theory to ionic solids, the subsystems corresponded to atoms and the LDA functionals were used together with additional approximations on the symmetry and localization of orbitals for each subsystem [28, 133]. In our adaptation of this formalism to molecular systems, LDA and GGA functionals can be used and no restrictions are made on symmetry or localization of orbitals in each subsystem [31].

The above numerical results concerning applicability of LDA and GGA functionals in the subsystem formulation of DFT leave us, therefore, with a number of questions of practical importance such as: *i*) In which class of systems LDA can be reliably applied to obtain interaction energies? *ii*) In which class of systems GGA can be reliably applied to obtain interaction energies? *iii*) How good are LDA and GGA equilibrium geometries?

LDA applied in Kohn-Sham framework to approximate the exchange-correlation energy is known to lead to rather unsatisfactory interaction energies for weakly bound intermolecular complexes. Therefore, the good performance of LDA applied to both exchange-correlation and non-additive kinetic energy functionals in the subsystem formulation of DFT indicates that errors in the corresponding functionals cancel each other to some extend. This brings up additional intriguing questions of more fundamental nature: *iv*) What are the physical conditions for such cancellation to take place? *v*) How to construct conjoint gradient-dependent approximations to the exchange-correlation- and non-additive kinetic energies assuring that such cancellation is maximal?

Moreover, since the overall accuracy of the interaction energy is determined by the errors in two types of quantities: functionals and their functional derivatives, it is important to asses the quality of these quantities independently for each considered approximation.

In this work, we report the results of numerical analysis addressing some of the above practical issues in detail. To this end, the equilibrium geometries are in focus of our analysis. The quality of this property is determined by the functional derivatives of the approximated density functionals. The practical importance of determining the range of applicability of LDA and GGA are obvious. This work complements the recently reported analysis of the interaction energies [131] calculated at equilibrium geometries obtained from benchmark wavefunction based calculations.

As far as accuracy of the kinetic-energy-functional dependent energy component is concerned, the Kohn-Sham results (LDA and GGA) are also discussed in this work. In the applied computational scheme, any differences between Kohn-Sham and subsystem-based calculations can be attributed to this functional (and its derivative).

For some intermolecular complexes of high symmetry, we reported already the equilibrium geometries derived from subsystem based calculations applying the functionals of the LDA and GGA type. The recent numerical implementation of the formalism makes it possible to study systems with more degrees of freedom such as the ones in the Zhao and Truhlar dataset comprising equilibrium geometries and interaction energies for a group of representative intermolecular complexes [134], obtained by means of high-level wave-function based type of calculations and intended to be used as benchmark. These authors used the same reference data to asses the performance of various approximations to the exchange-correlation energy functional applied within the Kohn-Sham framework.

The complexes in the test set are divided into the following groups [135, 136]:

- *hydrogen bonded* (HB6/04) NH_3-NH_3,HF-HF, H_2O-H_2O, NH_3-H_2O, $HCONH_2$-$HCONH_2$, and HCOOH-HCOOH,

- *dominated by dipolar interactions* (DI6/04): H_2S-H_2S, HCl-HCl, H_2S-HCl, CH_3Cl-HCl, HCN-CH_3SH, and CH_3SH-HCl,

- *weakly bonded* (WI9/04): He-Ne, He-Ar, Ne-Ne, Ne-Ar, CH_4-Ne, C_6H_6-Ne, CH_4-CH_4, C_2H_2-C_2H_2, and C_2H_4-C_2H_4.

It is worthwhile to underline that the strength of intermolecular interactions varies in a wide range (up to about 16 kcal/mol).

The numerical differences between our results and that in the compared database can be attributed to three factors: *i*) the used basis sets, *ii*) numerical procedures, and *iii*) the approximations to the relevant density functionals. The errors due to the first two factors can be easily controlled and reduced in our implementation of the formalism. The effect of using approximated functionals instead of the exact ones requires, however, dedicated studies on the case by case basis such as the ones reported in the present work.

6.2 Methods

6.2.1 Energy minimization

In this work the two subsystems are treated on equal footing within the subsystem formulation of Density-Functional Theory, solving the set of coupled one-electron equations (Eq. 2.11-2.12) to construct the densities ρ_A and ρ_B. In such a case, the local minimum at the Born-Oppenheimer potential energy surface corresponds to a minimum of the functional $\Xi^S[\{\phi_i^A\},\{\phi_i^B\}]$ with respect to several independent quantities: positions of nuclei in each subsystem $\{\mathbf{R}_A\}$ and $\{\mathbf{R}_B\}$ (geometrical degrees of freedom) and two electron densities ρ_A and ρ_B (electronic-structure related degrees of freedom). The electronic energy in the Born-Oppenheimer approximation corresponds to the numerical value of the Hohenberg-Kohn total energy functional $E^{HK}[\rho_A^o+\rho_B^o]$. All quantities needed to evaluate $E^{HK}[\rho_A^o+\rho_B^o]$ and its gradients with respect to nuclear positions are available at the end of the *freeze-and-thaw* procedure (schemes $A-C$ in Table 6.1).

If, however, only one component of the electron density (say ρ_A), is subject to optimization (Eq. 2.1) whereas ρ_B is not (schemes $D-E$ in Table 6.1), Eqs. 2.8 provides the upper bound for the electronic energy in the Born-Oppenheimer approximation:

$$E^{HK}[\rho_A^o(B)+\rho_B] \geq E^{HK}[\rho_A^o+\rho_B^o] = E^{HK}[\rho_{AB}^o] \tag{6.1}$$

where $\rho_A^o(B)$ denotes the electron density obtained from Eq. 2.1, in which a given ρ_B is used. On the virtue of the second Hohenberg-Kohn theorem, the equality is reached only if the assumed ρ_B added to $\rho_A^o(B)$ equals to

Table 6.1: The considered optimization schemes.

Label	Optimized	Frozen	Treatment of ρ_A and ρ_B
A	$\{\mathbf{R}_A\},\ \rho_A,\ \{\mathbf{R}_B\},\ \rho_B$		fully variational [28, 31]
B	$\{\mathbf{R}_A\},\ \rho_A,\ \rho_B$	$\{\mathbf{R}_B\}$	fully variational [28, 31]
C	$\rho_A,\ \rho_B$	$\{\mathbf{R}_A\}\ \{\mathbf{R}_B\}$	fully variational [28, 31]
D	$\{\mathbf{R}_A\},\ \rho_A$	$\{\mathbf{R}_B\},\ \rho_B$	partially variational [4]
E	ρ_A	$\{\mathbf{R}_A\},\ \{\mathbf{R}_B\},\ \rho_B$	partially variational [4]
F	$\{\mathbf{R}_A\},\ \{\mathbf{R}_B\}$	$\rho_A,\ \rho_B$	non-variational (frozen) [137]
G		$\{\mathbf{R}_A\},\ \{\mathbf{R}_B\},\ \rho_A,\ \rho_B$	non-variational (frozen) [137]

the ground-state electron density at this geometry (ρ^o_{AB}). Nevertheless, the orbitals obtained from Eq. 2.1 provide all necessary quantities to evaluate the numerical value of $E^{HK}[\rho^o_A(B) + \rho_B] - E^{HK}[\rho_B]$ and its gradients with respect to the coordinates of the nuclei in the subsystem A. Therefore, it is possible to optimize the geometry of subsystem A with frozen geometrical and electronic degrees of freedom of the subsystem B (scheme D in Table 6.1). For an assumed ρ_B its adequacy can be controlled by comparing the results obtained from schemes D and B (or E and C if the geometry is not the subject of investigation).

We notice also that the Gordon-Kim model [138, 139] (schemes F and G in Table 6.1), represents an extremely simplified optimization scheme, in which changes of ρ_A and ρ_B associated with inter-subsystem degrees of freedom are not taken into account. Such scheme is only applicable in some cases (rare gas dimers, for instance). For molecules, neglecting the complexation induced changes of the electron density is not a universally adequate approximation as reported previously [128, 93].

Our numerical implementation of Eqs. 2.11-2.12 makes it possible to perform the total energy optimization following each of the schemes listed in Table 6.1.

In this work, we focus on the adequacy of the used approximations to $E_{xc}[\rho]$ and $T_s^{nad}[\rho_A, \rho_B]$ functionals for obtaining equilibrium geometries. Therefore, scheme A (full optimization including geometrical and electronic degrees of freedom) is applied. In our previous works, concerning the interaction energies at some representative points at the potential energy surface, scheme C was applied [128, 93, 131]. We perform the search for the local minima in the vicinity of the reference equilibrium structures taken from the dataset of Zhao and Truhlar. The equilibrium geometries are obtained following an efficient minimization procedure, in which the structural and electronic degrees of freedom are optimized simultaneously (*Sequence II* in Table 6.2). For each geometry update either Eq. 2.11 or 2.12 is solved only once. Until the equilibrium energy is reached, the numerical value of $E^{HK}[\rho^o_A(B) + \rho_B]$ (or $E^{HK}[\rho_A + \rho^o_B(A)]$) does not correspond to the electronic energy in the Born-Oppenheimer approximation. Obtaining this energy at intermediate geometries would require performing the *freeze-and-thaw* procedure (see *Sequence I* in Table 6.2). Typically, the *freeze-and-thaw* procedure involves solving the pair of equations 2.11 and 2.12 two or three times. Therefore, *Sequence II* can be expected to reduce the computational effort by about factor five. In this work, we consider also an even more simplified optimization procedure, in which the exchange-correlation and non-additive kinetic energy functionals are linearized in either ρ_A (Eq. 2.11) or in ρ_B (Eq. 2.12). The errors associated with the linearization are small and they disappear by construction at the end of the *freeze-and-thaw* procedure [117] (and also at the end of the geometry optimization procedure). The resulting computational savings depend on the number of iterations in the self-consistent procedure to solve Eq. 2.1.

6.2.2 Approximations for $E_{xc}[\rho]$ and $T_s^{nad}[\rho_A, \rho_B]$

In this work, LDA and GGA density functionals are considered. We will use the labels KSCED LDA and KSCED GGA for the corresponding computational methods to solve Eqs. 2.1-2.12.

In the KSCED LDA calculations, the exchange functional is approximated using the expression for the uniform gas of non-interacting electrons by Dirac [16], the correlation energy is approximated using the Vosko, et al. [17] parametrization (equation [4.4] in Ref. [17] referred frequently as "VWN V") of the Ceperley-Alder [18] reference data for correlation energy in the uniform electron gas, and the non-additive kinetic energy is approximated using the Thomas-Fermi formula for the kinetic energy [9, 10]. Note that the above approximate functionals do not rely on any empirical data.

In the KSCED GGA calculations, the Perdew-Wang (PW91) [120, 121] exchange-correlation functional is used, whereas the non-additive kinetic energy bi-functional $T_s^{nad}[\rho_A, \rho_B]$ is approximated according to the formula: $T_s^{nad}[\rho_A, \rho_B] \approx T_s^{nad(GGA97)}[\rho_A, \rho_B] = T_s^{LC94}[\rho_A + \rho_B] - T_s^{LC94}[\rho_A] - T_s^{LC94}[\rho_B]$, where $T_s^{LC94}[\rho]$ denotes the Lembarki-Chermette [98] functional of the kinetic energy. The $T_s^{nad(GGA97)}[\rho_A, \rho_B]$ was shown to provide a good approximation to the non-additive kinetic energy functional and potential in the case of weakly overlapping densities [71, 72].

6.2.3 Computational details

The calculations are carried out using our numerical implementation of the subsystem formulation of DFT (the program deMon2K-KSCED [113]) based on the program deMon2K [126]. For geometry optimization, the following deMon2K-KSCED options and parameters are applied: 10^{-6} a.u. self-consistent field energy convergence criterion, adaptive grid (TOL=5.0E-07 "GUESS" grid [140]), and the atomic basis set MG3S [141] used within the

Table 6.2: Complete (sequence I) or partial (sequence II) optimization of the total electron density in one update of the coordinates of all atoms in the complex. The self-consistent procedure to solve Eqs. 2.11-2.12 is denoted with *freeze-and-thaw*. The procedure to update of the coordinates in one subsystem using analytic gradients obtained from Eq. 2.11 (for subsystem A) or Eq. 2.12 (for subsystem B) is denoted with *L-BFGS*.

sequence I	sequence II
$\overbrace{\{\mathbf{R}_B^k\},\{\mathbf{R}_A^i\},\ \underbrace{\rho_B \Longleftrightarrow \rho_A}_{\text{freeze−and−thaw}}}^{\text{Eq.9 or 10}}$ $\overset{i=1,N-1}{\Updownarrow \text{until } \mathbf{R}_\mathrm{A} \text{ convergence}}$ $\overbrace{\{\mathbf{R}_B^k\},\{\mathbf{R}_A^i\},\rho_A,\rho_B \Longrightarrow \{\mathbf{R}_A^{i+1}\}}^{\text{L−BFGS}}$	$\overbrace{\{\mathbf{R}_B^k\},\{\mathbf{R}_A^i\},\rho_B \Longrightarrow \rho_A^i}^{\text{Eq.9}}$ $\overset{i=1,N-1}{\Updownarrow \text{until } \mathbf{R}_\mathrm{A} \text{ convergence}}$ $\overbrace{\{\mathbf{R}_B^k\},\{\mathbf{R}_A^i\},\rho_A^i,\rho_B \Longrightarrow \{\mathbf{R}_A^{i+1}\}}^{\text{L−BFGS}}$
$\overbrace{\{\mathbf{R}_B^k\},\{\mathbf{R}_A^N\},\ \underbrace{\rho_A \Longleftrightarrow \rho_B}_{\text{freeze−and−thaw}}}^{\text{Eq.10 or 9}}$ $\overset{k=1,M-1}{\Updownarrow \text{until } \mathbf{R}_\mathrm{B} \text{ convergence}}$ $\overbrace{\{\mathbf{R}_B^k\},\{\mathbf{R}_A^N\},\rho_A,\rho_B \Longrightarrow \{\mathbf{R}_B^{k+1}\}}^{\text{L−BFGS}}$	$\overbrace{\{\mathbf{R}_B^k\},\{\mathbf{R}_A^N\},\rho_A^N \Longrightarrow \rho_B^k}^{\text{Eq.10}}$ $\overset{k=1,M-1}{\Updownarrow \text{until } \mathbf{R}_\mathrm{B} \text{ convergence}}$ $\overbrace{\{\mathbf{R}_B^k\},\{\mathbf{R}_A^N\},\rho_A^N,\rho_B^k \Longrightarrow \{\mathbf{R}_B^{k+1}\}}^{\text{L−BFGS}}$

monomer-centered expansion scheme (KSCED(m) - see below). The MG3S basis set is chosen based on our recently reported analysis [131] of the effect of changing the basis set on the interaction energies. In principle, two types of basis set expansions can be considered for orbitals corresponding to each subsystem: centered on the monomer or centered on the dimer. The corresponding calculations are labeled as KSCED(m) or KSCED(s), respectively, following the convention of Ref. [72]. For a given choice of the atomic basis sets, the KSCED(s) scheme leads to results closer to the complete basis set limit than the KSCED(m) one. However, if the atomic basis sets are sufficiently large, as the ones chosen for these studies, the two schemes lead to very similar interaction energies [131]. For the calculation of the interaction energies on the optimized geometries, the following deMon2K-KSCED program options are applied: 10^{-6} a.u. self-consistent field energy convergence criterion, and the pruned "MEDIUM" (75,302)p [142] grid, and the dimer-centered MG3S basis sets (KSCED(s) type of calculations).

Classical electron-electron interactions (Coulomb) are evaluated using auxiliary fitting functions referred to as GEN-A2*, which are automatically generated for any given orbital basis sets [125]. Further details concerning the formal framework of the applied computational methods and the numerical implementation can be found in Refs. [28, 31, 128].

The energy derivatives with respect to the coordinates of nuclei of the subsystem A are calculated using the deMon2K-KSCED program and passed together with the total energy and the coordinates of the subsystem A to the generic limited-memory quasi-Newton code for unconstrained optimization L-BFGS [143] (Broyden-Fletcher-Goldfarb-Shanno) using a Perl script, which controls the optimization process. The L-BFGS algorithms uses the following two non-default parameters: EPS=1.0E-05 (threshold for the norm of the gradient in [hartree/bohr]), and M=5 (the number of corrections used in the update of the inverse of the Hessian). Such an optimization procedure yields the precision of the intermolecular distances for the given set of molecules in order of 0.01 Å as tested by performing optimization starting from different geometries. The applied optimization procedure is very efficient in localizing the equilibrium intermolecular distance. For all systems discussed in this study, the equilibrium geometries[1] were obtained by performing multiple optimization runs, each starting from a different geometry

---

[1] The cartesian and z-matrix coordinates of optimized complexes are available at: http://pubs.acs.org/doi/suppl/10.1021/ct600367t/suppl_file/ct600367t-file009.pdf

of the complex. In the initial geometry, the original structure from the Zhao-Truhlar database was modified by changing the intermolecular distance (by a few Å) as well as mutual orientation of the monomers. Such a procedure leads to almost identical final geometries (they lie within 0.01 Å). Unfortunately, such a procedure fails to localize the minimum at flat potential energy surfaces, where some degrees of freedom are associated with very small gradients such as that corresponding to a parallel displacement of one monomer in the benzene dimer.

6.3 Results and Discussion

6.3.1 Geometries: LDA

Table 6.3 collects the selected representative parameters describing intermolecular degrees of freedom in the considered complexes derived from KSCED LDA calculations together with the corresponding reference data. The chosen two geometrical parameters are the intermolecular distance R between the closest two heavy atoms in two molecules forming the complex, and an representative angle describing the relative orientation of the monomers (ϕ). The labels given for the chosen angle allow one to identify it in a straightforward manner. For instance, the HOO denotes the H-O-O angle in the case of the water dimer.

KSCED LDA performs very well for hydrogen bonded complexes. This result is illustrated on Figs. 6.1-6.2 showing the optimized geometry superimposed on the reference one for two representative complexes: H_2O-H_2O, for which KSCED LDA optimized geometry deviates the least from the reference one, and NH_3-NH_3, from which the deviation from the reference is the largest among the hydrogen bonded complexes. The errors of intermolecular distances R are smaller than 0.10 Å for all complexes in this set except for NH_3-NH_3 and HCOOH-HCOOH. The errors in intermolecular distances tend to decrease with increasing binding energy.

In the set of the complexes of dipole character the most important difference with respect to the reference geometry is found for HCl-HCl. In the LDA optimized geometry, two monomers are in a parallel-like orientation and in the reference one they are almost perpendicular.

For the most weakly bound systems, noticeable errors in the intermolecular distance ($0.2 - 0.3$ Å) occur for He-Ne and He-Ar, and C_6H_6-Ne. Most

Table 6.3: The key parameters of the equilibrium geometry obtained from KSCED LDA calculations. R (in Å) denotes distance between the two closest heavy atoms of different monomers, and ϕ (in degrees)is a representative angle determining the relative orientation between the monomers. The reference values R_{ref} and ϕ_{ref} are taken from Ref. [135].

Compound	Def$_R$	R	R_{ref}	$R - R_{ref}$	Def$_\phi$	ϕ	ϕ_{ref}
NH_3-NH_3	dNN	3.14	3.27	-0.13	HNN	16	14
HF-HF	dFF	2.87	2.78	0.08	HFF	113	115
H_2O-H_2O	dOO	2.96	2.94	0.02	HOO	5	4
NH_3-H_2O	dNO	2.97	2.97	0.01	HON	5	6
$HCONH_2$-$HCONH_2$	dNO	2.89	2.88	0.01	ONC	114	116
HCOOH-HCOOH	dOO	2.81	2.70	0.11	OOC	127	125
H_2S-H_2S	dSS	4.03	4.12	-0.09	HSS	90	84
HCl-HCl	dClCl	3.60	3.79	-0.19	HClCl	47	8
H_2S-HCl	dSCl	3.81	3.76	0.05	HClS	92	88
CH_3Cl-HCl	dClCl	3.70	3.61	0.09	ClClC	81	82
HCN-CH_3SH	dCS	3.62	3.52	0.10	SNC	170	162
CH_3SH-HCl	dSCl	3.70	3.61	0.09	HClS	14	11
He-Ne	dHeNe	2.81	3.03	-0.22	-	-	-
He-Ar	dHeAr	3.16	3.48	-0.32	-	-	-
Ne-Ne	dNeNe	3.04	3.09	-0.05	-	-	-
Ne-Ar	dNeAr	3.47	3.49	-0.02	-	-	-
CH_4-Ne	dCNe	3.44	3.49	-0.05	HNeC	70	71
C_6H_6-Ne	dCNe	3.66	3.51	0.15	NeCC	79	79
CH_4-CH_4	dCC	3.61	3.61	-0.00	HCC	70	70
C_2H_2-C_2H_2	dCC	3.22	3.46	-0.23	CCC	122	123
C_2H_4-C_2H_4	dCC	3.83	3.83	0.00	CCC	80	80

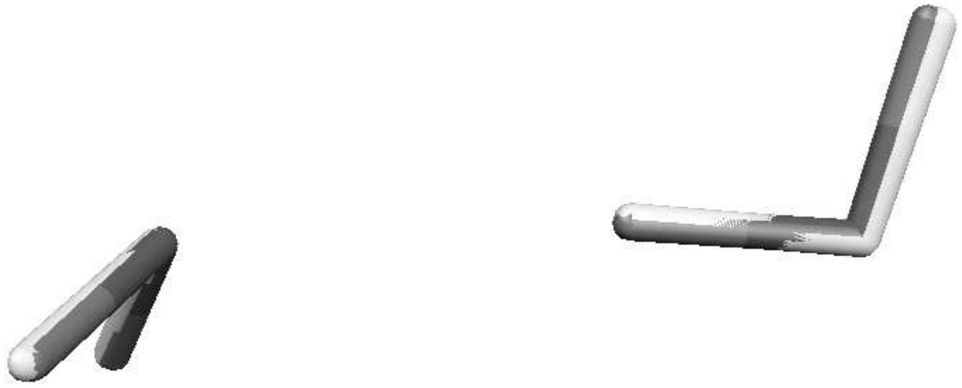


Figure 6.1: The optimized geometry (KSCED LDA) of the H_2O-H_2O complex superimposed on the reference equilibrium geometry (dark colors).

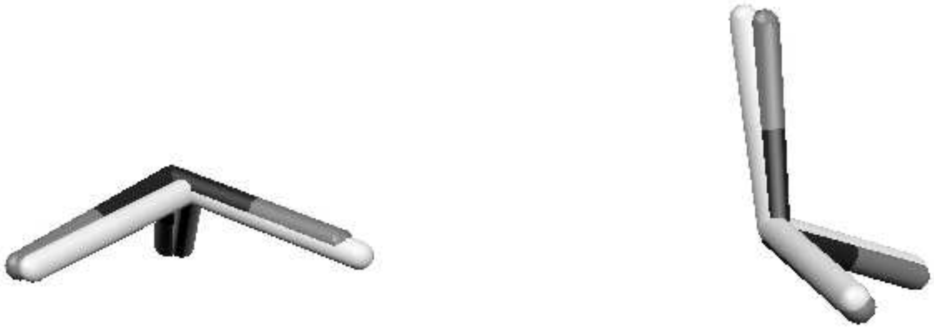


Figure 6.2: The optimized geometry (KSCED LDA) of NH_3-NH_3 superimposed on the reference equilibrium geometry (dark colors).

of the intermolecular equilibrium distances in this set are underestimated, which is an opposite tendency found in the other sets.

Analyzing the overall performance of local density approximation shows that it performs very well for hydrogen bonded complexes, which confirms the results reported elsewhere [128], and reasonably well for weakly bound complexes. In most cases, local density approximation overestimates intermolecular distances, except for the complexes in the W9/04 set (weakly bound complexes).

Local density approximation applied within the Kohn-Sham framework leads systematically to worse results. In the case of all considered intermolecular complexes, the Kohn-Sham LDA calculations lead to underestimated intermolecular equilibrium distances. For hydrogen bonded complexes, the errors reach -0.27 Å for NH_3-NH_3. In the case of dipole bound species, the maximal error occurs for H_2S-H_2S (-0.38 Å). Taking into account that the errors of equilibrium geometries are determined by the quality of the exchange-correlation effective potential whereas the differences between Kohn-Sham LDA and KSCED LDA results are due to the errors in the functional derivatives of the T_s^{nadd}, the superiority of KSCED LDA is the result of the compensation of errors in these quantities. For interaction energies, such compensation were reported previously for several systems [129, 93] (see also below).

6.3.2 Geometries: GGA

KSCED GGA calculations leads to underestimated intermolecular distances for all considered complexes (see Table 6.4).

For hydrogen bonded complexes, the errors in intermolecular distance are larger than the ones in the KSCED LDA case reaching -0.43 Å for NH_3-NH_3. Figs. 6.3-6.4 show the KSCED GGA optimized geometry superimposed on the reference one for two representative complexes: HCOOH-HCOOH, for which KSCED GGA optimized geometry deviates the least from the reference one, and NH_3-NH_3, for which the deviation from the reference is the largest among the hydrogen bonded complexes.

For the dipole-bound complexes, the largest difference with respect to the reference geometry is found for HCl-HCl. As in the case of KSCED LDA equilibrium geometry, the two monomers adopt a parallel-like orientation. In this group of complexes, the errors in the intermolecular distances are rather large reaching 0.44 Å for HCl-HCl.

Table 6.4: The key parameters of the equilibrium geometry obtained from KSCED GGA calculations. R (in Å) denotes distance between the two closest heavy atoms of different monomers, and ϕ (in degrees) is a representative angle determining the relative orientation between the monomers. The reference values R_{ref} and ϕ_{ref} are taken from Ref. [135].

Compound	Def$_R$	R	R_{ref}	$R - R_{ref}$	Def$_\phi$	ϕ	ϕ_{ref}
NH_3-NH_3	dNN	2.83	3.27	-0.43	HNN	40	14
HF-HF	dFF	2.73	2.78	-0.05	HFF	103	115
H_2O-H_2O	dOO	2.84	2.94	-0.10	HOO	5	4
NH_3-H_2O	dNO	2.86	2.97	-0.11	HON	6	6
$HCONH_2$-$HCONH_2$	dNO	2.77	2.88	-0.11	ONC	115	116
HCOOH-HCOOH	dOO	2.68	2.70	-0.02	OOC	128	125
H_2S-H_2S	dSS	3.84	4.12	-0.27	HSS	94	84
HCl-HCl	dClCl	3.35	3.79	-0.44	HClCl	46	8
H_2S-HCl	dSCl	3.62	3.76	-0.14	HClS	87	88
CH_3Cl-HCl	dClCl	3.49	3.61	-0.12	ClClC	78	82
HCN-CH_3SH	dCS	3.21	3.52	-0.31	SNC	146	162
CH_3SH-HCl	dSCl	3.47	3.61	-0.14	HClS	21	11
He-Ne	dHeNe	2.55	3.03	-0.48	-	-	-
He-Ar	dHeAr	2.97	3.48	-0.51	-	-	-
Ne-Ne	dNeNe	2.73	3.09	-0.36	-	-	-
Ne-Ar	dNeAr	3.22	3.49	-0.27	-	-	-
CH_4-Ne	dCNe	3.15	3.49	-0.34	HNeC	70	71
C_6H_6-Ne	dCNe	3.39	3.51	-0.12	NeCC	78	79
CH_4-CH_4	dCC	3.29	3.61	-0.33	HCC	70	70
C_2H_2-C_2H_2	dCC	2.95	3.46	-0.51	CCC	120	123
C_2H_4-C_2H_4	dCC	3.52	3.83	-0.31	CCC	79	80

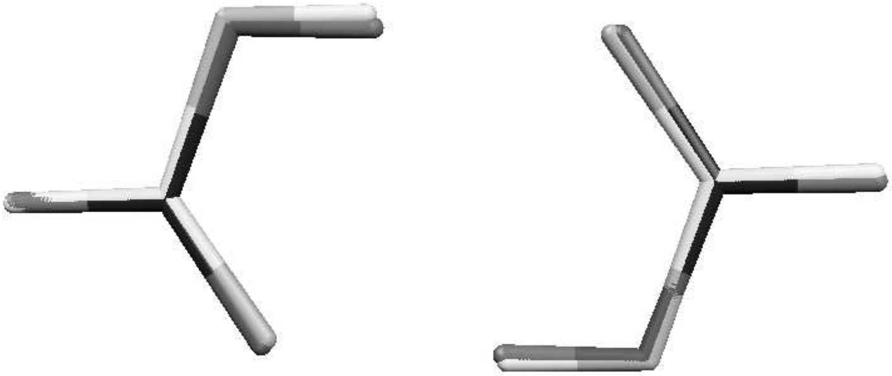


Figure 6.3: The optimized geometry (KSCED GGA) of the HCOOH-HCOOH complex superimposed on the reference equilibrium geometry (dark colors).

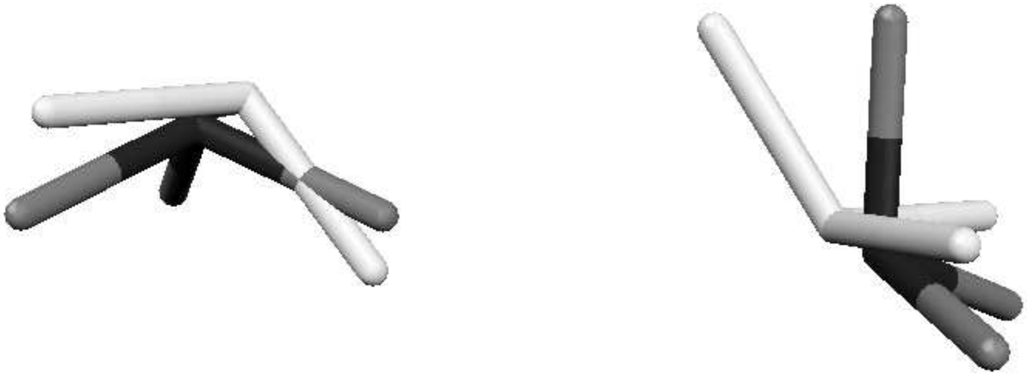


Figure 6.4: The optimized geometry (KSCED GGA) of NH_3-NH_3 superimposed on the reference equilibrium geometry (dark colors).

In the last group of complexes (weakly bound), the errors of the KSCED GGA equilibrium intermolecular distances are very large reaching 0.51 Å.

In view of the fact that the chosen GGA functional worsens significantly the equilibrium geometry for the complexes, for which KSCED LDA leads to rather good results, this approximation does not represent an universal improvement over LDA. Since, however, it leads to significantly better binding energies for (π-stacked systems [93, 129, 130]), it can be considered as a pragmatic choice for this type of complexes.

As far as Kohn-Sham calculations are concerned, the PW91 results are significantly and systematically better the LDA ones. For instance, the errors in the PW91 equilibrium intermolecular distances do not exceed 0.1 Å, whereas the LDA once reach -0.27 Å for hydrogen bonded complexes. Opposite to the tendencies discussed previously for LDA, KSCED GGA are not better than Kohn-Sham PW91 ones. The Kohn-Sham equilibrium geometries are slightly (hydrogen bonded complexes) or noticeably (dipole bound and van der Waals) better quality then the KSCED GGA results. This indicates that the error in the PW91 exchange-correlation potential is not compensated so well with the error in the GGA97 non-additive kinetic energy potential as it is the case of LDA.

6.3.3 Binding energies at optimized geometries

Binding energies discussed in this section are calculated at the optimized geometries and are obtained using the dimer centered expansion of the electron density of each subsystem (KSCED(s) type of calculations). The basis set superposition error and the errors resulting from the superposition of numerical grids are taken into account following the procedure of Ref. [131], which is also given in Supporting Information. We start the analysis with the LDA results. For most of the considered hydrogen bonded complexes, the binding energies are very good. For dipole-bound complexes, the errors in the binding energy are larger. The maximal relative overestimation of the binding energy for HCN-CH_3SH reaches 30%, whereas the binding energy in CH_3Cl-HCl is underestimated by 18%. For van der Waals complexes, KSCED LDA does not perform uniformly. The interactions of helium with other atoms is overestimated significantly. The accuracy of the KSCED LDA binding energies changes from excellent to mediocre along the series, Ne-Ne, Ne-Ar, Ne-CH_4, Ne-C_6H_6. For complexes involving saturated hydrocarbons, KSCED LDA performs reasonably well underestimating, however, the binding energy.

Results collected in Table 6.5 indicate clearly that the presence of a multiple bond of one molecule in the vicinity of the other molecule in the complex, leads systematically to significant errors in binding energies calculated at the KSCED LDA level. Except for C_2H_2-C_2H_2, they are underestimated by about factor two. This trend is in line with that for interaction energies calculated at reference intermolecular geometries for the same [131] or other complexes involving conjugated π systems [93, 129, 130].

The choice of the GGA functionals (exchange-correlation- and non-additive kinetic energies) used in this work was shown previously to lead to significant improvements of accuracy of the interaction energies in the cases where KSCED LDA fails: complexes between diatomic molecules and benzene [129], benzene dimer [130], and other complexes involving interactions with π bonds [93]. Results collected in Table 6.5 show that this choice of gradient-dependent functionals for exchange-correlation- and non-additive kinetic energies significantly worsens this quantity for all types of complexes considered in this work.

As far as the compensation of errors in the exchange-correlation- and non-additive kinetic energies are concerned, similar trends (for LDA, the compensation of errors in the energies occurs systematically whereas it is less systematic in the GGA case) can be seen as those for the accuracy of the effective potentials discussed in the previous section. The binding energies derived from Kohn-Sham LDA calculations are significantly worse than their KSCED LDA counterparts. For all the considered complexes, the deviations from the reference data are rather large, reaching 7.75 kcal/mol for $HCONH_2$-$HCONH_2$ and 12.01 kcal/mol for HCOOH-HCOOH (the corresponding KSCED LDA errors are 0.08 kcal/mol and 1.72 kcal/mol) for instance. As far as GGA is concerned such systematic trends cannot be identified. On the average, the Kohn-Sham PW91 binding energies are better than KSCED GGA.

6.3.4 Acceleration techniques for geometry optimization

In this section, we consider two optimization schemes (see Table 6.2), as well as their two modifications: *i*) *Sequence I'*, in which the converged *freeze-and-thaw* cycle is replaced by solving the pair or Eqs. 2.11 and 2.12 only once per geometry update, and *ii*) a modified *Sequence II*, in which the functional

Table 6.5: Binding energies ($-\Delta E$ in kcal/mol) calculated at optimized geometries. The relative error ($\frac{\Delta E - \Delta E^{Ref}}{\Delta E^{Ref}} \times 100\%$) is given in parenthesis.

Compound	$-\Delta E^{LDA}$		$-\Delta E^{GGA}$		$-\Delta E^{Ref}$
NH_3-NH_3	3.99	(27)	5.59	(77)	3.15
HF-HF	4.12	(-10)	5.00	(9)	4.57
H_2O-H_2O	4.97	(0)	5.94	(20)	4.97
NH_3-H_2O	6.72	(5)	8.03	(25)	6.41
$HCONH_2$-$HCONH_2$	15.03	(1)	17.85	(19)	14.94
HCOOH-HCOOH	14.43	(-11)	17.47	(8)	16.15
H_2S-H_2S	2.12	(28)	2.76	(66)	1.66
HCl-HCl	2.18	(8)	3.19	(59)	2.01
H_2S-HCl	3.44	(3)	4.34	(30)	3.35
CH_3Cl-HCl	2.89	(-19)	4.05	(14)	3.55
HCN-CH_3SH	4.68	(30)	5.73	(60)	3.59
CH_3SH-HCl	4.40	(6)	5.74	(38)	4.16
He-Ne	0.09	(125)	0.46	(1050)	0.04
He-Ar	0.15	(150)	0.47	(683)	0.06
Ne-Ne	0.08	(0)	0.52	(550)	0.08
Ne-Ar	0.12	(-8)	0.54	(315)	0.13
CH_4-Ne	0.15	(-32)	0.66	(200)	0.22
C_6H_6-Ne	0.23	(-51)	0.95	(102)	0.47
CH_4-CH_4	0.44	(-14)	1.03	(102)	0.51
C_2H_2-C_2H_2	1.84	(37)	2.72	(103)	1.34
C_2H_4-C_2H_4	1.06	(-25)	2.00	(41)	1.42

Table 6.6: The key parameters of the equilibrium geometry obtained using four optimization procedures considered in the text. Coordinates of only one molecule (A) in the complex (A-B) are optimized^a.

Compound/parameter		*Sequence I*	*Sequence I'*	*Sequence II*	*Sequence II linearized*
H_2O-H_2O	dOO	2.95	2.95	3.02	3.00
	HOO	6	6	6	6
HCl-CH_3SH	dSCl	3.66	3.69	3.76	3.77
	HClS	16	15	19	19

a) Starting parameters: dOO: 3.38 Å, HOO: 60°, dSCl: 2.65 Å, HClS: 108°

$E_{xc}[\rho_A + \rho_B] - E_{xc}[\rho_A] + T_s^{nad}[\rho_A, \rho_B]$ is linearized in ρ_A in the procedure to solve Eq. 2.11 in order to accelerate it (for Eq. 2.12, $E_{xc}[\rho_A + \rho_B] - E_{xc}[\rho_B] + T_s^{nad}[\rho_A, \rho_B]$ is linearized in ρ_B). By construction, linearization is exact at the end of the *freeze-and-thaw* cycle. In view of the fact that linearization might lead to noticeable savings in the computer time, it is worthwhile to investigate the effect of linearization applied without converging *freeze-and-thaw* cycle in the intermediate stages. For two intermolecular complexes (H_2O-H_2O, HCl-CH_3SH), performance of the four alternative optimization procedures is analyzed in either the complete or partial optimization of geometry. In the partial optimization case, the geometry of one monomer is optimized keeping the geometry of the other frozen (Scheme *B* in Table 6.1). In the geometry optimization, the same convergence criteria and the other optimization parameters as described in the previous section, and local density approximation are applied. The starting geometries were prepared by modifying the coordinates taken from Ref. [134] for one molecule in the complex (the selected intermolecular degrees of freedom are given in Table 6.6).

In the complete optimization calculations, all four optimization schemes lead to equivalent results. The Cartesian coordinates of corresponding atoms in optimized structures differ by less than 0.01 Å. This scatter of the minimized geometries corresponds to the precision of the optimization procedure itself. The key geometrical parameters obtained in the partial optimization (rigid geometry of one monomer) are collected in Table 6.6. All simplified schemes lead to the optimized geometries, which do not differ significantly from that derived using the Born-Oppenheimer surface type of optimization (*Scheme I* in Table 6.2).

The computational costs of the four considered optimization schemes differ significantly. In the case of the H_2O-H_2O dimer, the most expensive one (*Sequence I*) involves 88 geometry updates to converge the coordinates of the first subsystem, and solving the pair of Eqs. 2.11-2.12 two to three times per geometry update. *Sequence I'* converges after 91 geometry updates, however, the pair of Eqs. 2.11-2.12 is solved only once per geometry update. In *Sequence II*, the number of geometry updates increases to 117, but involves solving equation 2.11 only once per geometry update. The most effective among the studied optimization schemes is the one in which *Scheme II* is used and Eq. 2.11 are solved using linearized functionals. Linearization results in an additional reduction of the time of computations by 20-25% per geometry update. Optimization of geometry using this scheme involves 71 geometry updates.

6.4 Conclusions

The current study concerns the applicability of the subsystem formulation of density functional for studies of equilibrium geometries and binding energies in weakly bound intermolecular complexes. Two types of approximations are considered: *i*) local density approximation, which was shown in our previous studies to lead to good binding energies in hydrogen bonded systems [128], and several weakly bound complexes [93, 131] but fails for π-stacked complexes [129, 93, 131], and *ii*) our choice for gradient dependent approximation which improves significantly the interaction energies in π-stacked complexes [129, 93, 131]. In the present work, a systematic analysis of the accuracy of equilibrium geometries is made, complementing thus the previously obtained results concerning interaction energies and equilibrium geometries in complexes of high-symmetry.

Concerning the applicability of local density approximation in the subsystem formulation of DFT in deriving not only intermolecular energies but also equilibrium geometries, the present work confirms the adequacy of this approximation for hydrogen bonded complexes (the largest deviation between calculated and reference intermolecular distance amounts to 0.13 Å for NH_3-NH_3), group of dipole-bound complexes (the largest deviation between calculated and reference intermolecular distance amounts to 0.19 Å for HCl-HCl, for which also the relative orientation is the worst), and even very week intermolecular complexes involving Ne, Ar, and saturated hydrocarbons. In this group of complexes, the maximal relative errors in the binding energies reach 30% but they are significantly smaller in most cases. Using local density approximation in the subsystem formulation of DFT is, however, not recommended if the target of calculations is both the binding energy and equilibrium geometry in complexes involving molecules with conjugated bonds (benzene, ethylene). This trend is in line with our previously reported results concerning π stacked complexes. For the weakest bonds involving He, local density approximation leads to the parameters of the potential energy surface, of only qualitative value (binding energies are overestimated by factor two or three in He-Ne and He-Ar dimers, respectively, whereas the equilibrium distances are too short by 0.2-0.3 Å).

As far as the chosen gradient dependent approximation is concerned, it improves neither binding energies nor equilibrium geometries in the group of complexes for which local density approximation is adequate (hydrogen bonds, dipole-bonds, weak complexes involving, Ne, Ar, or saturated hydrocarbons). Its range of applicability seems, therefore, to be limited to π stacked systems determined in our previously reported studies.

Owing to the mathematical structure of the subsystem formulation of density functional theory, performing simultaneous optimization of different types of degrees of freedom (electron density and nuclear coordinates in each subsystem) is straightforward. An efficient optimization scheme is proposed, in which the system reaches the minimum on the Born-Oppenheimer surface only at the end of the procedure reducing thus the computational efforts in the intermediate geometries.

This work represents an intermediate step toward development of first-principles based multi-level simulation techniques for studying electronic structure in condensed matter systems. The orbital-free embedding formalism uses functionals and potentials defined in the Kohn-Sham formulation of density functional theory. However, they are applied for another basic descriptors of the whole investigated system: the embedded orbitals for one part and electron density only for another one. In the present work, we explore the applicability of the simplest approximation - local density approximation - to derive energetic and geometrical properties of weakly interacting systems. Confirming our previous analyzes based only on energies [128, 131], we can conclude that this approximation is especially suitable for deriving key parameters of the potential energy surface for such cases where the subsystems are linked by hydrogen bonds and no direct interactions involving π-electrons occur. The accuracy of the effective potential determines directly the quality of electronic properties obtained from orbital-free embedding calculations but also that of equilibrium geometries analyzed in this work. This brings us to the most important result as far as practical aspects are concerned that KSCED LDA provides a complete formalism, in which both electronic properties as well as potential energy surface are adequately described in a large class of embedded systems identified in this work. In this class, the balance of approximate terms is such, that the errors of the exchange-correlation- and non-additive kinetic energy functionals cancel to large extend. Practical applications of KSCED LDA framework in multi-scale numerical studies are currently curried out in our group. As far as going beyond LDA is concerned, the current study indicates clearly that the GGA functionals chosen based on our previous studies provide only a temporary solution. Although for systems involving interacting π-electrons this choice improves upon LDA, the improvement is not uniform for other systems. In particular, for hydrogen bonded complexes GGA results are even worse that the LDA ones. Development of a consistent GGA approximation retaining the strengths of LDA and providing an efficient compensation of errors in gradient-dependent terms is an objective of our current studies.

Appendix A: Numerical Implementation of the Subsystem Formulation of Density Functional Theory

The numerical implementation of the subsystem formulation of density functional theory in its Kohn-Sham with Constrained Electron Density (KSCED) form into the deMon2k_KSCED program is presented. Currently the self-consistent field (SCF) procedure, the SCF energy, and the analytical nuclear gradients of the SCF energy are available.

The formulas are given in atomic units for the spin-compensated case. The program flowchart shows the default VXCTYPE BASIS route for the self-consistent field orbital-free embedding calculations.

Symbols and definitions

The electron density ρ^A of the subsystem A (associated here with so called ACTIVE subsystem, what means the subsystem for which a set of KSCED equations (see section 6.4) is solved) is given by

$$\rho^A(\mathbf{r}) = \sum_{\mu,\nu} (\underline{\underline{P}}^A)_{\mu\nu} \mu^A(\mathbf{r}) \nu^A(\mathbf{r}), \tag{6.2}$$

where $\underline{\underline{P}}^A$ is the density matrix of the subsystem A:

$$(\underline{\underline{P}}^A)_{\mu\nu} = 2 \sum_{i}^{occ(A)} c^A_{\mu i} c^A_{\nu i}, \tag{6.3}$$

μ^A is the contracted atomic Gaussian orbital of the subsystem A, and $c^A_{\mu i}$ is the molecular orbital coefficient of the subsystem A. The approximated electron density $\tilde{\rho}^A$ of the subsystem A is given by

$$\tilde{\rho}^A(\mathbf{r}) = \sum_{\bar{k}=1}^{\bar{m}_A} x^A_{\bar{k}} \bar{k}^A(\mathbf{r}), \tag{6.4}$$

where $x^A_{\bar{k}}$ is the density fit coefficient of the subsystem A, and $\bar{k}^A$ is primitive Hermite Gaussians function of the subsystem A. The primitive Hermite Gaussians function (without normalization factor) at atom K with exponent ζ_k has the form:

$$\bar{k}(\mathbf{r}) = \left(\frac{\partial}{\partial K_x}\right)^{\bar{k}_x} \left(\frac{\partial}{\partial K_y}\right)^{\bar{k}_y} \left(\frac{\partial}{\partial K_z}\right)^{\bar{k}_z} e^{-\zeta_k(\mathbf{r}-\mathbf{K})^2} \tag{6.5}$$

Analogously for the subsystem B (associated here with so called FROZEN subsystem, what means the subsystem the electron density of which is kept frozen when a set of KSCED equations is solved) we define:

$$\rho^B(\mathbf{r}) = \sum_{\sigma,\tau} (\underline{\underline{P}}^B)_{\sigma\tau} \sigma^B(\mathbf{r}) \tau^B(\mathbf{r}), \tag{6.6}$$

where $\underline{\underline{P}}^B$ is the density matrix of the subsystem B:

$$(\underline{\underline{P}}^B)_{\sigma\tau} = 2 \sum_{i}^{occ(B)} c^B_{\sigma i} c^B_{\tau i}, \tag{6.7}$$

σ^B is the contracted atomic Gaussian orbital of the subsystem B, and $c^B_{\sigma i}$ is the molecular orbital coefficient of the subsystem B. The approximated electron density $\tilde{\rho}^B$ of the subsystem B is given by

$$\tilde{\rho}^B(\mathbf{r}) = \sum_{\bar{l}=1}^{\bar{m}_B} x^B_{\bar{l}} \bar{l}^B(\mathbf{r}), \tag{6.8}$$

where $x^B_{\bar{l}}$ is the density fit coefficient of the subsystem B, and $\bar{l}^B$ is primitive Hermite Gaussians function of the subsystem B. The number of electrons N^A in the subsystem A is given by

$$N^A = \sum_{\mu,\nu} (\underline{\underline{P}}^A)_{\mu\nu} (\underline{\underline{S}}^A)_{\mu\nu}, \tag{6.9}$$

where $(\underline{\underline{S}}^A)_{\mu\nu}$ is the element of the overlap matrix between the contracted atomic Gaussian orbitals of the subsystem A. The number of electrons N^B in the subsystem B is given by

$$N^B = \sum_{\sigma,\tau} (\underline{\underline{P}}^B)_{\sigma\tau} (\underline{\underline{S}}^B)_{\sigma\tau}, \tag{6.10}$$

where $(\underline{\underline{S}}^B)_{\sigma\tau}$ is the element of the overlap matrix between the contracted atomic Gaussian orbital of the subsystem B. The "core" Hamiltonian $\underline{\underline{H}}^A$ of the subsystem A is defined as:

$$(\underline{\underline{H}}^A)_{\mu\nu} = \langle \mu^A | -\frac{1}{2}\nabla^2 | \nu^A \rangle + \langle \mu^A | -\sum_{\alpha=1}^{N^A_{nuc}} \frac{Z^A_\alpha}{|\mathbf{r} - \mathbf{R}^A_\alpha|} | \nu^A \rangle \tag{6.11}$$

The "core" Hamiltonian $\underline{\underline{H}}^B$ of the subsystem B is defined as:

$$(\underline{\underline{H}}^B)_{\sigma\tau} = \langle \sigma^B | -\frac{1}{2}\nabla^2 | \tau^B \rangle + \langle \sigma^B | -\sum_{\beta=1}^{N^B_{nuc}} \frac{Z^B_\beta}{|\mathbf{r} - \mathbf{R}^B_\beta|} | \tau^B \rangle \tag{6.12}$$

The total nuclear repulsion energy of the system is defined as:

$$E_{N^A_{nuc} N^B_{nuc}} = \sum_{\alpha=1}^{N^A_{nuc}} \sum_{\beta=1}^{N^B_{nuc}} \frac{Z^A_\alpha Z^B_\beta}{|\mathbf{R}^A_\alpha - \mathbf{R}^B_\beta|} \tag{6.13}$$

The nuclear repulsion energy of the subsystem A is defined as:

$$E_{N^A_{nuc}N^A_{nuc}} = \frac{1}{2}\sum_{\alpha_1=1}^{N^A_{nuc}}\sum_{\alpha_2\neq\alpha_1}^{N^A_{nuc}} \frac{Z^A_{\alpha_1}Z^A_{\alpha_2}}{|\mathbf{R}^A_{\alpha_1}-\mathbf{R}^A_{\alpha_2}|} \tag{6.14}$$

The nuclear repulsion energy of the subsystem B is defined as:

$$E_{N^B_{nuc}N^B_{nuc}} = \frac{1}{2}\sum_{\beta_1=1}^{N^B_{nuc}}\sum_{\beta_2\neq\beta_1}^{N^B_{nuc}} \frac{Z^B_{\beta_1}Z^B_{\beta_2}}{|\mathbf{R}^B_{\beta_1}-\mathbf{R}^B_{\beta_2}|} \tag{6.15}$$

Total energy bifunctional

Total energy bifunctional (for the ground state noninteracting pure state v-representable electron densities ρ^A and ρ^B): reads [4, 72]:

$$\begin{aligned}
E[\rho^A,\rho^B] &= \sum_{\mu,\nu}(\underline{\underline{P}}^A)_{\mu\nu}\,(\underline{\underline{H}}^A)_{\mu\nu} + \frac{1}{2}\langle\rho^A\|\rho^A\rangle + E_{N^A_{nuc}N^A_{nuc}} \\
&+ \sum_{\sigma,\tau}(\underline{\underline{P}}^B)_{\sigma\tau}\,(\underline{\underline{H}}^B)_{\sigma\tau} + \frac{1}{2}\langle\rho^B\|\rho^B\rangle + E_{N^B_{nuc}N^B_{nuc}} \\
&+ \sum_{\mu,\nu}(\underline{\underline{P}}^A)_{\mu\nu}\langle\mu^A| - \sum_{\beta=1}^{N^B_{nuc}}\frac{Z^B_\beta}{|\mathbf{r}-\mathbf{R}^B_\beta|}|\nu^A\rangle \\
&+ \sum_{\sigma,\tau}(\underline{\underline{P}}^B)_{\sigma\tau}\langle\sigma^B| - \sum_{\alpha=1}^{N^A_{nuc}}\frac{Z^A_\alpha}{|\mathbf{r}-\mathbf{R}^A_\alpha|}|\tau^B\rangle \\
&+ \langle\rho^A\|\rho^B\rangle + E_{N^A_{nuc}N^B_{nuc}} \\
&+ E_{xc}[\rho^A+\rho^B] + T_s^{nad}[\rho^A,\rho^B],
\end{aligned} \tag{6.16}$$

where $T_s^{nad}[\rho^A,\rho^B] = T_s[\rho^A+\rho^B] - T_s[\rho^A] - T_s[\rho^B]$, and the symbol $\|$ stands for the Coulomb operator $\frac{1}{|\mathbf{r}_1-\mathbf{r}_2|}$. With the density fitting procedure [144] the following total energy bifunctional expression is obtained:

$$\begin{aligned}
E[\rho^A,\rho^B] &= \sum_{\mu,\nu}(\underline{\underline{P}}^A)_{\mu\nu}\,(\underline{\underline{H}}^A)_{\mu\nu} + \langle\rho^A\|\tilde{\rho}^A\rangle - \frac{1}{2}\langle\tilde{\rho}^A\|\tilde{\rho}^A\rangle + E_{N^A_{nuc}N^A_{nuc}} \\
&+ \sum_{\sigma,\tau}(\underline{\underline{P}}^B)_{\sigma\tau}\,(\underline{\underline{H}}^B)_{\sigma\tau} + \langle\rho^B\|\tilde{\rho}^B\rangle - \frac{1}{2}\langle\tilde{\rho}^B\|\tilde{\rho}^B\rangle + E_{N^B_{nuc}N^B_{nuc}} \\
&+ \sum_{\mu,\nu}(\underline{\underline{P}}^A)_{\mu\nu}\langle\mu^A| - \sum_{\beta=1}^{N^B_{nuc}}\frac{Z^B_\beta}{|\mathbf{r}-\mathbf{R}^B_\beta|}|\nu^A\rangle
\end{aligned}$$

$$
\begin{aligned}
&+ \sum_{\sigma,\tau}(\underline{\underline{P}}^{B})_{\sigma\tau}\langle\sigma^{B}| - \sum_{\alpha=1}^{N_{nuc}^{A}}\frac{Z_{\alpha}^{A}}{|\mathbf{r}-\mathbf{R}_{\alpha}^{A}|}|\tau^{B}\rangle \\
&+ \langle\rho^{A}\|\tilde{\rho}^{B}\rangle + \langle\tilde{\rho}^{A}\|\rho^{B}\rangle - \langle\tilde{\rho}^{A}\|\tilde{\rho}^{B}\rangle + E_{N_{nuc}^{A}N_{nuc}^{B}} \\
&+ E_{xc}[\rho^{A}+\rho^{B}] + T_{s}^{nad}[\rho^{A},\rho^{B}] \\
&- \lambda(\int\tilde{\rho}^{A}(\mathbf{r})d\,\mathbf{r} - N^{A}) \qquad (6.17)
\end{aligned}
$$

where $\int\tilde{\rho}^{A}(\mathbf{r})d\,\mathbf{r} = \sum_{\bar{k}=1}^{\bar{m}_{A}} x_{\bar{k}}^{A}\langle\bar{k}^{A}\rangle$ (with $\langle\bar{k}^{A}\rangle = \int\bar{k}^{A}(\mathbf{r})d\,\mathbf{r}$), and the Lagrange multiplier λ was introduced to ensure the normalization of the approximated ACTIVE electron density $\tilde{\rho}^{A}$ to the number of electrons in the ACTIVE subsystem. The derivation of Eq. 6.17 using the variational fitting of $\rho^{A}+\rho^{B}$ is presented in Ref. [145].

Total energy bifunctional partitioning

Total energy bifunctional can be partitioned into the two components:

$$
E[\rho^{A},\rho^{B}] = EFDFT1[\rho^{A},\rho^{B}] + EFDFT3[\rho^{B}], \qquad (6.18)
$$

where:

$$
\begin{aligned}
EFDFT3[\rho^{B}] &= \sum_{\sigma,\tau}(\underline{\underline{P}}^{B})_{\sigma\tau}\,(\underline{\underline{H}}^{B})_{\sigma\tau} + \langle\rho^{B}\|\tilde{\rho}^{B}\rangle - \frac{1}{2}\langle\tilde{\rho}^{B}\|\tilde{\rho}^{B}\rangle + E_{N_{nuc}^{B}N_{nuc}^{B}} \\
&+ E_{xc}[\rho^{B}], \qquad (6.19)
\end{aligned}
$$

$$
\begin{aligned}
EFDFT1[\rho^{A},\rho^{B}] &= E[\rho^{A},\rho^{B}] - EFDFT3[\rho^{B}] = \\
&\quad EFDFT3[\rho^{A}] - E_{xc}[\rho^{A}] \\
&- E_{xc}[\rho^{B}] + E_{xc}[\rho^{A}+\rho^{B}] + T_{s}^{nad}[\rho^{A},\rho^{B}] + E_{N_{nuc}^{A}N_{nuc}^{B}} \\
&+ \sum_{\mu,\nu}(\underline{\underline{P}}^{A})_{\mu\nu}\langle\mu^{A}| - \sum_{\beta=1}^{N_{nuc}^{B}}\frac{Z_{\beta}^{B}}{|\mathbf{r}-\mathbf{R}_{\beta}^{B}|}|\nu^{A}\rangle \\
&+ \sum_{\sigma,\tau}(\underline{\underline{P}}^{B})_{\sigma\tau}\langle\sigma^{B}| - \sum_{\alpha=1}^{N_{nuc}^{A}}\frac{Z_{\alpha}^{A}}{|\mathbf{r}-\mathbf{R}_{\alpha}^{A}|}|\tau^{B}\rangle \\
&+ \langle\rho^{A}\|\tilde{\rho}^{B}\rangle + \langle\tilde{\rho}^{A}\|\rho^{B}\rangle - \langle\tilde{\rho}^{A}\|\tilde{\rho}^{B}\rangle, \qquad (6.20)
\end{aligned}
$$

$$
\begin{aligned}
EFDFT3[\rho^{A}] &= \sum_{\mu,\nu}(\underline{\underline{P}}^{A})_{\mu\nu}\,(\underline{\underline{H}}^{A})_{\mu\nu} + \langle\rho^{A}\|\tilde{\rho}^{A}\rangle - \frac{1}{2}\langle\tilde{\rho}^{A}\|\tilde{\rho}^{A}\rangle + E_{N_{nuc}^{A}N_{nuc}^{A}} \\
&+ E_{xc}[\rho^{A}] - \lambda(\int\tilde{\rho}^{A}(\mathbf{r})d\,\mathbf{r} - N^{A}). \qquad (6.21)
\end{aligned}
$$

We identify $EFDFT3[\rho^A]$ as the SCF KSCED energy calculated using the ACTIVE density.

KSCED matrix

The derivative of the $EFDFT1[\rho^A, \rho^B]$ (see Eq. 6.20) with respect to the density matrix element $(\underline{\underline{P}}^A)_{\mu\nu}$ defines the ACTIVE KSCED matrix:

$$\begin{aligned}
(\underline{\underline{K}}^A)_{\mu\nu} &= \left(\frac{\partial EFDFT1[\rho^A, \rho^B]}{\partial (\underline{\underline{P}}^A)_{\mu\nu}}\right)_{\{x^A_{\bar{k}}\}} = \\
& \left(\frac{\partial EFDFT3[\rho^A]}{\partial (\underline{\underline{P}}^A)_{\mu\nu}}\right)_{\{x^A_{\bar{k}}\}} - \left(\frac{\partial E_{xc}[\rho^A]}{\partial (\underline{\underline{P}}^A)_{\mu\nu}}\right)_{\{x^A_{\bar{k}}\}} \\
+ & \left(\frac{\partial E_{xc}[\rho^A + \rho^B]}{\partial (\underline{\underline{P}}^A)_{\mu\nu}}\right)_{\{x^A_{\bar{k}}\}} + \left(\frac{\partial T_s^{nad}[\rho^A, \rho^B]}{\partial (\underline{\underline{P}}^A)_{\mu\nu}}\right)_{\{x^A_{\bar{k}}\}} \\
+ & \langle \mu^A | - \sum_{\beta=1}^{N^B_{nuc}} \frac{Z^B_\beta}{|\mathbf{r} - \mathbf{R}^B_\beta|} | \nu^A \rangle + \sum_{\bar{l}=1}^{\bar{m}_B} \langle \mu^A \nu^A \| \bar{l}^B \rangle x^B_{\bar{l}}.
\end{aligned} \qquad (6.22)$$

The element $\mu\nu$ of the KSCED matrix associated with the local energy functional $E_Y[\rho]$ is defined as the derivative of $E_Y[\rho]$ with respect to the density matrix element $(\underline{\underline{P}})_{\mu\nu}$:

$$\frac{\partial E_Y[\rho]}{\partial (\underline{\underline{P}})_{\mu\nu}} = \int \frac{\delta E_Y[\rho]}{\delta \rho(\mathbf{r})} \frac{\partial \rho(\mathbf{r})}{\partial (\underline{\underline{P}})_{\mu\nu}} d\mathbf{r} \equiv \int \nu_Y[\rho] \frac{\partial \rho(\mathbf{r})}{\partial (\underline{\underline{P}})_{\mu\nu}} d\mathbf{r}, \qquad (6.23)$$

where $\nu_Y[\rho]$ is the potential.

When the exchange-correlation and kinetic potentials are calculated using the approximated densities $\tilde{\rho}^A$ and $\tilde{\rho}^B$, the derrivatives of the approximated electron density $\tilde{\rho}^A$ with respect to the elements of the KSCED matrix are needed:

$$\frac{\partial \tilde{\rho}^A}{\partial (\underline{\underline{P}}^A)_{\mu\nu}} = \sum_{\bar{k}=1}^{\bar{m}_A} \frac{\partial x^A_{\bar{k}}}{\partial (\underline{\underline{P}}^A)_{\mu\nu}} \bar{k}^A(\mathbf{r}). \qquad (6.24)$$

Using the explicit expression for the fit coefficients (Eq. 8.2.7) from

Ref. [145] together with the definition 6.9 we obtain:

$$\begin{aligned} x_{\bar{k}}^{A} &= \sum_{\bar{j}=1}^{\bar{m}_A} (\underline{\underline{G}}^{A})^{-1}_{kj} \left(\sum_{\mu,\nu} (\underline{\underline{P}}^{A})_{\mu\nu} \langle \mu^A \nu^A \| \bar{j}^A \rangle + (\underline{J}^{AB})_j \right) \\ &+ (\underline{\underline{G}}^{A})^{-1}_{k(\bar{m}_A+1)} \sum_{\mu,\nu} (\underline{\underline{P}}^{A})_{\mu\nu} (\underline{\underline{S}}^{A})_{\mu\nu}. \end{aligned} \tag{6.25}$$

Employing that the matrix $(\underline{\underline{G}}^{A})^{-1}$ is symmetric we obtain the derrivative of the fit coeffient:

$$\frac{\partial x_{\bar{k}}^{A}}{\partial (\underline{\underline{P}}^{A})_{\mu\nu}} = \sum_{\bar{j}=1}^{\bar{m}_A} \langle \mu^A \nu^A \| \bar{j}^A \rangle (\underline{\underline{G}}^{A})^{-1}_{jk} + (\underline{\underline{S}}^{A})_{\mu\nu} (\underline{\underline{G}}^{A})^{-1}_{(\bar{m}_A+1)k}.$$

This result when first inserted into 6.25 and used in Eq. 6.24 yields to following expression for KSCED matrix element $(\underline{\underline{K}}^{A})_{\mu\nu}$:

$$\begin{aligned} (\tilde{\underline{\underline{K}}}^{A})_{\mu\nu} &= (\underline{\underline{H}}^{A})_{\mu\nu} + (\underline{\underline{S}}^{A})_{\mu\nu} (\lambda + \bar{z}_n^A) + \sum_{\bar{k}=1}^{\bar{m}_A} \langle \mu^A \nu^A \| \bar{k}^A \rangle (x_{\bar{k}}^A + z_{\bar{k}}^A) \\ &+ \langle \mu^A | - \sum_{\beta=1}^{N_{nuc}^B} \frac{Z_\beta^B}{|\mathbf{r} - \mathbf{R}_\beta^B|} | \nu^A \rangle + \sum_{\bar{l}=1}^{\bar{m}_B} \langle \mu^A \nu^A \| \bar{l}^B \rangle x_{\bar{l}}^B , \end{aligned} \tag{6.26}$$

where $z_{\bar{l}}^A$ is the exchange-correlation-kinetic fitting coefficient:

$$z_{\bar{l}}^A = \sum_{\bar{k}=1}^{\bar{m}_A} (\underline{\underline{G}}^{A})^{-1}_{\bar{l}\bar{k}} \langle \bar{k}^A \| \tilde{\nu}_{xcT} \rangle, \tag{6.27}$$

and $\tilde{\nu}_{xcT}$ is the effective embedding exchange-correlation-kinetic potential calculated using the approximated densities $\tilde{\rho}^A$ and $\tilde{\rho}^B$:

$$\begin{aligned} \tilde{\nu}_{xcT}[\tilde{\rho}^A, \tilde{\rho}^B] &= \tilde{\nu}_{xc}[\tilde{\rho}^A] + \tilde{\nu}_{xc}[\tilde{\rho}^A + \tilde{\rho}^B] - \tilde{\nu}_{xc}[\tilde{\rho}^A] + \left. \frac{\delta T_s^{nad}[\rho, \tilde{\rho}^B]}{\delta \rho} \right|_{\rho = \tilde{\rho}^A} \\ &\equiv \left. \frac{\delta E_{xcT}[\rho, \tilde{\rho}^B]}{\delta \rho} \right|_{\rho = \tilde{\rho}^A} = \tilde{\nu}_{xc}[\tilde{\rho}^A] + \tilde{\nu}\prime_{xcT}[\tilde{\rho}^A, \tilde{\rho}^B]. \end{aligned} \tag{6.28}$$

When the exchange-correlation and kinetic potentials are calculated using the “exact” densities ρ^A and ρ^B, the KSCED matrix element $(\underline{\underline{K}}^{A})_{\mu\nu}$ reads:

$$(\underline{\underline{K}}^{A})_{\mu\nu} = (\underline{\underline{H}}^{A})_{\mu\nu} + (\underline{\underline{S}}^{A})_{\mu\nu} \lambda + \sum_{\bar{k}=1}^{\bar{m}_A} \langle \mu^A \nu^A \| \bar{k}^A \rangle x_{\bar{k}}^A + \langle \mu^A | \nu_{xcT} | \nu^A \rangle$$

$$+ \quad \langle\mu^A| - \sum_{\beta=1}^{N_{nuc}^B} \frac{Z_\beta^B}{|\mathbf{r} - \mathbf{R}_\beta^B|}|\nu^A\rangle + \sum_{\bar{l}=1}^{\bar{m}_B}\langle\mu^A\nu^A\|\bar{l}^B\rangle x_{\bar{l}}^B, \tag{6.29}$$

where ν_{xcT} is the effective embedding exchange-correlation-kinetic potential calculated using the "exact" densities ρ^A and ρ^B:

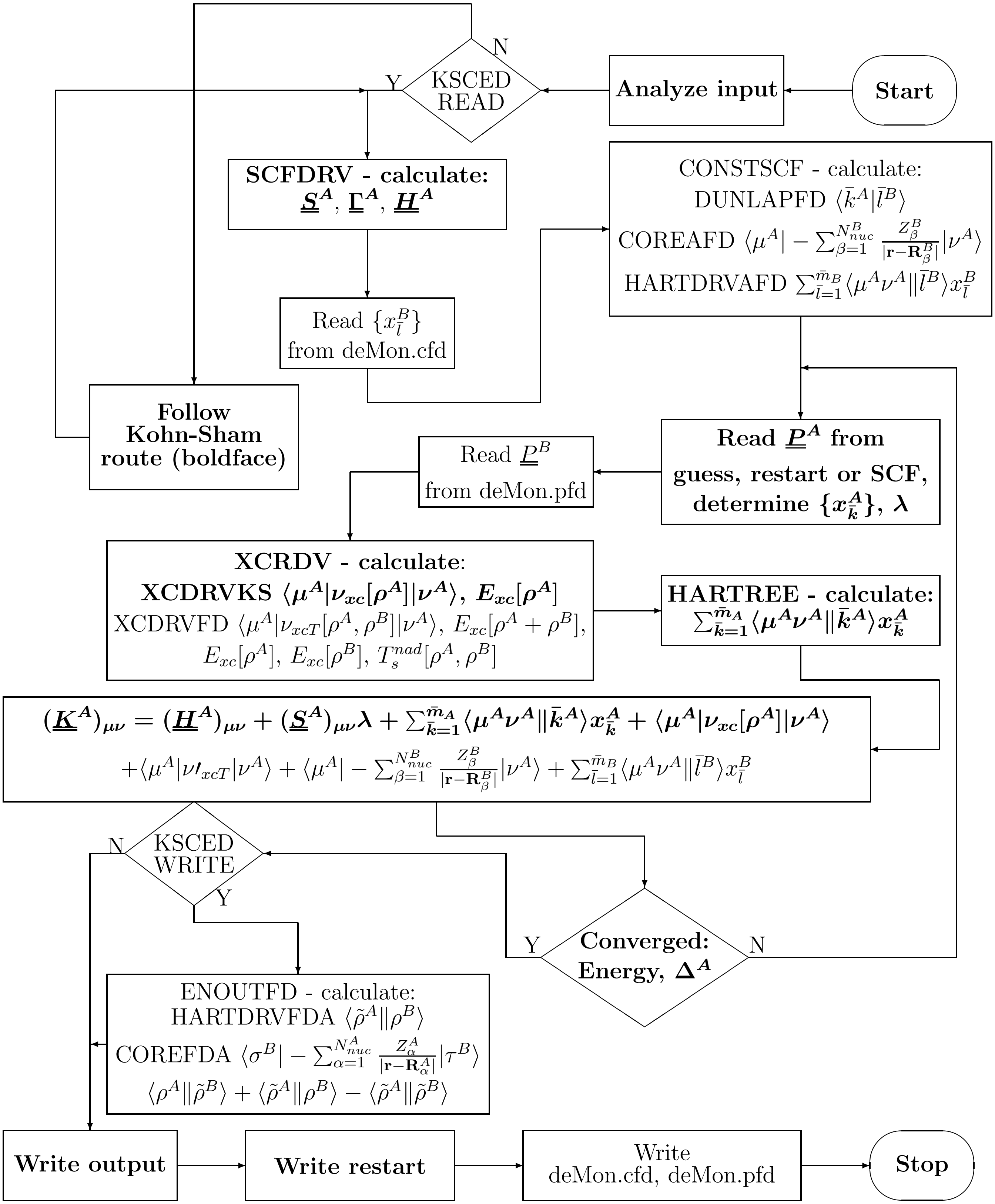
$$\begin{aligned}\nu_{xcT}[\rho^A, \rho^B] &= \nu_{xc}[\rho^A] + \nu_{xc}[\rho^A + \rho^B] - \nu_{xc}[\rho^A] + \left.\frac{\delta T_s^{nad}[\rho, \rho^B]}{\delta\rho}\right|_{\rho=\rho^A} \\ &= \nu_{xc}[\rho^A] + \nu\prime_{xcT}[\rho^A, \rho^B]. \end{aligned} \tag{6.30}$$

It should be noted that during the SCF procedure in deMon an additional criterion (Eq. 26 in Ref. [144]) of convergence is used:

$$\Delta^A = \sum_{\bar{i}=1}^{\bar{m}_A}\left[\sum_{\bar{j}=1}^{\bar{m}_A}\left(x_{\bar{j}}^A(new) - x_{\bar{j}}^A(old)\right)\langle\bar{j}^A\|\bar{i}^A\rangle\right]^2, \tag{6.31}$$

where $x_{\bar{j}}^A(new)$ and $x_{\bar{j}}^A(old)$ denote the fit coefficient of the approximated ACTIVE electron density of two successive SCF iterations.

Program SCF flowchart



Analytical gradient of the energy

In this section the expression for the component of the cartesian gradient of the SCF energy E^X (derivative of the SCF energy with respect to a cartesian coordinate of a nucleus of the ACTIVE system) is given for the default VXCTYPE BASIS route in the deMon2k_KSCED program. To facilitate the notation the superscript ^{X} denotes the partial derivative with respect to a cartesian coordinate of a nucleus of the ACTIVE system.

The derivation of E^X is straightforward. First, the component of the gradient of the exchange-correlation and kinetic energies can be written as:

$$\begin{aligned} E^X_{xcT}[\rho^A,\rho^B] &= \\ &\sum_{\mu,\nu}(\underline{\underline{P}}^A)_{\mu\nu}\langle(\mu^A\nu^A)^X|\nu_{xcT}\rangle + \sum_{\mu,\nu}(\underline{\underline{P}}^{AX})_{\mu\nu}\langle\mu^A|\nu_{xcT}|\nu^A\rangle, \end{aligned} \tag{6.32}$$

owing to the following equation, holding for a local energy functional $E_Y[\rho]$:

$$E^X_Y[\rho] = \int \frac{\delta E_Y[\rho]}{\delta\rho(\mathbf{r})}\rho^X(\mathbf{r})d\mathbf{r} = \int \nu_Y[\rho]\rho^X(\mathbf{r})d\mathbf{r}, \tag{6.33}$$

and the formula for the component the ACTIVE electron density gradient:

$$\rho^{AX}(\mathbf{r}) = \sum_{\mu,\nu}(\underline{\underline{P}}^A)_{\mu\nu}(\mu^A(\mathbf{r})\nu^A(\mathbf{r}))^X + \sum_{\mu,\nu}(\underline{\underline{P}}^{AX})_{\mu\nu}\mu^A(\mathbf{r})\nu^A(\mathbf{r}).$$

Using Eq. 6.32 E^X yields:

$$\begin{aligned} E^X &= EFDFT1^X[\rho^A,\rho^B] + EFDFT3^X[\rho^B] = EFDFT1^X[\rho^A,\rho^B] = \\ &= \sum_{\mu,\nu}(\underline{\underline{P}}^A)_{\mu\nu}\left[(\underline{H}^{AX})_{\mu\nu} + (\underline{S}^{AX})_{\mu\nu}\lambda + \sum_{\bar{k}=1}^{\bar{m}_A}\langle\mu^A\nu^A\|\bar{k}^A\rangle^X x^A_{\bar{k}} + \langle(\mu^A\nu^A)^X|\nu_{xcT}\rangle\right] \\ &+ \sum_{\mu,\nu}(\underline{\underline{P}}^{AX})_{\mu\nu}\left[(\underline{H}^A)_{\mu\nu} + (\underline{S}^A)_{\mu\nu}\lambda + \sum_{\bar{k}=1}^{\bar{m}_A}\langle\mu^A\nu^A\|\bar{k}^A\rangle x^A_{\bar{k}} + \langle\mu^A|\nu_{xcT}|\nu^A\rangle\right. \\ &+ \left.\langle\mu^A| - \sum_{\beta=1}^{N^B_{nuc}}\frac{Z^B_\beta}{|\mathbf{r}-\mathbf{R}^B_\beta|}|\nu^A\rangle + \sum_{\bar{l}=1}^{\bar{m}_B}\langle\mu^A\nu^A\|\bar{l}^B\rangle x^B_{\bar{l}}\right] \\ &- \frac{1}{2}\sum_{\bar{k}=1}^{\bar{m}_A}\sum_{\bar{l}=1}^{\bar{m}_A}\langle\bar{k}^A\|\bar{l}^A\rangle^X x^A_{\bar{k}}x^A_{\bar{l}} + E^X_{N^A_{nuc}N^A_{nuc}} \end{aligned}$$

$$
\begin{aligned}
&+ \sum_{\mu,\nu} (\underline{\underline{P}}^{A})_{\mu\nu} \left[\langle \mu^{A} | - \sum_{\beta=1}^{N_{nuc}^{B}} \frac{Z_{\beta}^{B}}{|\mathbf{r} - \mathbf{R}_{\beta}^{B}|} | \nu^{A} \rangle^{X} + \sum_{\bar{l}=1}^{\bar{m}_{B}} \langle \mu^{A} \nu^{A} \| \bar{l}^{B} \rangle^{X} x_{\bar{l}}^{B} \right] \\
&+ \sum_{\sigma,\tau} (\underline{\underline{P}}^{B})_{\sigma\tau} \left[\langle \sigma^{B} | - \sum_{\alpha=1}^{N_{nuc}^{A}} \frac{Z_{\alpha}^{A}}{|\mathbf{r} - \mathbf{R}_{\alpha}^{A}|} | \tau^{B} \rangle^{X} + \sum_{\bar{k}=1}^{\bar{m}_{A}} \langle \sigma^{B} \tau^{B} \| \bar{k}^{A} \rangle^{X} x_{\bar{k}}^{A} \right] \\
&- \sum_{\bar{k}=1}^{\bar{m}_{A}} \sum_{\bar{l}=1}^{\bar{m}_{B}} \langle \bar{k}^{A} \| \bar{l}^{B} \rangle^{X} x_{\bar{k}}^{A} x_{\bar{l}}^{B} + E_{N_{nuc}^{A} N_{nuc}^{B}}^{X}. \qquad (6.34)
\end{aligned}
$$

The result of Eq. 6.34 can be simplified using the 'Pulay relation' [146] (Eq. 6.37), where the calculation of explicit derivatives of the ACTIVE density matrix elements $(\underline{\underline{P}}^{AX})_{\mu\nu}$ is avoided.

The final expression for the component of the cartesian gradient of the SCF energy E^X reads:

$$
\begin{aligned}
E^{X} &= \sum_{\mu,\nu} (\underline{\underline{P}}^{A})_{\mu\nu} \left[(\underline{\underline{H}}^{AX})_{\mu\nu} + \sum_{\bar{k}=1}^{\bar{m}_{A}} \langle \mu^{A} \nu^{A} \| \bar{k}^{A} \rangle^{X} x_{\bar{k}}^{A} + \langle (\mu^{A} \nu^{A})^{X} | \nu_{xcT} \rangle \right] \\
&- \sum_{\mu,\nu} (\underline{\underline{W}}^{A\lambda})_{\mu\nu} (\underline{\underline{S}}^{AX})_{\mu\nu} - \frac{1}{2} \sum_{\bar{k}=1}^{\bar{m}_{A}} \sum_{\bar{l}=1}^{\bar{m}_{A}} \langle \bar{k}^{A} \| \bar{l}^{A} \rangle^{X} x_{\bar{k}}^{A} x_{\bar{l}}^{A} + E_{N_{nuc}^{A} N_{nuc}^{A}}^{X} \\
&+ \left\{ \sum_{\mu,\nu} (\underline{\underline{P}}^{A})_{\mu\nu} \left[\langle \mu^{A} | - \sum_{\beta=1}^{N_{nuc}^{B}} \frac{Z_{\beta}^{B}}{|\mathbf{r} - \mathbf{R}_{\beta}^{B}|} | \nu^{A} \rangle^{X} + \sum_{\bar{l}=1}^{\bar{m}_{B}} \langle \mu^{A} \nu^{A} \| \bar{l}^{B} \rangle^{X} x_{\bar{l}}^{B} \right] \right. \\
&+ \sum_{\sigma,\tau} (\underline{\underline{P}}^{B})_{\sigma\tau} \left[\langle \sigma^{B} | - \sum_{\alpha=1}^{N_{nuc}^{A}} \frac{Z_{\alpha}^{A}}{|\mathbf{r} - \mathbf{R}_{\alpha}^{A}|} | \tau^{B} \rangle^{X} + \sum_{\bar{k}=1}^{\bar{m}_{A}} \langle \sigma^{B} \tau^{B} \| \bar{k}^{A} \rangle^{X} x_{\bar{k}}^{A} \right] \\
&- \left. \sum_{\bar{k}=1}^{\bar{m}_{A}} \sum_{\bar{l}=1}^{\bar{m}_{B}} \langle \bar{k}^{A} \| \bar{l}^{B} \rangle^{X} x_{\bar{k}}^{A} x_{\bar{l}}^{B} + E_{N_{nuc}^{A} N_{nuc}^{B}}^{X} \right\}, \qquad (6.35)
\end{aligned}
$$

with the following, modified energy-weighted density matrix:

$$
(\underline{\underline{W}}^{A\lambda})_{\mu\nu} = 2 \sum_{i}^{occ(A)} (\epsilon_{i}^{A} - \lambda) c_{\mu i}^{A} c_{\nu i}^{A}, \qquad (6.36)
$$

and where the curly braces delimit the terms specific to KSCED.

Derivation of the 'Pulay relation'

The 'Pulay relation' reads:

$$\sum_{\mu,\nu} (\underline{\underline{P}}^{AX})_{\mu\nu} (\underline{\underline{K}}^{A})_{\mu\nu} = - \sum_{\mu,\nu} (\underline{\underline{W}}^{A})_{\mu\nu} (\underline{\underline{S}}^{AX})_{\mu\nu}, \tag{6.37}$$

where $\underline{\underline{W}}^{A}$ is the energy-weighted density matrix:

$$(\underline{\underline{W}}^{A})_{\mu\nu} = 2 \sum_{i}^{occ(A)} \epsilon_i^A c_{\mu i}^A c_{\nu i}^A . \tag{6.38}$$

Eq. 6.37 can be obtained as follows. From Eq. 6.3 we have:

$$\begin{aligned} \sum_{\mu,\nu} (\underline{\underline{P}}^{AX})_{\mu\nu} (\underline{\underline{K}}^{A})_{\mu\nu} &= \sum_{\mu,\nu} 2 \sum_{i}^{occ(A)} \left(c_{\mu i}^{AX} (\underline{\underline{K}}^{A})_{\mu\nu} c_{\nu i}^{A} + c_{\mu i}^{A} (\underline{\underline{K}}^{A})_{\mu\nu} c_{\nu i}^{AX} \right) \\ &= \sum_{\mu,\nu} 2 \sum_{i}^{occ(A)} 2\, c_{\mu i}^{AX} (\underline{\underline{K}}^{A})_{\mu\nu} c_{\nu i}^{A}, \end{aligned}$$

where the last line results because the KSCED matrix $\underline{\underline{K}}^{A}$ is symmetric. Using the KSCED eigenvalue problem $\sum_{\nu} (\underline{\underline{K}}^{A})_{\mu\nu} c_{\nu i}^{A} = \epsilon_i^A \sum_{\nu} (\underline{\underline{S}}^{A})_{\mu\nu} c_{\nu i}^{A}$, and noting that from the orthonormalization of orbitals $\sum_{\mu,\nu} c_{\mu i}^{A} (\underline{\underline{S}}^{A})_{\mu\nu} c_{\nu i}^{A} = \delta_{ij}$ it follows that $2 \sum_{\mu,\nu} c_{\mu i}^{AX} (\underline{\underline{S}}^{A})_{\mu\nu} c_{\nu i}^{A} = - \sum_{\mu,\nu} c_{\mu i}^{A} (\underline{\underline{S}}^{AX})_{\mu\nu} c_{\nu i}^{A}$, we obtain:

$$\begin{aligned} \sum_{\mu,\nu} (\underline{\underline{P}}^{AX})_{\mu\nu} (\underline{\underline{K}}^{A})_{\mu\nu} &= \sum_{\mu} 2 \sum_{i}^{occ(A)} \left(2\, c_{\mu i}^{AX} \epsilon_i^A \sum_{\nu} (\underline{\underline{S}}^{A})_{\mu\nu} c_{\nu i}^{A} \right) = \\ &- \sum_{\mu,\nu} 2 \sum_{i}^{occ(A)} \epsilon_i^A c_{\mu i}^{A} c_{\nu i}^{A} (\underline{\underline{S}}^{AX})_{\mu\nu} . \end{aligned}$$

Appendix B: Binding energies and Kohn-Sham optimized geometries

The interaction energies discussed here were corrected for the effects associated with the basis set superposition error (BSSE) and the grid superposition error (GSE). The origin and treatment of these two types of errors is similar as in all methods in which the interaction energy is evaluated as the difference of total energies. In the KSCED(s) case, BSSE originates from the fact that for each subsystem, the basis sets includes also atomic orbitals belonging to the other subsystem. Such orbitals are missing in the Kohn-Sham calculations to derive the total energy of the isolated subsystem. The Boys-Bernardi counterpoise technique [147] is used to correct for BSSE. In this case, the counterpoise technique eliminates also the grid superposition error because the same grid is used to calculate the energy of each isolated monomer as well as for the dimer. The situation is different in the case of KSCED(m) calculations. The set of centres of the atomic basis sets is the same for both the isolated- and embedded monomer. The grids differ however. A larger one - that corresponding to the whole dimer - is used for the embedded monomer whereas a smaller one is used for isolated monomer. This imbalance of the grid sizes gives rise to a small numerical effect which is nevertheless also accounted for.

In the KSCED(m) case, the same set of atomic orbitals are is used in both calculations. Because of these particularities in correcting both types of errors in the interaction energy, the procedure to evaluate the corrected interaction energy denoted as $\Delta E(BSSE)$ throughout the text is given below. We use the notation from Ref. [148] and start with the KSCED(s) case.

- the electronic energy of a molecular system M at geometry G computed with basis set σ is denoted by $E_G^\sigma(M)$,
- $\Delta E = E_{AB}^{\alpha\cup\beta}(AB) - \left[E_A^\alpha(A) + E_B^\beta(B)\right]$ is the interaction energy without BSSE correction,
- $\Delta E(FCP) = E_{AB}^{\alpha\cup\beta}(AB) - \left[E_{AB}^{\alpha\cup\beta}(A) + E_{AB}^{\alpha\cup\beta}(B)\right]$ is the interaction energy corrected by means of the function counterpoise (fCP) technique
- $\Delta E(BSSE) = \Delta E(FCP) + E_{rel}^\alpha(A) + E_{rel}^\beta(B) = \Delta E(FCP) + E_{rel}$ is the fCP corrected interaction energy including the relaxation effects of the geometry of each monomer,
- $E_{rel}^\alpha(A) = E_{AB}^\alpha(A) - E_A^\alpha(A)$,
- $E_{rel}^\beta(B) = E_{AB}^\beta(B) - E_B^\beta(B)$.

Note that, in the KSCED(m) case all the above formulas and definitions apply provided that the symbol $\alpha \cup \beta$ means that the α basis set is used for the subsystem A and β for the subsystem B. $\Delta E(FCP)$ defined above takes into account the correction for GSE in the KSCED(m) case.

The reported results are given as binding energies (the negative of the above-mentioned interaction energies) in kcal/mol.

In the statistical analysis the errors have the following definitions ($E_i^{Ref.}$ is the reference binding energy taken from Refs. [134, 135, 136], E_i is the calculated binding energy, and N is the number of the complexes in the given set):

- MD (Mean difference) $= \frac{1}{N}\sum_{i=1}^{N}(E_i^{Ref.} - E_i)$,
- MAD (Mean absolute difference) $= \frac{1}{N}\sum_{i=1}^{N}|E_i^{Ref.} - E_i|$,
- MRAD (Mean relative absolute difference) $= \frac{1}{N}\sum_{i=1}^{N}|E_i^{Ref.} - E_i|/E_i^{Ref.}$,
- MAX (Maximal error) $= \max_{i=1}^{N}|E_i^{Ref.} - E_i|$,
- MAXR (Maximal relative error) $= \max_{i=1}^{N}|E_i^{Ref.} - E_i|/E_i^{Ref.}$).

Table 6.7: Binding energies (in kcal/mol) calculated using the LDA and MG3S basis set. C denotes the number of compound and S its number within the given set, respectively. KSCED(m), KSCED(s), and Ref. denote monomolecular KSCED, supermolecular KSCED, and reference results respectively. NMAX and NMAXR are the numbers of the complexes with the given maximal (MAX and MAXR, respectively) errors.

		KSCED(m)			KSCED(s)				
C, S	Compound	$-\Delta E$	$-\Delta E(fCP)$	$-\Delta E(BSSE)$	$-\Delta E$	$-\Delta E(fCP)$	$-\Delta E(BSSE)$	$-E_{rel}$	Ref.
01, 01	NH_3-NH_3	3.88	3.91	3.88	4.17	4.01	3.99	-0.02	3.15
02, 02	HF-HF	4.04	4.05	4.04	4.51	4.13	4.12	-0.01	4.57
03, 03	H_2O-H_2O	5.06	5.08	5.06	5.44	4.99	4.97	-0.02	4.97
04, 04	NH_3-H_2O	6.53	6.58	6.53	7.03	6.77	6.72	-0.05	6.41
05, 05	$HCONH_2$-$HCONH_2$	14.31	14.99	14.31	15.46	15.71	15.03	-0.68	14.94
06, 06	HCOOH-HCOOH	13.54	14.62	13.53	15.00	15.52	14.43	-1.09	16.15
	MD (MAD)	0.47 (0.79)	0.16 (0.52)	0.47 (0.79)	-0.24 (0.64)	-0.16 (0.51)	0.15 (0.57)		
	MAX (NMAX)	2.61 (6)	1.53 (6)	2.62 (6)	1.15 (6)	0.86 (1)	1.72 (6)		
	MRAD (MAXR, NMAXR)	0.10 (0.23, 1)	0.08 (0.24, 1)	0.10 (0.23, 1)	0.11 (0.32, 1)	0.09 (0.27, 1)	0.09 (0.27, 1)		
07, 01	H_2S-H_2S	1.98	1.98	1.98	2.30	2.12	2.12	0.00	1.66
08, 02	HCl-HCl	2.02	2.02	2.02	2.31	2.19	2.18	-0.01	2.01
09, 03	H_2S-HCl	3.17	3.18	3.17	3.68	3.45	3.44	-0.01	3.35
10, 04	CH_3Cl-HCl	2.73	2.74	2.73	3.05	2.90	2.89	-0.01	3.55
11, 05	HCN-CH_3SH	4.43	4.46	4.44	4.83	4.70	4.68	-0.02	3.59
12, 06	CH_3SH-HCl	4.02	4.04	4.02	4.64	4.43	4.40	-0.03	4.16
	MD (MAD)	-0.01 (0.39)	-0.02 (0.38)	-0.01 (0.39)	-0.41 (0.58)	-0.24 (0.46)	-0.23 (0.45)		
	MAX (NMAX)	0.84 (5)	0.87 (5)	0.85 (5)	1.24 (5)	1.11 (5)	1.09 (5)		
	MRAD (MAXR, NMAXR)	0.12 (0.23, 5)	0.12 (0.24, 5)	0.13 (0.24, 5)	0.21 (0.39, 1)	0.16 (0.31, 5)	0.16 (0.30, 5)		
13, 01	He-Ne	0.08	0.08	0.08	0.14	0.09	0.09	0.00	0.04
14, 02	He-Ar	0.14	0.14	0.14	0.17	0.15	0.15	0.00	0.06
15, 03	Ne-Ne	0.07	0.07	0.07	0.20	0.08	0.08	0.00	0.08
16, 04	Ne-Ar	0.11	0.11	0.11	0.22	0.12	0.12	0.00	0.13
17, 05	CH_4-Ne	0.12	0.12	0.13	0.29	0.15	0.15	0.00	0.22
18, 06	C_6H_6-Ne	0.00	0.23	0.23	0.00	0.23	0.23	0.00	0.47
19, 07	CH_4-CH_4	0.42	0.42	0.42	0.45	0.44	0.44	0.00	0.51
20, 08	C_2H_2-C_2H_2	1.77	1.77	1.77	1.90	1.85	1.84	-0.01	1.34
21, 09	C_2H_4-C_2H_4	1.04	1.05	1.03	1.10	1.08	1.06	-0.02	1.42
	MD (MAD)	0.06 (0.18)	0.03 (0.15)	0.03 (0.15)	-0.02 (0.21)	0.01 (0.15)	0.01 (0.15)		
	MAX (NMAX)	0.47 (6)	0.43 (8)	0.43 (8)	0.56 (8)	0.51 (8)	0.50 (8)		
	MRAD (MAXR, NMAXR)	0.54 (1.33, 2)	0.48 (1.33, 2)	0.48 (1.33, 2)	0.96 (2.50, 1)	0.49 (1.50, 2)	0.49 (1.50, 2)		
	MD (MAD)	0.16 (0.41)	0.05 (0.32)	0.15 (0.40)	-0.20 (0.44)	-0.11 (0.34)	-0.02 (0.36)		
	MAX (NMAX)	2.61 (6)	1.53 (6)	2.62 (6)	1.24 (11)	1.11 (11)	1.72 (6)		
	MRAD (MAXR, NMAXR)	0.29 (1.33, 14)	0.27 (1.33, 14)	0.27 (1.33, 14)	0.50 (2.50, 13)	0.28 (1.50, 14)	0.28 (1.50, 14)		

Table 6.8: Binding energies (in kcal/mol) calculated using the GGA and MG3S basis set. C denotes the number of compound and S its number within the given set, respectively. KSCED(m), KSCED(s), and Ref. denote monomolecular KSCED, supermolecular KSCED, and reference results respectively. NMAX and NMAXR are the numbers of the complexes with the given maximal (MAX and MAXR, respectively) errors.

		KSCED(m)			KSCED(s)				
C, S	Compound	$-\Delta E$	$-\Delta E(fCP)$	$-\Delta E(BSSE)$	$-\Delta E$	$-\Delta E(fCP)$	$-\Delta E(BSSE)$	$-E_{rel}$	Ref.
01, 01	NH_3-NH_3	5.25	5.30	5.25	5.83	5.64	5.59	-0.05	3.15
02, 02	HF-HF	4.92	4.93	4.92	5.47	5.01	5.00	-0.01	4.57
03, 03	H_2O-H_2O	6.03	6.05	6.03	6.49	5.96	5.94	-0.02	4.97
04, 04	NH_3-H_2O	7.70	7.76	7.70	8.41	8.09	8.03	-0.06	6.41
05, 05	$HCONH_2$-$HCONH_2$	16.86	17.73	16.87	18.33	18.71	17.85	-0.87	14.94
06, 06	HCOOH-HCOOH	16.27	17.62	16.28	18.11	18.81	17.47	-1.34	16.15
	MD (MAD)	-1.14 (1.14)	-1.53 (1.53)	-1.14 (1.14)	-2.08 (2.08)	-2.00 (2.00)	-1.61 (1.61)		
	MAX (NMAX)	2.10 (1)	2.79 (5)	2.10 (1)	3.39 (5)	3.77 (5)	2.91 (5)		
	MRAD (MAXR, NMAXR)	0.22 (0.67, 1)	0.24 (0.68, 1)	0.22 (0.67, 1)	0.34 (0.85, 1)	0.29 (0.79, 1)	0.27 (0.77, 1)		
07, 01	H_2S-H_2S	2.55	2.55	2.55	2.95	2.76	2.76	0.00	1.66
08, 02	HCl-HCl	2.93	2.94	2.93	3.33	3.20	3.19	-0.01	2.01
09, 03	H_2S-HCl	3.96	3.97	3.96	4.57	4.35	4.34	-0.01	3.35
10, 04	CH_3Cl-HCl	3.80	3.83	3.80	4.21	4.07	4.05	-0.02	3.55
11, 05	HCN-CH_3SH	5.42	5.47	5.42	5.87	5.79	5.73	-0.05	3.59
12, 06	CH_3SH-HCl	5.21	5.24	5.21	5.96	5.77	5.74	-0.04	4.16
	MD (MAD)	-0.93 (0.93)	-0.95 (0.95)	-0.93 (0.93)	-1.43 (1.43)	-1.27 (1.27)	-1.25 (1.25)		
	MAX (NMAX)	1.83 (5)	1.88 (5)	1.83 (5)	2.28 (5)	2.20 (5)	2.14 (5)		
	MRAD (MAXR, NMAXR)	0.33 (0.54, 1)	0.34 (0.54, 1)	0.33 (0.54, 1)	0.51 (0.78, 1)	0.45 (0.66, 1)	0.44 (0.66, 1)		
13, 01	He-Ne	0.46	0.46	0.46	0.52	0.46	0.46	0.00	0.04
14, 02	He-Ar	0.45	0.45	0.45	0.49	0.47	0.47	0.00	0.06
15, 03	Ne-Ne	0.52	0.53	0.53	0.66	0.52	0.52	0.00	0.08
16, 04	Ne-Ar	0.51	0.51	0.51	0.67	0.54	0.54	0.00	0.13
17, 05	CH_4-Ne	0.60	0.60	0.60	0.82	0.66	0.66	0.00	0.22
18, 06	C_6H_6-Ne	0.91	0.92	0.92	1.24	0.95	0.95	0.00	0.47
19, 07	CH_4-CH_4	0.97	0.97	0.97	1.04	1.03	1.03	0.00	0.51
20, 08	C_2H_2-C_2H_2	2.62	2.64	2.62	2.79	2.74	2.72	-0.02	1.34
21, 09	C_2H_4-C_2H_4	1.98	2.00	1.97	2.07	2.03	2.00	-0.02	1.42
	MD (MAD)	-0.53 (0.53)	-0.53 (0.53)	-0.53 (0.53)	-0.67 (0.67)	-0.57 (0.57)	-0.56 (0.56)		
	MAX (NMAX)	1.28 (8)	1.30 (8)	1.28 (8)	1.45 (8)	1.40 (8)	1.38 (8)		
	MRAD (MAXR, NMAXR)	3.37 (10.50, 1)	3.39 (10.50, 1)	3.39 (10.50, 1)	4.17 (12.00, 1)	3.50 (10.50, 1)	3.50 (10.50, 1)		
	MD (MAD)	-0.82 (0.82)	-0.94 (0.94)	-0.82 (0.82)	-1.29 (1.29)	-1.18 (1.18)	-1.06 (1.06)		
	MAX (NMAX)	2.10 (1)	2.79 (5)	2.10 (1)	3.39 (5)	3.77 (5)	2.91 (5)		
	MRAD (MAXR, NMAXR)	1.60 (10.50, 13)	1.62 (10.50, 13)	1.61 (10.50, 13)	2.03 (12.00, 13)	1.71 (10.50, 13)	1.70 (10.50, 13)		

In the tables below, we present the key geometrical parameters of the equilibrium geometries obtained from Kohn-Sham calculations, as well as corresponding binding energies.

Table 6.9: The key parameters of the equilibrium geometry obtained from KS LDA calculations. R (in Å) denotes distance between the two closest heavy atoms of different monomers, and ϕ (in degrees) is the selected angle determining the relative orientation between the monomers. The labels used for these parameters specified in the fields Def_R and Def_ϕ identify the relevant atoms and are the same as the ones used in the Supplementary Material-1. The reference values R_{ref} and ϕ_{ref} are taken from Ref. [135].

Compound	Def_R	R	R_{ref}	$R - R_{ref}$	Def_ϕ	ϕ	ϕ_{ref}
NH_3-NH_3	dNN	3.00	3.27	-0.27	HNN	14	14
HF-HF	dFF	2.56	2.78	-0.22	HFF	100	115
H_2O-H_2O	dOO	2.73	2.94	-0.21	HOO	5	4
NH_3-H_2O	dNO	2.75	2.97	-0.22	HON	6	6
$HCONH_2$-$HCONH_2$	dNO	2.69	2.88	-0.19	ONC	118	116
HCOOH-HCOOH	dOO	2.47	2.70	-0.23	OOC	120	125
H_2S-H_2S	dSS	3.74	4.12	-0.38	HSS	84	84
HCl-HCl	dClCl	3.46	3.79	-0.33	HClCl	7	8
H_2S-HCl	dSCl	3.44	3.76	-0.32	HClS	88	88
CH_3Cl-HCl	dClCl	3.33	3.61	-0.28	ClClC	80	82
HCN-CH_3SH	dCS	3.33	3.52	-0.19	SNC	161	162
CH_3SH-HCl	dSCl	3.31	3.61	-0.30	HClS	10	11
He-Ne	dHeNe	2.47	3.03	-0.56	-	-	-
He-Ar	dHeAr	2.93	3.48	-0.55	-	-	-
Ne-Ne	dNeNe	2.56	3.09	-0.53	-	-	-
Ne-Ar	dNeAr	2.93	3.49	-0.56	-	-	-
CH_4-Ne	dCNe	3.24	3.49	-0.25	HNeC	70	71
C_6H_6-Ne	dCNe	3.09	3.51	-0.42	NeCC	78	79
CH_4-CH_4	dCC	3.33	3.61	-0.28	HCC	70	70
C_2H_2-C_2H_2	dCC	3.14	3.46	-0.32	CCC	119	123
C_2H_4-C_2H_4	dCC	3.54	3.83	-0.29	CCC	80	80

Table 6.10: The key parameters of the equilibrium geometry obtained from KS PW91 calculations. R (in Å) denotes distance between the two closest heavy atoms of different monomers, and ϕ (in degrees) is the selected angle determining the relative orientation between the monomers. The labels used for these parameters specified in the fields Def_R and Def_ϕ identify the relevant atoms and are the same as the ones used in the Supplementary Material-1. The reference values R_{ref} and ϕ_{ref} are taken from Ref. [135].

Compound	Def_R	R	R_{ref}	$R - R_{ref}$	Def_ϕ	ϕ	ϕ_{ref}
NH_3-NH_3	dNN	3.21	3.27	-0.06	HNN	14	14
HF-HF	dFF	2.72	2.78	-0.06	HFF	109	115
H_2O-H_2O	dOO	2.87	2.94	-0.07	HOO	4	4
NH_3-H_2O	dNO	2.88	2.97	-0.09	HON	6	6
$HCONH_2$-$HCONH_2$	dNO	2.84	2.88	-0.04	ONC	117	116
HCOOH-HCOOH	dOO	2.61	2.70	-0.09	OOC	122	125
H_2S-H_2S	dSS	4.01	4.12	-0.11	HSS	87	84
HCl-HCl	dClCl	3.70	3.79	-0.09	HClCl	7	8
H_2S-HCl	dSCl	3.60	3.76	-0.16	HClS	90	88
CH_3Cl-HCl	dClCl	3.54	3.61	-0.07	ClClC	88	82
HCN-CH_3SH	dCS	3.61	3.52	0.09	SNC	171	162
CH_3SH-HCl	dSCl	3.48	3.61	-0.13	HClS	6	11
He-Ne	dHeNe	2.76	3.03	-0.27	-	0	-
He-Ar	dHeAr	3.33	3.48	-0.15	-	0	-
Ne-Ne	dNeNe	2.99	3.09	-0.10	-	0	-
Ne-Ar	dNeAr	3.38	3.49	-0.11	-	0	-
CH_4-Ne	dCNe	3.48	3.49	-0.01	HNeC	70	71
C_6H_6-Ne	dCNe	3.53	3.51	0.02	NeCC	79	79
CH_4-CH_4	dCC	3.97	3.61	0.36	HCC	71	70
C_2H_2-C_2H_2	dCC	3.50	3.46	0.04	CCC	124	123
C_2H_4-C_2H_4	dCC	4.11	3.83	0.28	CCC	84	80

Table 6.11: Binding energies (in kcal/mol) calculated using the LDA and MG3S basis set. C denotes the number of compound and S its number within the given set, respectively. KS, and Ref. denote Kohn-Sham, and reference results respectively.

		KS				
C, S	Compound	$-\Delta E$	$-\Delta E(fCP)$	$-\Delta E(BSSE)$	$-E_{rel}$	Ref.
01, 01	NH_3-NH_3	6.25	6.15	6.03	-0.12	3.15
02, 02	HF-HF	8.54	8.17	8.01	-0.16	4.57
03, 03	H_2O-H_2O	9.19	8.80	8.60	-0.20	4.97
04, 04	NH_3-H_2O	11.50	11.63	11.10	-0.53	6.41
05, 05	$HCONH_2$-$HCONH_2$	23.29	26.24	22.69	-3.55	14.94
06, 06	HCOOH-HCOOH	29.18	43.72	28.16	-15.56	16.15
07, 01	H_2S-H_2S	4.31	4.18	4.08	-0.10	1.66
08, 02	HCl-HCl	5.00	4.85	4.73	-0.12	2.01
09, 03	H_2S-HCl	8.13	8.42	7.81	-0.61	3.35
10, 04	CH_3Cl-HCl	7.92	8.05	7.63	-0.42	3.55
11, 05	HCN-CH_3SH	6.96	7.08	6.78	-0.30	3.59
12, 06	CH_3SH-HCl	11.26	12.37	10.92	-1.45	4.16
13, 01	He-Ne	0.40	0.33	0.33	0.00	0.04
14, 02	He-Ar	0.38	0.36	0.36	0.00	0.06
15, 03	Ne-Ne	0.58	0.42	0.42	0.00	0.08
16, 04	Ne-Ar	0.72	0.55	0.55	0.00	0.13
17, 05	CH_4-Ne	0.67	0.52	0.52	0.00	0.22
18, 06	C_6H_6-Ne	1.53	1.11	1.14	0.04	0.47
19, 07	CH_4-CH_4	1.15	1.15	1.13	-0.02	0.51
20, 08	C_2H_2-C_2H_2	2.54	2.50	2.47	-0.02	1.34
21, 09	C_2H_4-C_2H_4	3.05	3.01	2.98	-0.03	1.42

Table 6.12: Binding energies (in kcal/mol) calculated using the GGA and MG3S basis set. C denotes the number of compound and S its number within the given set, respectively. KS, and Ref. denote Kohn-Sham, and reference results respectively.

		KS				
C, S	Compound	$-\Delta E$	$-\Delta E(fCP)$	$-\Delta E(BSSE)$	$-E_{rel}$	Ref.
01, 01	NH_3-NH_3	3.70	3.54	3.50	-0.04	3.15
02, 02	HF-HF	5.46	5.02	4.97	-0.05	4.57
03, 03	H_2O-H_2O	5.89	5.43	5.35	-0.08	4.97
04, 04	NH_3-H_2O	7.73	7.61	7.36	-0.25	6.41
05, 05	$HCONH_2$-$HCONH_2$	14.98	16.09	14.52	-1.56	14.94
06, 06	HCOOH-HCOOH	17.70	22.08	16.96	-5.12	16.15
07, 01	H_2S-H_2S	2.04	1.90	1.88	-0.02	1.66
08, 02	HCl-HCl	2.41	2.25	2.22	-0.04	2.01
09, 03	H_2S-HCl	4.72	4.70	4.45	-0.25	3.35
10, 04	CH_3Cl-HCl	3.89	3.80	3.67	-0.13	3.55
11, 05	HCN-CH_3SH	4.01	3.98	3.89	-0.09	3.59
12, 06	CH_3SH-HCl	6.46	6.86	6.19	-0.67	4.16
13, 01	He-Ne	0.32	0.27	0.27	0.00	0.04
14, 02	He-Ar	0.25	0.24	0.24	0.00	0.06
15, 03	Ne-Ne	0.45	0.33	0.33	0.00	0.08
16, 04	Ne-Ar	0.45	0.35	0.35	0.00	0.13
17, 05	CH_4-Ne	0.53	0.40	0.40	0.00	0.22
18, 06	C_6H_6-Ne	0.79	0.50	0.55	0.05	0.47
19, 07	CH_4-CH_4	0.40	0.39	0.40	0.01	0.51
20, 08	C_2H_2-C_2H_2	1.23	1.20	1.20	0.00	1.34
21, 09	C_2H_4-C_2H_4	0.82	0.78	0.79	0.01	1.42

Part IV

Multi-level simulations in condensed phase

Chapter 7

Modeling solvatochromic shifts using orbital-free embedding potential at statistically averaged solvent density

The correspondence between the exact embedding potential and the pair of the electron densities - that of the embedded molecule and that of its environment [Wesolowski and Warshel, *J. Phys. Chem.* **97** (1993), 8050] - is used to generate the average embedding potential and to calculate subsequently the solvatochromic shifts in a number of organic chromophores in solvents of various polarities. The averaged embedding potential is evaluated at a fictitious electron density of the solvent, which is obtained by means of "dressing up" with electrons the classical site distributions derived from the statistical-mechanical, 3D molecular theory of solvation (a.k.a. 3D-RISM method) [Kovalenko in *Molecular theory of solvation*, Ed. by Hirata, Understanding Chemical Reactivity, vol. 24, 2003], self-consistently coupled with the electronic structure of the solute. The proposed approach to modeling solvatochromic shifts can be situated between the implicit and explicit type of models for the solvent. Numerical examples are given for the lowest-lying $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ excitations.

7.1 Introduction

Solvents are known to affect such physicochemical properties of molecules as geometry, conformational equilibria, reaction rates, as well as their UV/Vis, IR, or NMR spectra, for instance [149]. A large number of computer modeling techniques were developed to study solvent effects on such properties. In the most simple ones, the environment of a solvated molecule is treated implicitly. It is represented as a cavity in a homogeneous dielectric which gives rise to the reaction field (for recent reviews see Ref. [150, 151]). The implicit models have been used extensively and describe adequately the solvent effects on many properties. One of their drawbacks is neglecting specific intermolecular interactions such as hydrogen bonds. Explicit solvent models, where both the solvent as well as the environment are described at the level of non-relativistic Schrödinger equation in Born-Oppenheimer approximation, can be situated at another end. In principle, such models can describe almost every property of solvated molecules but their practical applications are limited for two principal reasons: the fast increase of the computational cost with the increase of the size of the system and the need for averaging due to the statistical nature of the solvent. The explicit models can be simplified by treating the whole quantum mechanical system comprising the solvated molecule on one hand and the solvent molecules on another hand in a hybrid way, in which the system of interest is described at the quantum-mechanical (QM) level, whereas solvent at some lower level of description. Such schemes are known under such labels as QM/QM [152] or QM/MM [153]. Another group of approaches, where the solvent is described by means of site probabilities, can be situated between implicit and explicit levels. The statistical-mechanical molecular theory of liquids, three-dimensional reference interaction site model (3D-RISM) [154] in particular, provides a formal framework to obtain the probabilities (referred to in the present work as *site distributions*) of finding a solvent molecule at a given position and orientation with respect to the solute. At the 3D-RISM level, the solvent is not uniform and the specific interactions are accounted for as modifications of the three-dimensional site distributions, which are statistical quantities. The interactions between the subsystem described at the quantum-mechanical level and its 3D-RISM environment are taken into account in a straightforward manner by the electrostatic potential derived from 3D-RISM site distributions.

Turning back to the QM/MM and QM/QM models, the possible strategy to include the solvent in the calculation of molecular properties is based on the embedding strategy, in which only a selected part of the whole system is described using quantum-mechanical descriptors, whereas the remaining part (environment) is taken into account by means of a special potential (*embedding potential*) added to the gas phase Hamiltonian. The most commonly used embedding potentials comprise terms originating in the theory of intermolecular interactions. In the simplest version, the embedding potential comprises only terms representing classical Coulomb interactions, which are frequently further simplified using truncated multipole expansion, whereas more refined potentials account also for solvent polarization through a reaction field (discrete or continuous) [155, 156, 157, 152, 153] (for extensive review, see [78, 77, 150, 151]). The Hohenberg-Kohn variational principle provides another strategy to construct the embedding potential in a system-independent way [4, 36, 39]. Wesolowski and Warshel [4] showed that the optimal (i.e. minimizing the Hohenberg-Kohn energy functional [1]) embedding potential is uniquely determined by the pair of the electron densities - that of the investigated embedded system ($\rho_A(\vec{r})$) and that of its environment ($\rho_B(\vec{r})$). Using density functionals for: *i*) the kinetic energy of non-interacting reference system ($T_s[\rho]$) and *ii*) the exchange-correlation energy ($E_{xc}[\rho]$), known in the Kohn-Sham formulation [2] of density functional theory and their constrained search definitions [3], this correspondence reads:

$$\begin{aligned} v_{emb}^{KSCED}[\rho_A,\rho_B;\vec{r}] &= v_{ext}^{B}(\vec{r}) + \int \frac{\rho_B(\vec{r'})}{|\vec{r'}-\vec{r}|} d\vec{r'} \\ &+ v_{xc}^{nad}[\rho_A,\rho_B](\vec{r}) + v_{t}^{nad}[\rho_A,\rho_B](\vec{r}) \end{aligned} \tag{7.1}$$

where:

$$v_t^{nad}[\rho_A,\rho_B](\vec{r}) = \left.\frac{\delta T_s[\rho]}{\delta\rho}\right|_{\rho=\rho_A+\rho_B} - \left.\frac{\delta T_s[\rho]}{\delta\rho}\right|_{\rho=\rho_A} \tag{7.2}$$

and

$$v_{xc}^{nad}[\rho_A,\rho_B](\vec{r}) = \left.\frac{\delta E_{xc}[\rho]}{\delta\rho}\right|_{\rho=\rho_A+\rho_B} - \left.\frac{\delta E_{xc}[\rho]}{\delta\rho}\right|_{\rho=\rho_A} \tag{7.3}$$

are non-additive kinetic-energy and exchange-correlation potentials, respectively. Term $v_{ext}^{B}(\vec{r})$ is external potential generated by nuclei of the environment.

The above equation provides system-independent, exact expression for the universal embedding potential coupling two subsystems. The Kohn-Sham-like equations for embedded orbitals ϕ_A:

$$\left[-\frac{1}{2}\nabla^2 + v_{eff}^{KS}\left[\rho_A, \vec{r}\right] + v_{eff}^{KSCED}\left[\rho_A, \rho_B; \vec{r}\right]\right]\phi_i^A = \epsilon_i^A \phi_i^A \qquad i = 1, N^A \tag{7.4}$$

provide a self-consistent way to minimize of the energy of the total system for a given $\rho_B(\vec{r})$, i.e., to optimize the embedded electron density in the subsystem A ($\rho_A = 2\sum_i^{N_A}\left|\phi_i^A\right|^2$). For the complete description of the formalism based on the above equation see Ref. [29]. The environment is described only by its electron density ρ_B, which is an assumed quantity. No additional information about orbital structure of solvent is required. Note also, that density ρ_B remains constant during iterative process of solving Eq. 7.4. For this reasons we commonly refer to such method as Orbital-Free Embedding (OFE) or Frozen Density Embedding (FDE). The label KSCED in Eq. 7.4 stands for the Kohn-Sham Equations with Constrained Electron Density and is introduced to indicate that the total effective potential in Eq. 7.4 differs from the one in the Kohn-Sham equations.

Although $v_{emb}^{KSCED}[\rho_A, \rho_B; \vec{r}]$ was derived for embedding a reference system of non-interacting electrons, the same or very similar correspondence between the optimal embedding potential and the pair of electron densities ρ_A and ρ_B exist for embedding other quantum mechanical objects such as embedded wavefunction [36] or embedded one-particle reduced density matrix [39]. The results of practical applications of the embedding potential given in Eq. 7.1 in combination with wavefunction based models are encouraging [37, 38, 158]. Considering applications of Eq. 7.4, various studies are available in the literature [127, 53, 86, 84]. In all of them the quality of the results hinges on the choice of the density ρ_B corresponding to the environment. Owing to the fact that this density is an assumed quantity, a variety of theoretical methods working at different level of approximation can be applied to obtain ρ_B. In our common practice, depending on the studied system or on the observable of interest, we adopted several strategies to generate ρ_B, all of them based on the Kohn-Sham formalism. The lowest level of description of the electron density of the environment is to construct ρ_B as the sum of the atomic or molecular densities of atoms or molecules in the environment [159]. This level might be adequate if neither the solvent molecules nor the embedded molecule are polar. This model can be further refined if $\rho_B(\vec{r})$ is obtained in one Kohn-Sham calculation for the whole environment but with the absence of the embedded molecule. Such treatment of the solvent is adequate only if the electron density of the environment is not significantly modified by the presence of the solute (electronic polarization of the solvent molecules, for instance). In such a way, ρ_B includes the interactions between molecules constituting it, however, it does not account for the polarization of the environment by the embedded subsystem. This can be partly overcome by performing the Kohn-Sham calculations of the whole environment including point charges or dipole moments representing the field of embedded system, leading to *pre-polarized* ρ_B, i.e., to use as ρ_B the electron density of the molecules in the environment, which is polarized by the electric field generated by the solute. The most accurate approach is not to do any assumptions concerning ρ_B but to obtain it variationally, i.e., to find such a pair of densities ρ_A and ρ_B, that minimizes the Cortona functional for the total energy [28]. Such minimization can be performed using the the *freeze-and-thaw* algorithm [31]. In *freeze-and-thaw* calculations, both the embedded molecule and the environment are treated on equal footing and the notion of embedded system and the environment is not applicable. In our applications of Eq. 7.4, the *freeze-and-thaw* calculations are used to verify the adequacy of simplifications made for $\rho_B(\vec{r})$ in large scale simulations. Numerical examples of the convergence of the calculated properties along the above hierarchy of methods to generate the electron density of the environment can be found in our previous works on ligand-fields splitting of f-levels in lanthanide impurities [84] or on solvent effect on the UV/Vis spectra of organic chromophores [53, 86]. In the discussed above schemes, $\rho_B(\vec{r})$ has always a quantum-chemical origin and can be associated with some molecular system at a given geometry. In the present work, an alternative approach to generate density $\rho_B(\vec{r})$ is introduced. The quantum-chemical descriptors for the electron density of the environment are not used at all. Instead, the electron density of the environment is obtained using the classical-theory-of-liquids derived site distributions, which are subsequently "dressed up" with electrons. The resulting electron density - the averaged electron density of the solvent denoted with $\langle\rho_B(\vec{r})\rangle$ throughout this work - is used subsequently in the evaluation of the orbital-free embedding potential given in Eq. 7.1. We exploit here explicitly the fact that Eq. 7.1 provides an unique correspondence between the exact embedding potential and charge densities even if $\rho_B(\vec{r})$ is not obtained from any quantum mechanical method [4, 36, 39]. $\langle\rho_B(\vec{r})\rangle$ cannot be associated with the electron density of any molecular system as it represents a local function giving the ensemble averaged probability to find an electron in a given place in space. This is an entirely new strategy in the orbital-free embedding calculations. In the subsequent discussion, we will refer to such potential as the *average orbital-free embedding potential* or just *average embedding potential*. It is worthwhile to recall that the idea of averaging the potential generated by the solvent is not new in molecular simulations [160, 161, 162, 163]. The name - "Averaged Solvent Electrostatic Potential" - used for such a potential in Ref. [160], for instance, illustrates clearly the difference between such averaged potentials and the average embedding potential in the present work, where it is not the potential which is averaged but the solvent charge density, and where the potential in question includes also a non-electrostatic part. The difference between averaging the potential directly or calculating it at the averaged solvent charge density arises from the fact that the potential given by Eq. 2.7 is non-linear in ρ_B [164].

The stress on the embedding potential not just on the expression for the total energy of an embedded system is made here in view of our interest in properties directly related to the electronic structure. The target property analyzed in the present work is the shift of the absorption band center of a chromophore induced by the interactions with the solvent molecules (solvatochromic shift). A straightforward approach to simulate solvatochromic shifts uses an ensemble representing the solvent and evaluation of the excitation energy for each configuration in the ensemble. The quality of the spectral shifts obtained in this way depends on the quality of the evaluated excitation energy at each geometry but also on the used sample of geometries of the solvent. Concerning the latter factor, molecular dynamics simulations are frequently used for this purpose [53, 86, 165, 166]. Such simulations allow one to obtain not only the spectral shifts but also the shape of the absorption bands.

As far as the former factor is concerned, the spectral shifts can be obtained following the linear-response time-dependent density-functional-theory (LR-TDDFT) strategy [51] applied for the embedded orbitals (i.e. orbitals derived from Eq. 7.4). For a description of the details of such calculations and benchmark results, see Refs. [52, 127]. The absorption band in such a case is just:

$$I(\omega) = \sum_i f_i \left[\rho_B(\{R_j^B\})\right] \delta\left(\omega - \omega_i\left[\rho_B(\{R_j^B\})\right]\right) \tag{7.5}$$

where f_i and ω_i are the oscillator strength and the resonance frequency for the transition i, $\{R_j^B\}$ denotes the positions of the nuclei in the environment (j changes from 1 to the total number of atoms in the environment), f_i and ω_i are functionals of the electron density $\rho_B(\vec{r})$ of the environment of the chromophore, and δ is the Dirac delta function. Note that the square brackets in the above equation denote the fact that the evaluated quantity is a functional of the electron density, i.e., is uniquely determined by the function inside the square brackets.

For an environment of a flexible structure, such as the molecules surrounding a chromophore in the liquid phase, the picture based on single arrangement of nuclei in the environment ($\{R_j^B\}$) is not adequate. Due to difference in time-scale of the electronic transitions and the motions of the nuclei in the environment, Eq. 7.5 can be generalized by averaging over the ensamble of configurations $\{R_j^{B(n)}\}$, where n is the index for the configurations of the nuclei in the solvent.

$$I(\omega) = \left\langle \sum_i f_i \left[\rho_B(\{R_j^{B(n)}\})\right] \delta\left(\omega - \omega_i \left[\rho_B(\{R_j^{B(n)}\})\right]\right) \right\rangle_n \tag{7.6}$$

In our previously reported work on acetone [53] and on aminocoumarin [86], the ensemble of structures of the solvent $\{R_j^{B(n)}\}$ was obtained from either classical, i.e., pair-potential based or quantum-mechanical (Car-Parrinello [167]) molecular dynamics simulations. For such chromophores, which are characterized by broad and diffuse absorption bands, the strategy based on Eq. 7.6 implies, however, averaging over a large number of orientations, otherwise the sample of the used configurations is not representative. The CPU costs of the calculation of the excitation energy at one configuration of the solvent can be significantly reduced by using the aforementioned hierarchy of methods to generate $\rho_B(\vec{r})$ in a simplified manner. In practice, it is the number of configurations used to represent the solvent, that determines the overall CPU costs of the simulation of the absorption band.

In the present work, we consider an alternative strategy to obtain the spectral shifts. Instead of averaging f_i and ω_i over a large number of configurations, these quantities are evaluated for a fictitious density $\langle\rho_B(\vec{r})\rangle$, which is the ensemble averaged electron density of the environment. Simulating the absorption bands by performing the orbital-free embedding calculations of the excitation energy at the averaged solvent density $\langle\rho_B(\vec{r})\rangle$ is proposed here principally as the method to overcome the difficulty in obtaining the excitation energy for a sufficient number of solvent configurations to represent adequately the statistical ensemble. This problem arises from the fact that the overall computational costs of the simulation increases linearly with the number of configurations used in the averaging procedure and is proportional to the computational cost of the evaluation of the excitation energy for a given configuration of the solvent.

The strategy to obtain the averaged quantities applied in the present work corresponds to the following approximation:

$$\begin{aligned} I(\omega) &= \left\langle \sum_i f_i \left[\rho_B(\{R_j^{B(n)}\})\right] \delta\left(\omega - \omega_i \left[\rho_B(\{R_j^{B(n)}\})\right]\right) \right\rangle_n \\ &\approx \sum_i f_i \left[\langle\rho_B\rangle\right] \delta(\omega - \omega_i \left[\langle\rho_B\rangle\right]) \end{aligned} \tag{7.7}$$

Note that the averaged electron density of the solvent ($\langle\rho_B\rangle$) depends on the position but it does not involve any explicit variables describing the geometry of the solvent molecules. Simplifying the description of the environment by abandoning the atomistic treatment of the ensemble of the solvent and using, instead, the average density of the solvent $\langle\rho_B(\vec{r})\rangle$ in the evaluation of excitation energy aims, therefore, at obtaining better averages. It is, however, an approximation introduced in Eq. 7.7. Moreover, $\langle\rho_B(\vec{r})\rangle$ provides less complete descriptor of the solvent as the statistical ensemble of configurations. For instance, the shape of the absorption band cannot be obtained from such calculations but only individual lines (see Eq. 7.7). Finally, the position of the absorption band depends on the quality of the used $\langle\rho_B(\vec{r})\rangle$.

Eq. 7.7 is the main new element in the computational protocol for calculating the solvatochromic shifts proposed in this work and it is also one of the two main approximations (the other is the procedure to obtain $\langle\rho_B(\vec{r})\rangle$ described in the subsequent sections). Besides the fact that, the approximation given in Eq. 7.7 cannot describe the shapes of the absorption lines, even the position of the absorption band can be affected by this approximation. Averaging the electron density of the solvent, neglects the coupling between the instantaneous electronic structure of the solute and the structure of the solvent. Such coupling can be expected to play an important role for nonpolar but polarisable solutes in polar solvents. The unique correspondence between the pair of electron densities (solvent and solute) and the embedding potential given in Eq. 7.1 brings also another important consequence for the linear-response time-dependent density-functional-theory calculations which hinge on functional derivatives of the effective potential. Such correspondence makes it possible not only to replace averaging the potential by potential calculated at the average density $\langle \rho_B(\vec{r}) \rangle$ but also do the same for response kernel, it is instead of averaging the kernel, calculation of the kernel at $\langle \rho_B(\vec{r}) \rangle$.

In the present work, we obtain $\langle \rho_B(\vec{r}) \rangle$ not from quantum-mechanical calculations, but by "dressing up" with electrons the classical site distributions (quantities related to nuclear positions rather than to electrons) obtained from a particular variant of the Ornstein-Zernike equations of the statistical theory of liquids, i.e., the three-dimensional reference interacting site model with the Kovalenko-Hirata closure approximation (3D-RISM-KH) [154]. Such densities are subsequently "dressed up" with electron densities leading to the average electron density of the solvent, which is denoted with $\langle \rho_B(\vec{r}) \rangle$. The overall computational protocol combining: the average electron density of the liquid obtained as "dressed up" 3D-RISM-KH site distributions, orbital-free embedding with the applied approximants in Eq. 7.4, and the evaluation of the spectral shifts from linear-response calculations for embedded subsystem, will be labeled with OFE/RISM for the sake of simplicity in the present work.

Several chromophores, for which the strong dependency of the UV/Vis spectra bands on the surrounding media is known, are chosen for the study. Our previous works validated the use of the orbital-free embedding strategy to obtain hydrogen-bonding induced shifts of the excitation energies in the case of a rigid environment of the chromophore. The present work addresses rather the issue of efficient calculation of averages for flexible environments such as the considered solvents.

7.2 Methodology

7.2.1 Spectral shifts from the orbital-free embedding calculations at a rigid geometry of the environment

The applications of the orbital-free embedding potential to evaluate the complexation-induced shifts of the low-lying excitations in organic chromophores follows the same framework as the one introduced in Ref. [52]

and subsequently applied in several works [53, 86, 127]. In this framework, the linear-response time-dependent density-functional strategy to obtain electronic excitation is applied not for the Kohn-Sham orbitals and the Kohn-Sham effective potential but for the corresponding quantities in Eq. 7.4, i.e., the embedded orbitals and the total effective potential. Such computational scheme is especially suited for modeling the complexation induced shifts and not the absolute values of the excitation energies because the latter depend on the quality of the excitation energy for the isolated chromophore derived from LR-TDDFT. The straightforward application of the supermolecular strategy, i.e., evaluation of the shifts as the difference between the LR-TDDFT excitation energies calculated for the chromophore in the complex and for the isolated chromophore, hinges on the cancellation of errors in the effective potential and in its functional derivative (response kernel). The errors in these quantities result in the errors in excitation energies of the magnitude reaching the 0.1-0.2 eV [168], i.e., the same order of magnitude as the investigated spectral shifts. The situation is quite different in the orbital-free embedding case. Owing to the unique correspondence between the pair of the electron densities (chromophore and its environment) and the embedding potential (Eq. 7.1), also the embedding component of the response kernel is uniquely determined by this pair of densities. The errors in the calculated spectral shifts at a given rigid geometry of the environment arise from the fact that non-electrostatic terms in Eq. 7.1 are defined as density functionals of unknown analytical form and they are evaluated using approximants. In the hydrogen bonded complexes, the electrostatic part of the orbital-free embedding potential dominates and the errors in the part evaluated using approximants can be expected to play smaller relative role. As a consequence, the errors in the spectral shifts can be expected to be smaller than the ones in the absolute excitation energies. Indeed, numerical results confirm this [52, 127]. The absolute values of the excitation energies in the complex are as good as their counterparts for the isolated component and can vary depending on the choice for the approximant for the exchange-correlation potential and its functional derivative. The applied approach is by construction only applicable for excitations localized in the chromophore [58]. Otherwise, the non-dynamical (frequency dependent) response of the electron density in the environment has to be taken into account. In the case of localized excitations, the embedding strategy offers a serious advantage over a possible alternative, in which the LR-TDDFT calculations are applied for the whole system including the chromophore and the environment. Such calculations not only scale unfavorably with the size of the environment but are known to suffer from artificial low lying charge-transfer excitations [54, 55, 56, 57]. The importance of such spurious excitations increases with increasing size of the system. Restricting the orbital space to the confined system provides thus a practical remedy for this problem.

The applied scheme to evaluate the spectral shifts in the case of a structurally rigid environment of a chromophore is based on a number of approximations/simplifications of various nature. Some of them have clear physical interpretation, others are due to the use of approximants for the density functionals defined in the exact embedding theory, and some are of purely technical character. They are summarized below:

- NDRE - neglect of the dynamic response in the environment.

  By construction, neither in solving Eq. 7.4 nor in solving response equations in LR-TDDFT, the electron density ρ_B is allowed to vary. It is fixed (frozen) at some initial value. As the consequence, the dynamic response of the environment to the electron excitation is neglected. Only virtual embedded orbitals (ϕ_A in Eq. 7.4) which are localized on chromophore are used to construct its response. Therefore, only interfaces in which the environment does not absorb in the considered spectral range can be accurately described. In the considered cases, NDRE is an adequate approximation as confirmed in our dedicated analysis of complexation induced shifts in similar systems [127]. NDRE can be lifted by using the full framework combining the subsystem formulation of DFT with LR-TDDFT [51]. In practice, the most efficient way to lift this simplification is to allow for coupled responses [58].

- The use of the same geometry of the solute with- and without the environment.

  It was already shown in previous studies, that influence of the solvent on the geometry and absorption spectrum of the solute can be very important [169, 170]. For the examples provided in this work, the solvatochromic shifts were evaluated assuming the same geometry of the solute in each of the considered solvents. This is obviously an approximation. Dedicated studies on solvatochromic shifts in a representative sample of the chromophores indicate that this effects affects the shifts in vertical excitation energy on average by 20% [171]. This bring us to another approximation made commonly in the simulations of the solvatochromic shifts, where it is assumed that the shifts in the maxima of the absorption bands from the gas phase to a solvent (or between two different solvents) is the same as the shifts in the vertical excitation energy. In this work, the same assumption is made (for detailed analysis see Ref. [172]).

- The use of approximants instead of the exact density functionals.

  In practical calculations, the kinetic end exchange-correlation components of the effective potential in Eq. 7.4 and of the response kernel are obviously not evaluated exactly but by means of some approximants to these quantities. The choices for the approximants made for these quantities were tested previously for similar systems [127, 52]. The term $v_t^{nad}[\rho_A, \rho_B](\vec{r})$ in Eq. 7.4 vanishes with the increasing distance from chromophore. The applied approximant for this term satisfies this exact condition. It can be expected, therefore, that the asymptotic behavior of the embedding potential used in practical calculations is correct. Note that the electrostatic terms in Eq. 7.1 are exact. In our previously reported study on the DNA base pairs [52], it was shown that maximum deviation in the shifts of excitation energies obtained applying embedding strategy does not exceed 0.05 eV when compared to supermolecular approach. The respective deviation with the experimental result was recently established to not exceed 0.04 eV [127]. The magnitude of such discrepancies is, however, smaller than the magnitude of the solvatochromic shifts in most of the cases analyzed in the present work. For these reasons we will not discuss solvent induced shifts of absorption bands smaller than these values.

- Numerical implementation.

  All the calculations are made using finite basis sets, numerical grid integrations, and the iterative procedure to diagonalize the response matrix in LR-TDDFT. The choices made for these numerical parameters (see the "computational details section" are based on the accumulated numerical experience with the used solvent and are expected to affect the calculation spectral shifts less than the aforementioned approximants and approximations.

The overall accuracy of the solvatochromic shifts is not only determined by the above factors, which determine the accuracy of spectral shifts at a rigid environment, but also on the quality of the description of the statistically fluctuating solvent. In particular, it is affected by the main approximation of this work given in Eq. 7.7. The factors affecting the quality of the average electron density of the environment ($\langle \rho_B(\vec{r}) \rangle$) used in Eq. 7.7 will be overviewed in the subsequent sections.

7.2.2 Classical site distributions from the 3D-RISM-KH theory

The classical site distributions for a molecular solvent around a solute of arbitrary shape are obtained by using the 3D-RISM-KH molecular theory of solvation [154, 173]. The 3D-RISM integral equation can be derived from the six-dimensional, molecular Ornstein-Zernike integral equation [174] for the solute-solvent correlation functions by averaging out the orientation degrees of freedom of solvent molecules while keeping the orientation of the solute macromolecule described at the three-dimensional level [154, 173]. It has the form

$$h^{\mathrm{uv}}_{\gamma}(\vec{r}) = \sum_{\alpha} \int d\vec{r}' c^{\mathrm{uv}}_{\alpha}(\vec{r} - \vec{r}') \chi^{\mathrm{vv}}_{\alpha\gamma}(r'), \tag{7.8}$$

where $h^{\mathrm{uv}}_{\alpha}(\vec{r})$ is the 3D total correlation function of solvent site γ around the solute macromolecule (the superscripts "u" and "v" denoting the solute and solvent, respectively) giving the normalized deviation of the solvent density from its bulk value, which is related to the 3D solute-solvent site distribution function $g_{\gamma}(\vec{r}) = h_{\gamma}(\vec{r}) + 1$, and $c^{\mathrm{uv}}_{\alpha}(\vec{r})$ is the 3D direct correlation function representing "direct" correlations between the solute and solvent site γ, which has the long-range asymptotics of the 3D solute-solvent site interaction potential: $c^{\mathrm{uv}}_{\gamma}(\vec{r}) \sim -\beta u^{\mathrm{uv}}_{\gamma}(\vec{r})$, where $\beta = 1/k_{\mathrm{B}}T$ is the inverse temperature with the Boltzmann constant k_{B}. The site-site susceptibility of pure solvent $\chi^{\mathrm{vv}}_{\alpha\gamma}(r) = \omega^{\mathrm{vv}}_{\alpha\gamma}(r) + \rho^{\mathrm{v}}_{\alpha} h^{\mathrm{vv}}_{\alpha\gamma}(r)$ gives the response, in terms of the distributions in pure bulk solvent, of site γ to the presence of site α at separation r. It consists of the intramolecular matrix $\omega^{\mathrm{vv}}_{\alpha\gamma}(r) = \delta(r - l^{\mathrm{vv}}_{\alpha\gamma})/(4\pi(l^{\mathrm{vv}}_{\alpha\gamma})^2)$ specifying the intramolecular correlations of solvent molecules with the geometry given by the z-matrix of site separations $l^{\mathrm{vv}}_{\alpha\gamma}$, and the intermolecular part given by the bulk solvent site number density $\rho^{\mathrm{v}}_{\alpha}$ times the site-site radial correlation functions of pure bulk solvent $h^{\mathrm{vv}}_{\alpha\gamma}(r)$. The latter is obtained in advance to the 3D-RISM calculations from the dielectrically consistent

RISM theory (DRISM) developed by Perkyns and Pettitt [175] which provides a consistent description of the dielectric properties for ions in polar solvent. The DRISM integral equation is complemented with the KH closure approximation (see below). The DRISM-KH theory ensures appropriate treatment of systems with strongly associating components, such as aqueous electrolyte solutions and solvent/co-solvent mixtures, in a wide range of concentrations [154, 176, 177, 178, 179, 180, 181, 182, 183, 184]. While yielding the dielectric properties of polar solvent around the solute macromolecule at the level of a macroscopic dielectric constant, the bulk solvent susceptibility $\chi_{\alpha\gamma}(\vec{r})$ obtained from the DRISM-KH equations and then inserted in the 3D-RISM equation 7.8 properly accounts for short-range solvation structure effects, such as chemical specificities of solute and solvent molecules, hydrogen bonding, hydrophobicity, steric effects, nanoconfinement, etc.

The convolution in the 3D-RISM Eq. 7.8 (as well as that in the DRISM integral equation) is calculated as a product in the reciprocal space. In calculating the forward and backward Fourier transforms between the direct and reciprocal space by using the 3D fast Fourier transform (3D-FFT) technique, the long-range electrostatic parts of the direct and total correlation functions are separated out and handled analytically for both the radial solvent-solvent and 3D solute-solvent correlation functions to ensure their proper asymptotics. For ionic or highly polar solutes, simply truncating the electrostatic asymptotics would lead to catastrophic errors. Using the Ewald summation in a 3D periodic supercell causes huge distortions in the solvation structure and thermodynamics, but the corresponding analytical corrections restore the proper non-periodic asymptotics of the 3D site correlation functions and accurately eliminate the errors in the solvation chemical potential [185, 186, 154]. Here we use a somewhat different, non-periodic version of the analytical account for the electrostatic asymptotics that we had implemented in the 3D-RISM-KH coupled with KS-DFT in the ADF program package [187]. The 3D solute-solvent site interaction potential $u_{\gamma}^{\mathrm{uv}}(\vec{r})$ and the 3D site correlation functions c_{γ}^{uv} and h_{γ}^{uv} are specified on a 3D linear grid in a nonperiodic rectangular box. The electrostatic asymptotics are specified analytically and subtracted from c_{γ}^{uv} and h_{γ}^{uv} before applying the 3D-FFT to the remaining short-range parts and then added back after the transform. The 3D box has to be large enough to to have the short-range parts of the 3D site correlation functions around the macromolecular solute decay at its boundaries. It has to be padded with zeros to double size in each dimension when performing the 3D-FFT to prevent aliasing, but can be truncated back to the original size in the direct space after calculating the product $c^{\rm uv}_\gamma(\vec{r})\chi^{\rm vv}_{\alpha\gamma}(k)$ for the convolution in Eq. 7.8 and performing the backward 3D-FFT. The electrostatic asymptotics of the 3D direct correlation function is given by that of the 3D interaction potential, $c^{\rm uv(as)}_\gamma(\vec{r}) = -\beta u^{\rm uv(as)}_\gamma(\vec{r})$. (The asymptotics $c^{\rm uv(as)}_\gamma(\vec{r})$ and $u^{\rm uv(as)}_\gamma(\vec{r})$ cancel out in the closure relation (7.14) - see below, and the equations are in fact being solved for the short-range part of $c^{\rm uv}_\gamma(\vec{r})$.) It is chosen as the potential of the site charges gauss-smeared with half-width η which has a convenient form in both the direct and reciprocal space (with the suppressed singularity at $r \to 0$ and gaussian decay for large k),

$$c^{\rm uv(as)}_\gamma(\vec{r}) \;=\; -\beta \sum_i \frac{\tilde{Q}^{\rm u}_i q^{\rm v}_\gamma}{|\vec{r}-\vec{R}_i|} \operatorname{erf}\Big(\frac{|\vec{r}-\vec{R}_i|}{\eta}\Big), \tag{7.9}$$

$$c^{\rm uv(as)}_\gamma(\vec{k}) \;=\; -\beta \sum_i \frac{4\pi\tilde{Q}^{\rm u}_i q^{\rm v}_\gamma}{k^2} \exp\Big(-\frac{1}{4}k^2\eta^2 + i\vec{k}\cdot\vec{R}_i\Big), \tag{7.10}$$

where $\tilde{Q}^{\rm u}_i$ are point-like partial site charges representing the charge distribution of both the solute electrons and atomic cores, and the smearing parameter is set as $\eta \approx 1$ Åto ensure smoothness of the form (7.9) within a molecular core and quick decay of the form (7.10) with k. Substitution of the long-range $c^{\rm uv(as)}_\gamma(\vec{k})$ given by (7.10) into the 3D-RISM integral equation (7.8) yields the renormalized, long-range $\tilde{h}^{\rm uv(as)}_\gamma(\vec{k}) = \sum_\alpha c^{{\rm uv}(as)}_\alpha(\vec{k})\chi^{\rm vv}_{\alpha\gamma}(k)$ which has an involved form rather not amenable to analytical transformation to the direct space. However, it suffices instead to introduce a simple analytical form but with the same asymptotic behavior. The electrostatic asymptotics of the 3D total correlation function $h^{\rm uv}_\gamma$ represents the distribution of solvent site charges resulting in dielectric screening of the solute charge density by solvent polar molecules and Debye screening by solvent ions. Thus, it is dipole-dipole for a polar solute in polar solvent, ion-dipole for an ionic solute in polar solvent, and Debye-screened ion-ion for an ionic solute in electrolyte solution. The latter cannot be neglected in the case of electrolyte solution at medium and especially low ionic concentration, and the analytical electrostatic asymptotics subtracted from $h^{\rm uv}_\gamma$ before applying the backward 3D-FFT and then added back are thus specified as

$$h^{\rm uv(as)}_\gamma(\vec{k}) \;=\; -\beta \sum_i \frac{4\pi\tilde{Q}^{\rm u}_i q^{\rm v}_\gamma}{k^2+\kappa_{\rm D}^2} \exp\Big(-\frac{1}{4}k^2\eta^2 + i\vec{k}\cdot\vec{R}_i\Big), \tag{7.11}$$

$$
\begin{aligned}
h_\gamma^{\mathrm{uv(as)}}(\vec{r}) \;\; = \;\; & -\beta \sum_i \frac{\tilde{Q}_i^{\mathrm{u}} q_\gamma^{\mathrm{v}}}{\varepsilon |\vec{r} - \vec{R}_i|} \exp\left(-\frac{1}{4}\kappa_{\mathrm{D}}^2 \eta^2\right) \\
& \times \frac{1}{2} \left\{ \exp\left(-\kappa_{\mathrm{D}} |\vec{r} - \vec{R}_i|\right) \left[1 - \mathrm{erf}\left(\frac{\kappa_{\mathrm{D}}\eta}{2} - \frac{|\vec{r} - \vec{R}_i|}{\eta}\right)\right] \right. \\
& \quad \left. - \exp\left(\kappa_{\mathrm{D}} |\vec{r} - \vec{R}_i|\right) \left[1 - \mathrm{erf}\left(\frac{\kappa_{\mathrm{D}}\eta}{2} + \frac{|\vec{r} - \vec{R}_i|}{\eta}\right)\right] \right\}
\end{aligned}
\tag{7.12}
$$

where ε in the dielectric constant of the polar solvent (specified as an input parameter into the DRISM theory for bulk solvent), and $\kappa_{\mathrm{D}} = \left(4\pi\beta \sum_j \rho_j^{\mathrm{v}} q_j^{\mathrm{v}} 2/\varepsilon\right)^{1/2}$ is the inverse Debye length of the electrolyte solution with ionic species j of concentrations ρ_j^{v}.

The 3D-RISM integral equation 7.8 must be complemented with a closure relating the total and direct correlation functions. Although the exact closure to the Ornstein-Zernike (OZ) integral equation of liquid state theory is known, it has the form of an extremely complex non-local functional of the total correlation function [174]. Therefore, it is replaced in practice with approximations adequately representing the physics and/or chemistry phenomena occurring in the system, among the most popular being the so-called hypernetted chain (HNC) closure, mean-spherical approximation (MSA), Percus-Yevick (PY) closure, Verlet closure, Martynov-Sarkisov (MS) closure, and others [174]. It is not trivial difficulty to represent, with just one of these approximations, both short-range features of solvation structure (such as steric effects and excluded volume of the repulsive cores) and long-range effects (association due to chemical specificities in complex liquids and long-range fluctuations in near-critical fluids and liquid mixtures near separation lines). To overcome such difficulties, a renormalization approach was developed in which the interaction potential between the particles is subdivided into a ”reference” subsystem amenable to an accurate treatment with one of the closures (for example, liquid of hard-sphere or specially defined repulsive cores) and “the rest” of the interaction potential, typically the long-range part of Lennard-Jones and the electrostatics, and the OZ integral equation is re-written accordingly. The resulting combined closure provides a considerable improvement, for example, the reference HNC closure (RHNC) [174] very accurately describes a system of Lennard-Jones particles. However, a problem occurs for a strong crossover between short- and long-range features of the solvation structure: repulsive core exclusion areas, association peaks at different distances, and long-range attraction tails of the distribution functions; unfortunately, this is typically the case for complex liquids and liquid mixtures, and a huge variety of macromolecules and nanostructures in solution.

The closures to the RISM and 3D-RISM integral equations for the site-site and 3D site correlations are constructed in analogy with the above closure approximations to the OZ integral equation for simple liquids. The RISM-PY theory pioneered by Chandler and Andersen [188] was successful in describing hard-body molecular fluids. Hirata and co-workers extended the description to polar molecular liquids and electrolyte solutions by proposing the RISM-HNC theory [189, 190, 191]. The 3D-RISM-HNC theory was conceptually sketched by Chandler and co-workers in their derivation of density functional theory for classical site distributions of molecular liquids [192, 193] and then introduced in that way by Beglov and Roux for polar liquids [194]. Kovalenko and Hirata derived the 3D-RISM integral equation from the six-dimensional molecular Ornstein-Zernike integral equation [174] for the solute-solvent correlation functions by averaging out the orientation degrees of freedom of solvent molecules while keeping the orientation of the solute macromolecule described at the three-dimensional level [154, 173].

The 3D analogue of the so-called hypernetted chain (HNC) closure to the 3D-RISM equation 7.8 is constructed as

$$g_{\gamma}^{\mathrm{uv}}(\vec{r}) = \exp(-\beta u_{\gamma}^{\mathrm{uv}}(\vec{r}) + h_{\gamma}^{\mathrm{uv}}(\vec{r}) - c_{\gamma}^{\mathrm{uv}}(\vec{r})), \tag{7.13}$$

where $u_{\gamma}^{\mathrm{uv}}(\vec{r})$ is the 3D interaction potential between solvent site γ and the whole solute. Beglov and Roux [194] have obtained the 3D-RISM/HNC equations, Eqs. 7.8 and 7.13, within the density functional method by reduction of the generalized closure of Chandler, McCoy, and Singer for nonuniform polyatomic systems [195]. In the case of a relatively strong attractive potential between the solute and solvent sites, the 3D-HNC closure 7.13 can become divergent [173, 185, 186, 154].

A closure approximation appropriate and successful for the description of the solvation structure and thermodynamics of various inorganic and organic solutes and macromolecules with multiple partial charges in different non-polar and polar liquids, mixtures, and electrolyte solutions, as well as solid-liquid interfaces has been proposed by Kovalenko and Hirata (KH approximation) [173, 154]:

$$
\begin{gathered}
g_\gamma^{\mathrm{uv}}(\vec{r}) = \begin{cases} \exp\bigl(\mathcal{X}_\gamma^{\mathrm{uv}}(\vec{r})\bigr) & \text{for} \quad \mathcal{X}_\gamma^{\mathrm{uv}}(\vec{r}) \le 0 \\ 1 + \mathcal{X}_\gamma^{\mathrm{uv}}(\vec{r}) & \text{for} \quad \mathcal{X}_\gamma^{\mathrm{uv}}(\vec{r}) > 0 \end{cases}, \\
\mathcal{X}_\gamma^{\mathrm{uv}}(\vec{r}) = -\beta u_\gamma^{\mathrm{uv}}(\vec{r}) + h_\gamma^{\mathrm{uv}}(\vec{r}) - c_\gamma^{\mathrm{uv}}(\vec{r}).
\end{gathered}
\tag{7.14}
$$

Eq. 7.14 combines the exponential HNC approximation for the regions of depletion of the distribution function, $g_\gamma^{\mathrm{uv}}(\vec{r}) < 1$, the mean spherical approximation (MSA) for the regions of enrichment, $g_\gamma^{\mathrm{uv}}(\vec{r}) > 1$, with the function and its first derivative continuous at the joint point $\mathcal{X}_\gamma^{\mathrm{uv}}(\vec{r}) = 0$. The 3D-KH approximation 7.14 enforces the proper long-range asymptotics of the direct correlation function $c_\gamma^{\mathrm{uv}}(\vec{r})$ in the same way as the original HNC and MSA closures. The MSA-type linearization prevents the artifact of the distribution function diverging in the regions with a large potential. This partial linearization somewhat reduces and widens high peaks of the distribution functions, whereas it much less affects the coordination numbers of the solvation shells.

The 3D-HNC approximation 7.13 leads to the excess chemical potential of solvation in the closed analytical form [173, 154]

$$
\Delta\mu^{HNC} = k_{\mathrm{B}}T \sum_\gamma \rho_\gamma^{\mathrm{v}} \int d\vec{r} \left[\frac{1}{2}(h_\gamma^{\mathrm{uv}}(\vec{r}))^2 - \frac{1}{2} h_\gamma^{\mathrm{uv}}(\vec{r}) c_\gamma^{\mathrm{uv}}(\vec{r}) - c_\gamma^{\mathrm{uv}}(\vec{r})\right] \tag{7.15}
$$

equivalent to that derived by Singer and Chandler for the site-site RISM/HNC equations [196]. For the 3D-KH closure 7.14, the analytical expression is somewhat different [173, 154],

$$
\Delta\mu^{KH} = k_{\mathrm{B}}T \sum_\gamma \rho_\gamma^{\mathrm{v}} \int d\vec{r} \left[\frac{1}{2}(h_\gamma^{\mathrm{uv}}(\vec{r}))^2 \Theta(-h_\gamma^{\mathrm{uv}}(\vec{r})) - \frac{1}{2} h_\gamma^{\mathrm{uv}}(\vec{r}) c_\gamma^{\mathrm{uv}}(\vec{r}) - c_\gamma^{\mathrm{uv}}(\vec{r})\right], \tag{7.16}
$$

where Θ in the Heaviside step function which puts the term h^2 in effect in the regions of density depletion only.

Accurate calculation of the integral in the excess chemical potential of solvation (7.16) (or equally in (7.15)) requires analytical treatment of the asymptotics of the 3D correlation functions (7.9) and (7.12). In the last term in the expression (7.16), the electrostatic asymptotics of $c_\gamma^{\mathrm{uv}}(\vec{r})$ cancel out upon the summation over the site index γ because of the electroneutrality condition. With the short-range parts of the 3D correlation functions decaying at the 3D box boundaries, the asymptotic forms of the first two terms in (7.16) are subtracted from it and then added back, forming the short-range and long-range integrands, the former integrated numerically over the 3D box volume V and the latter containing just the analytical asymptotics integrated over the whole space,

$$\begin{aligned}
\Delta\mu^{KH} &= k_{\mathrm{B}}T\sum_{\gamma}\rho^{\mathrm{v}}_{\gamma}\int\limits_{V} d\vec{r}\left[\frac{1}{2}(h^{\mathrm{uv}}_{\gamma}(\vec{r}))^2\Theta(-h^{\mathrm{uv}}_{\gamma}(\vec{r})) - \frac{1}{2}h^{\mathrm{uv}}_{\gamma}(\vec{r})c^{\mathrm{uv}}_{\gamma}(\vec{r}) - c^{\mathrm{uv}}_{\gamma}(\vec{r})\right.\\
&\qquad\left. -\frac{1}{2}(h^{\mathrm{uv(as)}}_{\gamma}(\vec{r}))^2\Theta(Q^{\mathrm{u}}q^{\mathrm{v}}_{j}) + \frac{1}{2}h^{\mathrm{uv(as)}}_{\gamma}(\vec{r})c^{\mathrm{uv(as)}}_{\gamma}(\vec{r})\right]\\
&+ k_{\mathrm{B}}T\sum_{\gamma}\rho^{\mathrm{v}}_{\gamma}\int d\vec{r}\left[\frac{1}{2}(h^{\mathrm{uv(as)}}_{\gamma}(\vec{r}))^2\Theta(Q^{\mathrm{u}}q^{\mathrm{v}}_{j}) - \frac{1}{2}h^{\mathrm{uv(as)}}_{\gamma}(\vec{r})c^{\mathrm{uv(as)}}_{\gamma}(\vec{r})\right],
\end{aligned}\tag{7.17}$$

where $Q^{\mathrm{u}} = \sum_i Q^{\mathrm{u}}_i$ is the total charge of the solute, and $q^{\mathrm{v}}_j = \sum_{\gamma\in j} q^{\mathrm{v}}_{\gamma}$ is the total charge of solute ionic species j which site γ belongs to. The factor $\Theta(Q^{\mathrm{u}}q^{\mathrm{v}}_j)$ switches the asymptotics term $(h^{\mathrm{uv(as)}}_{\gamma}(\vec{r}))^2$ on for like charges of the solute and solvent ion Q^{u} and q^{v}_j, and off for unlike ones, according to the term $(h^{\mathrm{uv}}_{\gamma}(\vec{r}))^2\Theta(-h^{\mathrm{uv}}_{\gamma}(\vec{r}))$ being switched on at long range in the excess chemical potential (16) by the depleted long-range tale of $h^{\mathrm{uv}}_{\gamma}(\vec{r})$ for like charges and off by its enhanced long-range tale for unlike charges. The factor $\Theta(Q^{\mathrm{u}}q^{\mathrm{v}}_j)$ is dropped for the excess chemical potential (15) in the 3D-HNC approximation (13). In the expression (17), the latter integral of the asymptotics over the whole space is analytically reduced to 1D integrals easy to compute,

$$\begin{aligned}
&k_{\mathrm{B}}T\sum_{\gamma}\rho^{\mathrm{v}}_{\gamma}\int d\vec{r}\left[\frac{1}{2}(h^{\mathrm{uv(as)}}_{\gamma}(\vec{r}))^2\Theta(Q^{\mathrm{u}}q^{\mathrm{v}}_{j}) - \frac{1}{2}h^{\mathrm{uv(as)}}_{\gamma}(\vec{r})c^{\mathrm{uv(as)}}_{\gamma}(\vec{r})\right]\\
&= \frac{1}{\pi}\sum_{\gamma\in\mathrm{v}}\frac{\kappa^2_{\gamma}}{\kappa_{\mathrm{D}}}\int\limits_{0}^{1} dx \sum_{ii'\in\mathrm{u}} Q^{\mathrm{u}}_i Q^{\mathrm{u}}_{i'}\, J_0(\kappa_{\mathrm{D}} l^{\mathrm{uu}}_{ii'} x)\,\exp\Big(-\frac{1}{2}(\kappa_{\mathrm{D}}\eta x)^2\Big)\\
&\qquad\times\left[\frac{x^2}{\varepsilon\,(x^2+(1-x)^2)^2} - \frac{1}{x^2+(1-x)^2}\right],
\end{aligned}\tag{7.18}$$

where $\kappa^2_{\gamma} = 4\pi\beta\rho^{\mathrm{v}}_{\gamma}q^{\mathrm{v}}_{\gamma}q^{\mathrm{v}}_{j}/\varepsilon$ is the partial contribution of site γ of solution ionic species j to the Debye length squared, $J_0(x)$ is the zero-order Bessel function of the first kind, and $l^{\mathrm{uu}}_{ii'}$ are the site separations between the solute point-like partial site charges $\tilde{Q}^{\mathrm{u}}_i$.

7.2.3 Self-consistent combination of the orbital-free embedding potential and the 3D-RISM-KH site distributions

The principal goal of the present work is the calculation the average solvent dependent shifts in the absorption spectra. To this end, we construct the embedding potential defined in Eq. 7.1, which is determined by the averaged electron density of the solvent and use Eq. 7.7 to derive the solvent-induced spectroscopic shifts. The 3D-RISM-KH site distributions cannot be directly used to evaluate the orbital-free embedding potential, as they represent not the electron densities needed in Eq. 7.1 but the classical nuclear probabilities. To combine the 3D-RISM-KH method with orbital-free embedding framework, two issues must be, therefore, resolved: *i*) the procedure to "dress up" the site densities with electrons and *ii*) the method to take into account the solute-solvent interactions in generating the classical 3D-RISM-KH site distributions.

"Dressing up" classical site distributions in electron density

The 3D-RISM-KH method is used to obtain the equilibrium density distributions of their classical interaction sites γ. Such distributions do not suffice to evaluate the orbital-free embedding potential as they are essentially related to nuclei and not electrons. The second key new approximation introduced in this work, besides that in Eq. 7.7, is the applied procedure to "dress up" the site distributions with electrons.

First of all, it is assumed that the electron density attributed to each site ($q_\gamma^{\mathrm{v}}(\vec{r})$) is "rigid". This assumption was used already in the first application of the orbital-free embedding potential in simulating solvated system [159] where it was assumed that a "rigid" electron density cloud moves together with the nuclei. In other words, the inhomogeneous average electron density of solvent around the solute molecule is obtained as the convolution of the inhomogeneous probability density of the solvent nuclei and the electron density around each of them. Note that the electron density around each nucleus is assumed to be translationally and rotationally invariant. Therefore, inhomogeneities of the average solvent charge density arise due to inhomogeneities of the nuclear distributions. Such construction of $\langle\rho_B(\vec{r})\rangle$ neglects, therefore, any instantaneous fluctuations of the electron density of each solvent molecule due to fluctuations of its induced dipole moments. Note, however, that the "rigid" density used for each solvent molecule does not correspond to the gas phase but to the liquid. The averaged electron density of the solvent is obtained thus as:

$$\langle \rho_B(\vec{r}) \rangle = \sum_{\gamma} \int d\vec{r}' q_{\gamma}^{\mathrm{v}}(|\vec{r} - \vec{r}'|) \rho_{\gamma}^{\mathrm{v}} g_{\gamma}^{\mathrm{uv}}(\vec{r}'), \tag{7.19}$$

where the angle brackets $\langle \ldots \rangle$ denote the statistical ensemble average, and $g_{\gamma}^{\mathrm{uv}}(\vec{r})$ is the 3D solute-solvent site distribution function obtained from the 3D-RISM-KH theory described in the previous section.

Such simplified construction of the electron density of the environment was used in a number of our subsequent applications in liquids [53, 86]. In the present work, a further simplification is introduced: the "rigid" electron densities are orientationally averaged, spherically symmetric distributions centered at atoms (O, H in the case of water, for instance) or groups of atoms (O, H, and CH_3, in the case of methanol, for instance). The third approximation made in the procedure to dress-up the RISM site distributions concerns the evaluation of the convolution (Eq. 7.19). We note that the functions $q_{\gamma}^{\mathrm{v}}(\vec{r})$ are short ranged whereas the RISM grid at which the site distributions are evaluated is large (0.5 Å). Therefore, to a good approximation, the integration can be replaced by summation involving only diagonal elements, with the weights corresponding to the number of electrons associated to each site (8.8 e for oxygen in water and 0.6 e for a hydrogen in water for instance). This assumption means that the weights represent the average charge distribution in the solvent molecule which is not affected by the instantaneous configuration of the solvent. Lifting this simplification by using an orientation-dependent charge density $q_{\gamma}^{\mathrm{v}}(r, \Omega)$ would involve the orientation-dependent site distributions $g_{\gamma}^{\mathrm{uv}}(\vec{r}, \Omega)$. As a consequence, solving 3D-RISM equations would become significantly more involved. It is important to underline that the average charge distribution differs, however, from its gas-phase counterpart.

The three above assumptions lead to the following expression for $\langle \rho_B(\vec{r}) \rangle$ at the 3D-RISM grid, which reads:

$$\langle \rho_B(\vec{r}) \rangle = \sum_{\gamma} q_{\gamma}^{\mathrm{v}} \rho_{\gamma}^{\mathrm{v}} g_{\gamma}^{\mathrm{uv}}(\vec{r}). \tag{7.20}$$

Solute-solvent interactions in the 3D-RISM-KH approach

To account for the solute-solvent coupling in the 3D-RISM-KH calculations aimed at obtaining the site distributions g^{uv}, we adopt the same approach as the one introduced in our previous work on embedding a Kohn-Sham system in the average solvent potential generated by the 3D-RISM-KH method [187, 197]. Here, the solute-solvent coupling is represented in a classical way. Its basic features are outlined below.

The use of the average electron density of the solvent, i.e.,$\langle\rho_B(\vec{r})\rangle$ in Eq. 7.4, is based on the generalization of the Hohenberg-Kohn energy functional to the ensembles of solvent conformations. This can be made similarly to the case of density functional theory for ensembles of electronic states in which ensemble operators and mean values for the system quantities are defined [7, 198]. The whole system is thus described by means of the Helmholtz free energy functional

$$\langle A\rangle[\langle\rho_A\rangle, \{\rho_\gamma^{\mathrm{v}}\}] = \langle E_A\rangle[\langle\rho_A\rangle, \langle\rho_B\rangle] + \Delta\mu_{\mathrm{solv}}^{\mathrm{KH}}[\{\rho_\gamma^{\mathrm{v}}\}] \tag{7.21}$$

of the mean electronic density of the embedded molecule $\langle\rho_A(\vec{r})\rangle$ and the set of the classical 3D solvent site density distributions $\rho_\gamma^{\mathrm{v}}(\vec{r}) = \rho_\gamma^{\mathrm{v}} g_\gamma^{\mathrm{uv}}(\vec{r})$ for all solvent sites γ. At this stage, assuming the coupling between the embedded subsystem A and environment B to be weak enough we apply the same above-described OFE functionals to the mean densities $\langle\rho_A(\vec{r})\rangle$ and $\langle\rho_B(\vec{r})\rangle$. For simplicity, we will drop the ensemble averaging brackets in all notations below, keeping in mind that all quantities are mean values averaged over the ensemble of the environment.

The environment electron density obtained from the 3D-RISM theory in the form 7.20 can be used right away to calculate the non-electrostatic components $v_{xc}^{nad}[\rho_A, \rho_B](\vec{r})$ and $v_{\mathrm{t}}^{nad}[\rho_A, \rho_B](\vec{r})$ of the embedding potential $v_{\mathrm{emb}}^{\mathrm{KSCED}}[\rho_A, \rho_B; \vec{r}]$ defined in Eq. 7.1. Furthermore, the electrostatic terms in the embedding potential $v_{\mathrm{emb}}^{\mathrm{KSCED}}[\rho_A, \rho_B; \vec{r}]$ explicitly dependent on the positions of the environment nuclei in the original OFE are replaced in our ensemble approach by the statistical-mechanical average of the electrostatic potential of solvent sites acting on the solute, defined as a variational derivative of the system free energy 7.21 with respect to the embedded density $\rho_A(\vec{r})$:

$$v_{\mathrm{solv}}^{\mathrm{elec}}(\vec{r}) \equiv \frac{A[\rho_A, \{\rho_\gamma^{\mathrm{v}}\}]}{\delta\rho_A(\vec{r})} = \sum_\gamma \rho_\gamma^{\mathrm{v}} \int d\vec{r}' v_\gamma^{\mathrm{el}}(|\vec{r} - \vec{r}'|) h_\gamma^{\mathrm{uv}}(\vec{r}'), \tag{7.22}$$

where $v_\gamma^{\mathrm{el}}(\vec{r})$ is the electrostatic potential created by the site electronic charge density $q_\gamma^{\mathrm{v}}(\vec{r})$, and $h_\gamma^{\mathrm{uv}}(\vec{r}) = g_\gamma^{\mathrm{uv}}(\vec{r}) - 1$ is the 3D solute-solvent site total correlation function obtained from the 3D-RISM-KH theory. Note that the expression 7.22 is obtained by definition with the variational differentiation of the expression for the solvation chemical potential of the embedded molecule with respect to its electronic density $\rho_A(\vec{r})$ [173].

The 3D-RISM-KH integral equations 7.8 and 7.14 are solved for the 3D solute-solvent site correlation functions describing the classical solvation structure around the embedded solute molecule, using the input of the classical interaction potentials $u^{\mathrm{uv}}_{\gamma}(\vec{r})$ of the solute acting on solvent site γ that enter the closure relation 7.14. In a self-consistent field loop, these interaction potentials are defined by the variational differentiation of the system free energy functional 7.21 with respect to the classical 3D solvent site density distributions $\rho^{\mathrm{v}}_{\gamma}(\vec{r})$. We further break them up into the short-range interaction part represented by the sum of the Lennard-Jones potentials $u^{\mathrm{uv(LJ)}}_{i\gamma}(\vec{r})$ between solvent site γ and all the atoms i centered at positions $\vec{R}_i$ in the embedded solute molecule, the electrostatic interaction potential between the solvent site charge density $q^{\mathrm{v}}_{\gamma}(\vec{r})$ and the solute atomic cores with charges Q^{u}_i at $\vec{R}_i$, and the electrostatic interaction potential between the solvent site and the valence electron density of the embedded molecule $\rho_A(\vec{r})$ [173, 187],

$$\begin{aligned} u^{\mathrm{uv}}_{\gamma}(\vec{r}) \equiv \frac{A[\rho_A, \{\rho^{\mathrm{v}}_{\gamma}\}]}{\delta \rho^{\mathrm{v}}_{\gamma}(\vec{r})} = \sum_i \left(u^{\mathrm{uv(LJ)}}_{i\gamma}(|\vec{r} - \vec{R}_i|) + \int d\vec{r}' \frac{Q^{\mathrm{u}}_i q^{\mathrm{v}}_{\gamma}(|\vec{r}' - \vec{r}|)}{|\vec{r}' - \vec{R}_i|} \right) \quad (7.23) \\ - e \int d\vec{r}' d\vec{r}'' \frac{\rho_A(\vec{r}') q^{\mathrm{v}}_{\gamma}(|\vec{r}'' - \vec{r}|)}{|\vec{r}'' - \vec{r}'|}. \end{aligned}$$

The latter term in Eq. 7.23 is calculated in the density fitting procedure [114]. Following the standard force fields, the solvent site charge densities $q^{\mathrm{v}}_{\gamma}(\vec{r})$ can be represented with point-like partial charges q^{v}_{γ} to simplify this expression to

$$u^{\mathrm{uv}}_{\gamma}(\vec{r}) = \sum_i \left(u^{\mathrm{uv(LJ)}}_{i\gamma}(|\vec{r} - \vec{R}_i|) + \frac{Q^{\mathrm{u}}_i q^{\mathrm{v}}_{\gamma}}{|\vec{r} - \vec{R}_i|} \right) - e \int d\vec{r}' \frac{\rho_A(\vec{r}') q^{\mathrm{v}}_{\gamma}}{|\vec{r}' - \vec{r}|}. \qquad (7.24)$$

Note that the analytical form for the free energy of solvation (Eq. 7.16) and for the coupling potentials (Eq. 7.22 and 7.23) yield the analytical expressions for the derivatives of the system free energy 7.21 with respect to nuclei coordinates R_i of the embedded molecule. As a consequence, efficient algorithms based on analytic gradients can be used in the geometry optimization and reaction pathway studies of solvated systems [187, 197].

Practical applications of Eq. 7.4 involve numerical integration on the grid. The adequate grid used for this purpose extends on both the embedded system and its environment and, similarly to grids used in solving the Kohn-Sham equations, is not uniform. The 3D-RISM-KH calculations use also a grid which is, however, different. The classical site distributions are more homogeneous than the electron density. To facilitate evaluation of the Fourier transforms, it is uniform and cubic. Calculations, which involve

the combination of these two grids, are implemented in our local version of the ADF 2008 program [114, 199] and involve four principal stages: *i*) The OFE module is used to generate the OFE grid. *ii*) The 3D-RISM-KH site distributions are evaluated on the 3D-RISM grid and "dressed up" with electrons. The density $\langle\rho_B(\vec{r})\rangle$ and the corresponding electrostatic potential are evaluated thus on the 3D-RISM grid. The numerical values of these quantities are interpolated on the OFE grid (obtained in stage *i*)) and imported into the 3D-RISM module. *iii*) The interpolated density and potential are imported back to OFE module [37], which solves Eq. 7.4, and obtains all the quantities needed to evaluate the excitation energies following the approach introduced in Ref. [52], i.e., embedded orbitals, their energies, and the contribution to the response kernel due to the orbital-free embedding potential. *iv*) Evaluation of the excitation energies. The modification of the ADF 2008 code allowing for such calculations concern primarily the stage *ii*), which involves "dressing up" the classical site distributions with electrons and interfacing the grid generated at stage *i*) with that needed at stage *ii*). Concerning the solver of Eq. 7.4 and the solver of LR-TDDFT response equations for embedded system the numerical implementation of the methodology introduced in the present work do not require any specific code modifications.

Finally, we point out that difference between the approach taken in the present work and that in Refs. [187, 197] lies in the embedding potential used in quantum mechanical part of the applied computational protocol and not in the evaluation of the classical site distributions which are evaluated in the same way as in Refs. [187, 197]. Compared to the embedding potential of Wesolowski and Warshel [4] (Eq. 7.1), the embedding potential used in the previous studies comprises only the first two electrostatic terms of $v_{emb}^{KSCED}[\rho_A, \rho_B; \vec{r}]$. Such simplification corresponds to a purely electrostatic embedding. The neglected terms $v_{xc}^{nad}[\rho_A, \rho_B](\vec{r})$ and $v_{\mathrm{t}}^{nad}[\rho_A, \rho_B](\vec{r})$, which account for the Pauli repulsion, were shown to be indispensable in many types of embedded systems [52, 200, 84].

7.3 Computational details

7.3.1 3D-RISM-KH

To obtain the solvent site distributions and atomic charges of the solvent which are self-consistent, the Kohn-Sham DFT/3D-RISM-KH scheme introduced and recently implemented by Gusarov et al. [187] in the ADF program package [114] was used. The size of the 3D-FFT grid was chosen to 64 × 64 × 64 points in cell size 32 × 32 × 32 Å (64 × 64 × 128 in cell size 32 × 32 × 64 Å for benzophenone). In the KS-DFT/3D-RISM-KH scheme the partial charges used for the solute are taken from the multipole derived population analysis (labelled as mdc-q in ADF output) [201]. In order to assure the high quality of these charges, STO(TZP) basis set with modified fit functions to give better multipole moments was used [201]. For the exchange-correlation functional LDA approximation [16, 18, 17] was used, integration parameter was set to 6.0. The van der Waals parameters required to solve 3D-RISM-KH equations were taken from OPLS force field [202, 203]. If not stated otherwise, in all calculations the geometry of considered chromophores was optimized in gas-phase in KS-DFT framework using BP86 exchange-correlation functional [19, 21].

7.3.2 Orbital-free embedding

The effective embedding potential given in Eq. 7.1 is evaluated for each instantaneous pair of densities, $\rho_A(\vec{r})$ and $\langle\rho_B(\vec{r})\rangle$, using the following approximants:

- For the $v_t^{nad}[\rho_A, \rho_B](\vec{r})$ component of the embedding potential, its analytic form was obtained using Eq. 7.2 applied to the Thomas-Fermi expression for kinetic energy functional[9, 10].

- For the $v_{xc}^{nad}[\rho_A, \rho_B](\vec{r})$ component of the embedding potential, local density approximation was used for its exchange- [16] and correlation parts [18, 17].

The electrostatic component was evaluated using the monomer expansion of the 3D-RISM potential using all the centers of the 3D-RISM grid (see Eq. 7.22).

The exchange-correlation component of the total effective potential in Eq. 7.4, which is generated by the embedded density ρ_A, i.e., $\frac{\delta E_{xc}[\rho_A]}{\delta \rho_A}$, is approximated using the SAOP potential [204, 205]. Such hybrid treatment of the total exchange-correlation potential is motivated by the fact that SAOP is known to describe better the electronic excitations in the isolated chromophore. Note that SAOP cannot be used for approximating the corresponding component of the orbital-free embedding potential because it depends explicitly on orbitals which are not available for $\rho_A + \rho_B$.

In all calculations, the orbitals of the chromophore are expanded using the Slater type of atomic orbitals (TZ2P basis set) which included only atomic centers localized in the chromophore (the (KSCED(m) variant of the finite basis version of Eq. 7.4 [71]). The integration parameter for the grid generation was set to 6.0.

7.4 Results and discussion

For the following chromophores: aminocoumarin C151, acetone, acrolein, and benzophenone, the solvatochromic shifts are discussed in separate sections below. The lowest lying excitations of the $\pi \rightarrow \pi^*$ and/or $n \rightarrow \pi^*$ type are discussed. The solvatchromism of acetone and amnocoumarin in water, was already studied using the orbital-free embedding strategy [53, 86]. The previous simulations did not use the key approximation introduced in the present work, given in Eq. 7.7, and can be used as the reference for establishing adequacy of this approximation.

The analysis of the significance of the non-electrostatic components of the embedding potential is provided in the separate section.

7.4.1 Aminocoumarin C151

We start the analysis of the solvatochromic shifts for the $\pi \rightarrow \pi^*$ excitations in aminocoumarin C151 (see Figure 7.1). Experimental [206, 207, 208] as well as theoretical studies [209, 210] including previous work applying the embedding potential of Eq. 7.1 [86] indicate that lowest absorption band is of the $\pi \rightarrow \pi^*$ character showing pronounced red-shift in polar solvents.

Aminocoumarin C151 is especially suited for the analysis of the key approximation in the computational approach introduced in Eq. 7.7 of the present work, i.e., the replacement of a large number of calculations spectral

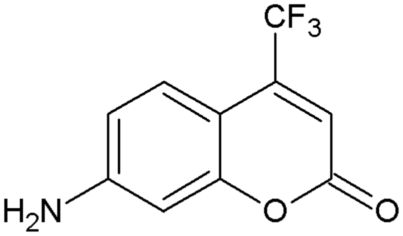


Figure 7.1: Structure of aminocoumarin C151

Table 7.1: Solvatochromic shifts from water to gas phase of aminocumarin $\pi \rightarrow \pi^*$ absorption band calculated using orbital-free effective embedding potential and the two strategies to account for the statistical nature of the solvent: explicit and thorough 3D-RISM. Data taken from Neugebauer et al. [86]. For description see text.

Model of solvent	$\Delta\varepsilon$ $[eV]$
explicit: 300 H_2O, 50 snapshots (Ref. [86])	-0.28
explicit: 300 H_2O, 400 snapshots (Ref. [86])	-0.25
3D-RISM	-0.25
Experiment	-0.22[a]

[a] Ref. [206]

shifts for different configurations of the solvent to obtain averages by only one calculation but using the average embedding potential. In simulations reported in Ref. [86] as well as in the present work, the orbital-free embedding potential given in Eq. 7.1 was used to obtain the shifts between the absorption maxima in the gas phase and in water. Therefore, the solvatochromic shifts obtained following these two strategies can be used to analyze the adequacy of the approximation made in Eq. 7.7.

The spectral shifts were evaluated in Ref. [86], using either 50 or 400 configurations of 300 solvent molecules. The configurations were taken from either one or eight 50 ps long equilibrated molecular dynamic trajectory. The calculated spectral shifts reported in Ref. [86] amount to -0.28 eV and -0.25 eV for the 50 or 400 configuration based statistics, respectively. These values are in remarkable agreement with the OFE/RISM results (see Table 7.1). This agreement calls for a more detailed analysis. The OFE/RISM results could be considered as the ensemble average only if the orbital-free embedding potential was independent on ρ_A. In such a case, Eq. 7.7 would

Table 7.2: Calculated excitation energies and solvatochromic shifts (in [eV]) to the gas phase of $\pi \rightarrow \pi^*$ absorption band in aminocoumarin. The numbers in parentheses correspond to the case when electrostatic-only embedding is used.

Solvent	$\varepsilon_{OFE/RISM}$	$\Delta\varepsilon_{OFE/RISM}$	$\Delta\varepsilon^a_{exp}$
H_2O	2.988 (2.985)	-0.25 (-0.25)	-0.22
Methanol	3.028 (3.029)	-0.21 (-0.21)	-0.31
Diethyl Ether	3.102 (3.103)	-0.13 (-0.13)	-0.21
n-Hexane	3.236 (3.235)	0.00 (0.00)	-
Gas phase	3.234	-	-

[a] Ref. [206]

not be an approximation but an exact expression. Our dedicated studies on the dependence of the orbital-free embedding potential on ρ_A showed that indeed this dependence is weak [164] for the type of intermolecular interactions as the ones occurring in the systems studied here. This justifies treating the OFE/RISM shifts as a target and interpreting the agreements between the shifts obtained as averages over 400 ps long MD trajectory and the OFE/RISM results as the evidence that the trajectory obtained in Ref. [86] represents adequately the statistical ensemble. Note that the OFE/RISM strategy requires only one evaluation of the excitation energies whereas the explicit strategy requires 400 of such evaluations.

The results of the OFE/RISM calculations together with the survey of available experimental data are gathered in Table 7.2. In this table (and in respective tables for other examples considered in the present work), the values in parentheses are obtained using the electrostatic-only embedding, i.e., with the $v_{\text{xc}}^{nad}[\rho_A, \rho_B](\vec{r})$ and $v_{\text{t}}^{nad}[\rho_A, \rho_B](\vec{r})$ terms in Eq. 7.1 neglected. They will be discussed in a separate section of this work.

In all of considered solvents, the calculated lowest-lying excitations have the $\pi \rightarrow \pi^*$ character, are red-shifted compared to the free chromophore, and agree reasonably well with experiment. For water, OFE/RISM overestimates the solvatochromic shift by only 0.03 eV. In the case of the other two considered solvents, 65% (for methanol) and 63% (for dimethyl ether) of the experimental shifts is recovered. Discrepancies between the experimental and calculated shifts can be attributed to several factors. The most probable ones are the errors due to approximants to the density functionals in Eq. 7.1, the neglect of dynamic response of the environment, and the treatment of the statistical nature of the solvent. The agreement between the shifts obtained for aminocoumarin in water in the present work with the previously published data [86] suggest that the last factor is less important. The analyses of Ref. [86] indicate also that NDRE is most likely the main contributor to the error.

The widely used COSMO model [211], is based on the continuum representation of the solvent. It is useful to compare the results obtained with the two levels of description of the solvent: COSMO (Table 7.3) and OFE/RISM (Table 7.2). Similarly to OFE/RISM, the COSMO shifts are in reasonable agreement with experiment. OFE/RISM is clearly superior for the non-polar solvent (n-hexane), whereas the two methods lead to very similar results for more polar solvents.

Neither OFE/RISM nor COSMO describe the fact that the magnitude of the solvatomic shift is smaller in water than in methanol. The failure to describe the larger magnitude of the solvatochromic shifts of the $\pi \rightarrow \pi^*$ transition in aminocoumarin in methanol than in water by mean of either methods remains to be clarified.

Considering practical use of discussed here methods, the LR-TDDFT calculations of isolated aminocoumarin take 2.5h or single core machine, while including solvent by either COSMO or OFE/RISM takes 3h and 7h, respectively. It can be seen that compared to COSMO, OFE/RISM calculations involve about twofold increase in the time of calculations performed using the ADF implementation of these two methods. This indicates that introducing a more realistic level of the description of the solvent by replacing the simplest continuum by non-uniform 3D-RISM site probabilities involves acceptable additional CPU costs.

7.4.2 Acetone

Acetone is commonly used as a model system to test the accuracy of theoretical methods to predict solvatochromic shifts [212, 53, 165, 213, 214, 215, 216, 217, 218, 219, 220, 221]. For the survey and comparison of available computational studies, see Ref. [165]. Acetone is also very well characterized experimentally [222, 223, 224, 225, 226]. For the overview of the experimental data, see the recent analyses by Renge [227]. The previous studies, where the orbital-free embedding strategy to calculate the solvatochromic shifts of acetone in water was also applied but the averaging was made without the approximation introduced in Eq. 7.7, provide excellent reference [53].

Table 7.3: Excitation energies and solvatochromic shifts to the gas-phase of $\pi \rightarrow \pi^*$ absorption band in aminocoumarin calculated with the COSMO model (excitation energies and shifts in [eV]).

Solvent	ε_{COSMO}	$\Delta\varepsilon_{COSMO}$	$\Delta\varepsilon^{a}_{exp}$
H_2O	3.041	-0.19	-0.22
Methanol	3.043	-0.19	-0.31
Diethyl Ether	3.089	-0.14	-0.21
n-Hexane	3.157	-0.08	-
Gas phase	3.234	-	-

[a] Ref. [206]

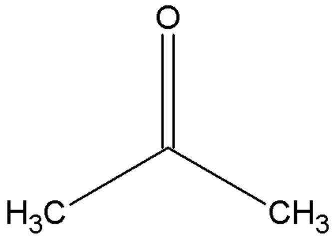


Figure 7.2: Structure of acetone

The experimental studies indicate that the lowest transition in acetone is of the $n \rightarrow \pi^*$ character. The solvatochromic shift in water is well estimated to be between 0.19-0.22 eV [222, 223, 224, 225, 226, 227]. Our results are gathered in Table 7.4.

For all of the considered solvents, OFE/RISM predicts the lowest excitation (HOMO to LUMO) to be the $n \rightarrow \pi^*$ transition. The magnitude of the excitation energy increases with the polarity of the solvent. Comparison of solvatochromic shifts with the experimental results reveals good agreement in all of the cases. For non-polar solvents, like n-hexane, the calculated shift is very small, which corroborates with the recent experimental reports [227]. In the case of the diethyl ether, the experimental result is slightly overestimated by 0.04 eV. For polar solvents the results are also close to the reference data. For methanol, the OFE/RISM combination predicts shift to be equal to 0.12 eV, which is in remarkable agreement wit the most recent experimental studies [227]. In the case of the water, the estimated solvatochromic shift is within the range of the reference values. In summary, the proposed computational approach describes acetone in a proper way.

Table 7.4: Calculated excitation energies and solvatochromic shifts (in [eV]) to the gas phase of $n \rightarrow \pi^*$ absorption band of acetone. The numbers in parentheses correspond to the case when electrostatic-only embedding is used.

Solvent	$\varepsilon_{OFE/RISM}$	$\Delta\varepsilon_{OFE/RISM}$	$\Delta\varepsilon^{a}_{exp}$
H_2O	4.768 (4.727)	0.19 (0.15)	0.22, (0.19 - 0.21^b)
Methanol	4.702 (4.683)	0.12 (0.10)	0.12, (0.15^c)
Diethyl Ether	4.638 (4.616)	0.06 (0.04)	0.02, (0.04^d)
n-Hexane	4.592 (4.585)	0.01 (0.00)	-0.004^d
Gas phase	4.581	-	-

^{a} The most recent re-examination of experiment data provided in Ref. [227] is used as reference. The results of previous experimental studies are given in parentheses: ^{b}Ref. [222, 223, 224, 225, 226] ^{c}Ref. [225] ^{d}Ref. [228].

Recently, Neugebauer et al. [53] applied the orbital-free embedding formalism and explicit solvent level for acetone in water. The instantaneous electron densities of the solvent (about 175 atoms) were calculated for 220 configurations extracted from equilibrated CPMD trajectory and used to evaluate the instantaneous excitation energies. The reported average solvatochromic shift equal to 0.20 eV evaluated in this tedious and laborious protocol requiring considerable amount of computer resources, is in excellent agreement with the shift obtained in the present work. Using 3D-RISM-KH to generate $\langle\rho_B\rangle$ reduces the time of computations to only 25 minutes and leads to the spectral shift (0.19 eV) also in good agreement with experiment. Owing to the approximation introduced in Eq. 7.7, the repetitive generation solvent density for each solvent configurations is avoided. The obtained shifts with- and without the approximation given in Eq. 7.7 agree excellently with the most recent experimental data [227] (0.22 eV). Moreover, the most recent theoretical value for this shift obtained using high-level ab-initio description (CCSD) of the acetone are also very similar (0.20 eV [165]).

In the same work [53], the authors identify and report the $\sigma - \pi^*$ valence transition, which is not known experimentally due to the low intensity, however, calculations reveal strong red-shift in excitation energy when going from gas-phase to water. For instance, Neugebauer et al. report the shift of -0.32 eV for a single configuration of 57 water molecules taken from equilibrated MD trajectory. The authors do not report, however, the averaged value of this shift, making direct comparisons with our OFE/RISM result (-0.23 eV) impossible. In view of this, both calculated shifts appear consistent.

7.4.3 Acrolein

s-trans-acrolein (Fig. 7.3) is a simple aldehyde, for which the lowest transition has the $n \rightarrow \pi^*$ character. The absorption spectra are available [229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240]. Most recent measurements estimate that the absorption energy of acrolein solvated in water is blue-shifted by 0.25 eV with respect to gas-phase [229]. The theoretical studies for this spectrum were reported [241, 242, 243, 244, 245, 246] (for survey see Ref. [229]). In the work by Ten-no [220], where SCF-RISM method was applied to obtain excitation energies, the authors report the change in the absorption band upon solvation to be 0.20 eV, in good agreement with the experiment. The spectral shift obtained in the present study (0.33 eV) overestimates slightly the experimental result.

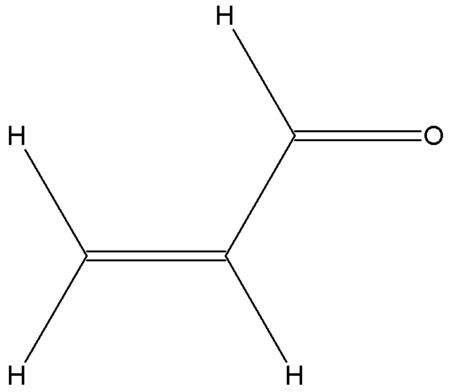


Figure 7.3: Structure of *s-trans*-acrolein

7.4.4 Benzophenone

Benzophenone (see Fig. 7.4) is an interesting molecule for theoretical studies. The experimental UV/Vis absorption spectra [247] is characterized by two main bands - the first one of the $n \rightarrow \pi^*$ character, and the second broad one is of the $\pi \rightarrow \pi^*$ type. Experiment reveals also that, upon solvation, the $n \rightarrow \pi^*$ excitation is blue-shifted, whereas the $\pi \rightarrow \pi^*$ band shows pronounced red-shift.

The OFE/RISM solvatochromic shifts in benzophenone combination are given in Table 7.5. The geometry of the solute was taken from Ref. [248]. The broad $\pi \rightarrow \pi^*$ band originates from four transitions. The intensity weighted average of the calculated absorption lines is, therefore, used for comparisons with experiment. For methanol and diethyl ether, the OFE/RISM describes very well the both blue- and red-shifts in $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ transitions. The deviation from the experimental findings does not exceed 0.02 eV. For water, the calculated shifts are less satisfactory. The OFE/RISM results recover only 59% and 50% of the experimental shifts, in $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ transitions, respectively. This underestimation might originate from both the interpretation of the experimental data or from the many simplifications/approximations in the computational protocol applied in this work. The experimental studies [247] indicate that with the increasing polarity of the solvent the formation of the shoulders in absorption spectrum is observed. In the case of water, this effect is already pronounced and it leads to the uncertainty in experimental absorption maxima as large as 0.1 eV. Concerning the possible deficiencies in the calculations reported here, the discrepancy between the experimental and calculated data could be attributed to the underestimated average polarization of the benzophenone molecule by the solvent. The detailed analysis of the simulations reported in [248] support such attribution. In the approach applied in Ref. [248], the induced dipole moment of the solute was obtained in a self-consistent manner to correspond to the structure of the solvent molecules and the authors stress the importance of proper accounting of this effect in proper modeling of the solvatochromic shifts. It is worth noting that this requirement of self-consistency between the electronic and solvation structures of the solute is met in the present work due to the electrostatic coupling between the solute charges (from quantum mechanical calculations) and the solvent site distributions (from classical statistical-mechanical calculations). The self-consistent solute charges are evaluated taking into account the electrostatic coupling with the solvent. The charges accounting for the electron distribution on the solute are self-consistently recalculated for every new distribution of the solvent sites in the self-consistent field loop. The converged dipole moment of benzophenone solvated in water given by Georg et al. is equal to 5.84 D which is significantly larger than that of the isolated solute by 2.73 D (MP2/6-311++G(d,p) calculations). In the present work, the corresponding values are 5.89 D and 2.16 D, respectively. The solute-solvent interaction induced increase of the dipole moment of benzophenone is, probably, underestimated in the present OFE/RISM calculations. The interpretation that this underestimation as the principal source of the underestimation of the OFE/RISM solvatochromic shifts in the $n \rightarrow \pi^*$ transition, is supported by the fact that the shift obtained in Ref. [248] (0.25 eV) is in excellent agreement with experiment (0.27 eV), whereas the OFE/RISM result is significantly underestimated (0.16 eV). Turning back to the possible underestimation of the solvation induced dipole moment of benzophenone in OFE/RISM calculations, it probably results from the basic approximation made in evaluating averages made in this work, i.e., Eq. 7.7.

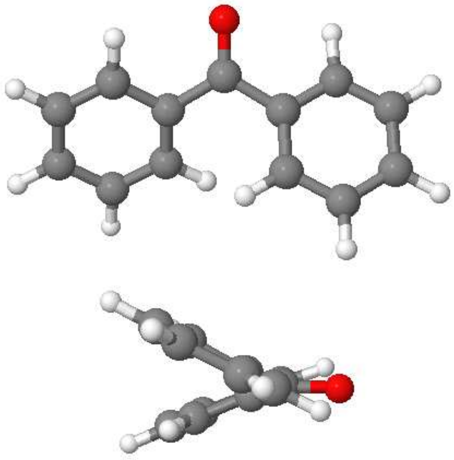


Figure 7.4: Structure of benzophenone taken from Ref. [248]

7.4.5 The role of the non-electrostatic terms in the orbital-free embedding potential

Various embedding schemes are currently in use for simulations of condensed matter. In view of the studies aimed at the electronic structure, it is worthwhile to underline that it is determined by the way the environment is represented at the level of the electronic Hamiltonian rather than in the final expression used for the energy. According to the exact embedding theory [4, 36, 39], the embedding potential comprises non-electrostatic terms. Very commonly, only electrostatic potential is used to represent the environment in the Hamiltonian. Moreover, this potential is frequently evaluated not exactly but it is simplified using the multi-center multipole expansion of

Table 7.5: Solvatochromic shifts (in [eV]) to the gas phase for benzophenone. The shift in the $\pi \rightarrow \pi^*$ transition is calculated as the average $<\varepsilon>=\frac{\sum_i \varepsilon_i \omega_i}{\sum_i \omega_i}$, where ε denotes the transition energy and ω the corresponding intensity. The numbers in parentheses correspond to the case when electrostatic-only embedding is used.

Solvent	Transition	$\Delta\varepsilon_{OFE/RISM}$	$\Delta\varepsilon^a_{exp}$
H_2O	$n \rightarrow \pi^*$	0.16 (0.14)	0.27
	$\pi \rightarrow \pi^*$	-0.10 (-0.12)	-0.20
Methanol	$n \rightarrow \pi^*$	0.16 (0.16)	0.16
	$\pi \rightarrow \pi^*$	-0.08 (-0.09)	-0.10
Diethyl Ether	$n \rightarrow \pi^*$	0.06 (0.06)	0.03
	$\pi \rightarrow \pi^*$	-0.02 (-0.02)	-0.03
Hexane	$n \rightarrow \pi^*$	0.00 (0.00)	-
	$\pi \rightarrow \pi^*$	0.00 (0.00)	-

[a] Ref. [247]

the electrostatic potential generated by the environment, for instance. Such approach is adequate especially if the target of the simulations is the potential energy surface. The non-electrostatic contributions to the total energy are added *a posteriori* as additional terms in the energy. This approach is know in literature to provide reliable results and is commonly used to treat problems in the fields of chemistry and biochemistry [249]. In studies of the electronic structure, electrostatic embedding leads sometimes to artifacts. It may lead, for instance, to artificial charge transfer from the embedded system to the environment caused by strong Coulomb attraction in the vicinity of solvent nuclei [29, 200, 250]. The full embedding potential given in Eq. 7.1 comprises the non-electrostatic components in addition to the electrostatic potential. The role of the $v_{\mathrm{t}}^{nad}[\rho_A, \rho_B](\vec{r})$ component of the embedding potential is to provide counterbalancing repulsive force [200] eliminating such artifacts.

In the present work, the electrostatic potential of environment in Eq. 7.1 is obtained from 3D-RISM-KH method as a statistical-mechanical average of the electrostatic potentials of solvent sites acting on embedded system (Eq. 7.22). This potential was already used to account for solvent effects in connection with Kohn-Sham method [173, 187, 197] or with Hartree-Fock [163, 220] and multiconfigurational methods [251]. To estimate the role of the non-electrostatic components in the embedding potential, each of the numerical value discussed in the previous section is complemented by its counterpart obtained with neglecting non-electrostatic terms in Eq. 7.1. The numerical values are given in parentheses in tables 7.2, 7.4, 7.5. We start analysis from acetone (table 7.4). In the case of this molecule the effect of the non-electrostatic terms is rather noticeable for polar solvents, especially for water. It increases the magnitude of the solvatochromic shift by 0.04 eV, i.e., by 21%. The relative effects decreases with decreasing polarity of the solvent: 16% for methanol and decreases further for diethyl ether and hexane.

In the case of benzophenone (table 7.5), the strongest effect of neglecting non-electrostatic terms occurs for water. The solvatochromic shift is affected by 0.03 eV, i.e., 12% and 27% of the whole shift for the $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ transitions, respectively). For other solvents, the electrostatic-only embedding potential seems to be already accurate as it leads to practically the same shifts as the full $v_{emb}^{KSCED}[\rho_A, \rho_B; \vec{r}]$ potential.

The solvatochromic shifts in aminocoumarin are not noticeably affected by neglecting non-electrostatic terms.

The above results indicate that non-electrostatic terms in the embedding potential affect most significantly the calculated shifts in such cases where the environment constitutes polar molecules with large dipole moment. This result is in line with our previous analyses of the embedding potential of charged systems, for which a significant charge-transfer from embedded molecule to the embedded system can be expected to happen in the absence of two terms in the embedding potential representing the intermolecular Pauli repulsion [102].

7.5 Conclusions

In this work, the ensemble average of the charge density of the solvent is used to represent the solvent surrounding a chromophore. Such treatment can be situated between the explicit atomistic- and the continuum type of models of the solvent. The average charge density accounts for inhomogeneity in the statistically averaged electron density but does not correspond to the electron density of any molecular system at a given geometry. Nevertheless, owing to the unique correspondence between the pair of electron densities (that of the chromophore and of its environment) and the embedding potential [4, 36, 39], the formalism used previously to evaluate spectral shifts in cases where the environment of a chromophore was structurally rigid [127, 52] can be also used for cases where the electron density of the environment lacks such a clear interpretation.

The averaged electron density of the solvent, was obtained by "dressing up" with electron density the site distributions derived from classical theory of liquids. The 3D-RISM theory using the Kovalenko-Hirata [173, 154] closure is used to obtain site distributions.

Using previously reported shifts obtained in explicit simulations for aminocoumarin in water, it is shown that the averaging of the absorption lines corresponding to instantaneous conformations of the solvent can be efficiently replaced by a single evaluation of the excitation energy of a solvent molecules in the average embedding potential. In atomistic simulations, the shifts of the lowest-lying $\pi \rightarrow \pi^*$ excitations depend strongly on the number of instantaneous configurations. They were reported to be equal to -0.28 eV and -0.25 eV for one- and eight 50 ps trajectories, respectively [86]. The shift obtained in the present work, which involves only one evaluation of the excitation energy of the solvated molecule, amounts to -0.25 eV which is in remarkable agreement with the results of the atomistic simulations. A similar agreement between results obtained with the orbital-free embedding potential but with- and without the approximation concerning the averaging procedure introduced in this work was found for solvatochromic shifts of the acetone in water. Abandoning the explicit level of description of the solvent makes it, however, impossible to get the shape of the absorption band. Only the position of the maximum of each band maximum can be evaluated. The accuracy of the calculated spectral shifts depends primarily on the accuracy of the method used to calculate the excitation energy shifts at a given geometry of the environment (based on the combination of the linear-response time-dependent density functional theory LR-TDDFT and orbital-free embedding [51, 52] in the present case). Our benchmarking studies on hydrogen-bonding induced spectral shifts for similar systems [127, 52] indicate that such error is in the range of 0.02-0.05 eV. Taking as the study case, the spectral shifts for the $n \rightarrow \pi^*$ excitations for acetone in solvents of various hydrogen-bonding properties, (water, methanol, diethyl ether, n-hexane), shows that the used method to account for solvent structural flexibility does not lead to deterioration of the quality of the calculated shifts. For these solvents, the discrepancies between the calculated and measured maxima of the absorbtion bands in liquid phase are in the range of 0.04 eV (or smaller). These rather small errors can be thus attributed to the intrinsic accuracy of the applied method to calculate spectral shifts at a rigid geometry. For the ensemble of the studied chromophores and solvents, the calculated solvatochromic shifts agree rather well with experimental measurements. In the worst case, i.e., $n \rightarrow \pi^*$ excitation for benzophenone in water, the calculations reproduce 50% of the experimental shifts (-0.10 eV vs. -0.20 eV). This underestimation is probably due to the approximation made in Eq. 7.7 which neglects the instantaneous coupling between the induced dipole moment of the chromophore and the structure of the solvent, which is averaged out in the applied method. In the absence of such coupling, the contributions to the absorption band due to such configurations of the solvent which generate strong electric field are underrepresented.

The proposed strategy to evaluate the solvatochromic shifts fails, however, to describe the increase of the magnitude of the experimental shift in the $\pi \rightarrow \pi^*$ transition in aminocoumarin C151 from -0.22 eV to -0.31 eV when going from water to methanol. Whereas the calculated spectral shift in water is rather good (-0.25 eV), it is underestimated in methanol (-0.21 eV).

The applied approach offers significant computational saving due to replacing a great number of evaluations of the excitation energy at different embedding potentials (one for each solvent configurations) by one evaluation of the excitation energy but with the average embedding potential. The possible savings are proportional to the number of configurations used in the statistical ensemble. This might depend on the system and on technique to generate the instantaneous geometries for the statistical ensemble but can be expected in the order of 1000 configurations at least. Additional savings are possible due to the fact that the generation of the averaged electron density from "dressed up" 3D-RISM site distributions is faster than explicit evaluation of the frozen electron density of solvent molecules. Note that a number of simplifications can be applied for this purpose in atomistic simulations [53, 86].

Finally, it is worthwhile to underline that the orbital-free embedding potential of Wesolowski and Warshel, similarly to most of the embedding potentials reported in the literature, comprises the electrostatic component. It comprises, however, also the non-electrostatic terms derived in the exact embedding theory [29, 36, 39]. These terms are not given as exact analytic expressions because such expressions are not known. They are evaluated by means of some approximants to the exact density functionals. The numerical results obtained with neglecting non-electrostatic components in the exact embedding potential are included in the present work for comparison purposes. They indicate that non-electrostatic terms lead typically to improved agreement between calculated and measured spectral shifts, although the non-electrostatic components of the embedding potential lead to negligible effect on the excitation energies in some cases. It should be underlined that the non-electrostatic terms in the orbital-free embedding potential (Eq. 7.1) assure that the embedded orbitals (occupied and non-occupied) are obtained from variational calculations. In practice, it manifests itself by the fact that their shapes and energies depend less on the choice of the basis set used to construct the embedded orbitals than in the electrostatic-only case [73, 84, 52, 200].

Part V

Other applications of DFT based methods

Chapter 8

$LiSc(BH_4)_4$: A Novel Salt of Li^+ and Discrete $Sc(BH_4)_4^-$ Complex Anions

$LiSc(BH_4)_4$ has been prepared by ball milling of $LiBH_4$ and $ScCl_3$. Vibrational spectroscopy indicates the presence of discrete $Sc(BH_4)^{4-}$ ions. DFT calculations of this isolated complex ion confirm that it is a stable complex, and the calculated vibrational spectra agree well with the experimental ones. The four BH^{4-} groups are oriented with a tilted plane of three hydrogen atoms directed to the central Sc ion, resulting in a global 8 + 4 coordination. The crystal structure obtained by high resolution synchrotron powder diffraction reveals a tetragonal unit cell with a = 6.076 Å, c = 12.034 Å (space group P-42c). The local structure of the $Sc(BH_4)^{4-}$ complex is refined as a distorted form of the theoretical high symmetry structure. The Li ions are found to be disordered along the z axis.

8.1 Introduction

The development of suitable hydrogen storage materials is a key barrier to the realization of a hydrogen economy. It is currently thought that a viable onboard hydrogen carrier must have very high gravimetric and volumetric hydrogen capacities as well as kinetic and thermodynamic properties that are within rather stringent limits. [252] To date, no directly reversible hydrogen material has been identified that meet all of these criteria. In view of their high theoretical hydrogen storage capacities, metal borohydrides have recently been the subject of intensive investigation. [253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267] The dehydrogenation of lithium and Group II borohydrides is plagued by severe kinetic limitations and irreversibility that precludes the utilization of these compounds under practical conditions [253, 254, 255, 256, 257, 258, 259] even as components of binary hydride mixtures. [261, 262, 263, 264, 265, 266, 267] Many transition metal borohydride complexes also have suitable gravimetric hydrogen densities. [268] However, neutral transition metal borohydride complexes, such as $Zr(BH_4)_4$, can be eliminated a priori from consideration as practical hydrogen carriers due to their high volatility under the conditions required for dehydrogenation. [268] Additionally, the elimination of diborane often competes with dehydrogenation of the neutral complexes. [269] In an attempt to circumvent these problems, we have explored the hydrogen storage properties of Group I salts of anionic transition metal borohydride complexes. Preliminary studies have shown these materials to hold promise as practical hydrogen storage materials. [269] Efforts to develop these materials would be greatly aided by an understanding of their molecular structure. However, the characterization of this class of complexes has, to date, been limited to ^{11}B NMR spectroscopic studies of $LiSc(BH_4)_4$. [270] Thus structural basics, such as the coordination mode of the $[BH_4]$ ligands have not been determined. In the present work, we have used a combination of vibrational spectroscopy, DFT calculations, and X-ray diffraction to elucidate the structure of this new potential hydrogen storage material.

8.2 Experimental details

The compound has been prepared by ball milling of $LiBH_4$ with $ScCl_3$ in 4:1 molar ratio respectively. Raman spectra were obtained using 2 different set-ups: the first one is an Ar ion laser (488nm) with a Kaiser Optical Instruments Holospec Monochromator and a liquid nitrogen cooled CCD camera and finally an Ar-Kr laser (568nm) in conjunction with a Spex 1404 double monochromator and a PM detector. The spectral resolution is 3-4 cm^{-1} for all Raman experiments. IR spectra were obtained using a Perkin Elmer Spectrum One instrument in conjunction with a "Golden Gate" ATR set-up. The spectral resolution is 2 cm^{-1} for the IR experiments. The samples were handled in a glove box and were sealed in glass capillaries for the Raman experiments.

The Kohn-Sham [2] DFT calculations (PBE [25] exchange-correlation functional in TZP basis set) were applied to optimize the geometry of the studied complexes. For such obtained minima, the vibrational frequencies were calculated. All the computations were performed using the ADF [106, 271] program package. Preliminary powder diffraction data (PXD) were measured on the diffractometer STOE IP-PSD equipped with curved image plate detector and using $KCu_{\alpha 1}$ radiation. The sample was filled in 1 mm boron-glass spinning capillary. High resolution synchrotron radiation powder diffraction (SR-PXD) data were collected at the Swiss-Norwegian Beamlines (SNBL, BM01B) at the ESRF, Grenoble, France. A beam with λ=0.5000 Å was provided by a double Si monochromator. The angular range $2\theta = 1.00^o$ to 25.50^o was scanned with a detector bank consisting of 6 scintillation detectors mounted in series with 1.1^o separation. The data were binned to the step size $\Delta 2\theta = 0.002^o$. The sample was contained in a 0.5 mm boron-glass spinning capillary. The observed reflections of $LiSc(BH_4)_4$ were indexed with DICVOL04 [272], the structure was solved with the direct space method program FOX [273] and refined with the Rietveld method by using the program TOPAS [274].

8.3 Results and discussion

IR spectra of the product confirm that $LiBH_4$ has completely reacted. This is also supported by the powder x-ray diffraction which shows only the presence of LiCl and $LiSc(BH_4)_4$ (see below).

As a first structural hypothesis for the new sample (and following the proposal by Nakamori et al [275]), the spectra were compared with those reported previously for $Al(BH_4)_3$ [276, 277], as one expects from crystallochemical considerations some similarities between Al and Sc. In $Al(BH_4)3$, the central Al has a sixfold coordination of hydrides in D3 symmetry.

The literature experimental spectra of $Al(BH_4)3$ present a strong IR band at 1531 cm^{-1} which corresponds to a H-B-H bending mode of E symmetry (labelled υ_{19} in [276]) of the bridging hydrides [276]. This band is not observed in our Sc compound.

A further difference is seen in the Al-B stretching modes, the corresponding strong IR band is observed at 598 cm^{-1}, while the strong Raman Al-B band appears at lower frequencies, namely 495 cm^{-1}. The spectra of the Sc

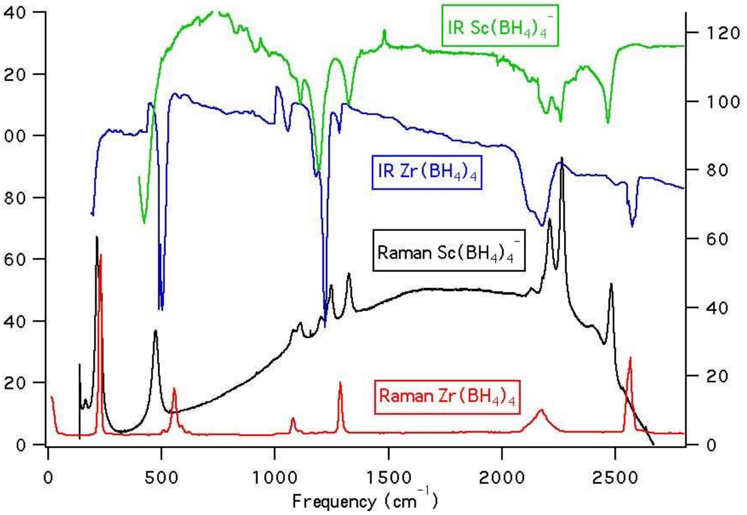


Figure 8.1: Experimental Raman and IR spectra of $LiSc(BH_4)_4$ (this work) and $Zr(BH_4)_4$

complex, however (see Figure 8.1), show in this spectral range a strong IR band at lower frequency than the Raman band (423 vs 474 cm^{-1}).

The observation of the highest B-H stretching mode around 2468 (IR) and 2483 (Raman) as well as a doublet around 2200 -2270 cm^{-1} suggests rather the presence of a tridentate B-H binding to the central metal ion [277] .The complex $Zr(BH_4)_4$ had been shown to present this type of B-H binding, and its vibrational spectra have been published previously [278]. These old data are compared to our new experiments of the Scandium complex in 8.1 and Table 8.1. Analysis of the results reveals indeed many similarities between both IR and Raman spectra of the two compounds.

We then performed a DFT calculation on the isolated $[Sc(BH_4)_4]^-$ complex. The optimization of the geometry without symmetry constraints and the calculation of the vibrational spectra confirm that this ion has a stable structure, similar to the one found for $Zr(BH_4)_4$.

The calculated structure (see Figure 8.2 and Table 8.2) shows that each BH^{4-} ion has a plane of three hydrides oriented partly towards the central Sc atom, resulting in a global 8 + 4 coordination with 8 Sc-H distances of 2.11-2.15 and 4 Sc-H distances of 2.31 Å. The last B-H bond points to the outside, yielding the high calculated B-H stretching frequencies between

Table 8.1: Comparison between Experimental IR and Raman Spectra of $\text{LiSc(BH}_4\text{)}_4$. All values in cm^-1.

Raman Zr [278]	Raman Sc	Δ	IR Zr [278]	Δ	Δ
233	216	17	198		
237	231	6	489	423	66
507			504		
556	474	82	1057	1113	-56
589			1180	1194	-14
	1079		1221		
1080	1112	-32	1284	1325	-41
	1247		2123	2199	-76
1289	1325	-36	2180	2259	-79
2176	2211	-35	2576	2468	108
	2264				
2567	2483	84			

2485 and 2498 cm^{-1}. Figure 8.3 compares the calculated and observed IR spectrum.

The calculated frequencies are typically within less than 30 -50 cm^{-1} of the experimental ones. The agreement between those results is surprisingly good taking into account the three factors influencing the DFT calculations: use of an approximated exchange-correlation functional for the exchange-correlation energy, harmonic approximation, and the complete neglect of the effects due to the crystal packing forces.

In view of the previous DFT studies on the alkali borohydrides [279], which showed that the effect of environment on vibrations of the borohydride ion reaches up to 100 cm^{-1}, this overall good agreement is probably the effect of compensation of these three types of errors.

The PXD data showed as a major phase in the sample LiCl (60 wt%), the second phase corresponded to the title compound. A primitive cubic unit cell (a = 6.064 Å) without systematic extinctions explained the positions of the observed reflections of $\text{LiSc(BH}_4\text{)}_4$ in PXD data. The structure was solved in the space group P23 providing one lithium atom (position 1a), one scandium atom (position 1b) and four BH_4 groups (position of the boron atom 4e) within a unit cell. The structure was solved with the BH_4 groups

Figure 8.2: Structure of the $Sc(BH_4)_4^-$ complex from the DFT calculations

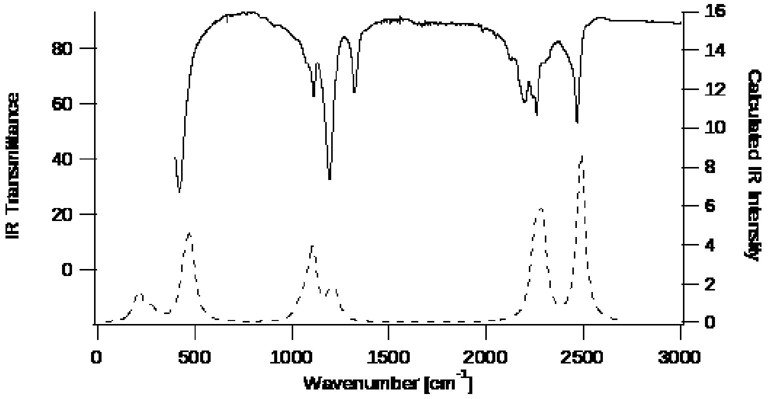


Figure 8.3: Comparison between observed IR spectrum of $LiSc(BH_4)_4$ and (full line) and calculated theoretical DFT spectrum (dashed line) for individual $Sc(BH_4)_4^-$ ions.

Table 8.2: Selected bond distances and angles from the DFT calculations on the Isolated $Sc(BH_4)_4^-$ ion and as refined from synchrotron PXD. The angles (Sc-H-B) are calculated only on the three hydrogen atoms pointing towards scandium.

	DFT	X-ray
d(B-H_1) [Å]	1.21	1.08(2)
d(B-H_2) [Å]	1.24	1.08(2)
d(B-H_3) [Å]	1.24	1.08(2)
d(B-H_4) [Å]	1.24	1.08(2)
$(Sc\text{-}H\text{-}B)_{min}$ [deg]	75	63
$(Sc\text{-}H\text{-}B)_{max}$ [deg]	83	95
$d(Sc\text{-}B)_{min}$ [Å]	2.33	2.28(1)
$d(Sc\text{-}B)_{max}$ [Å]	2.33	2.28(1)
$(B\text{-}Sc\text{-}B)_{min}$ [deg]	105	96.5(5)
$(B\text{-}Sc\text{-}B)_{max}$ [deg]	113	125.4(5)

as rigid bodies with fixed B-H distances of 1.2 Å.

The high resolution SR-PXD data showed that the symmetry is lower than cubic as some peaks were split. Indexing with the first 20 observed peaks from SR-PXD data showed excellent agreement with a tetragonal unit cell (a = 6.076 Å, c = 12.034 Å). The observed extinction rules were in agreement with the space groups P42mc, P42mmc and P-42c. Due to the good agreement between the PXD data and the cubic structure model, it was assumed that CsCl-type ordering of the Li-Sc sublattice was basically correct. It can be described in any of the three candidate space groups. However, only P-42c has site symmetry for Sc that allows for tetrahedral coordination by four BH_4 tetrahedra. Thus, this space group was assumed as correct one.

An approximate structure model was obtained by fixing Li and Sc in the positions corresponding to the cubic model (Li on 2e and Sc on 2c), optimizing both the position and orientation of rigid BH_4 tetrahedra. The final Rietveld refinement ($\chi^2 = 1.1$, Rietveld plot in the supplementary material) of the model showed the disordering of Li around the 2e position to the 4k position with z 0.1 (Figure 8.4 top). The disordering can be understood from the analysis of interatomic distances. The distance between the Li atom and four B atoms from the coordinating BH_4 groups would be 3.00 Å if Li were on the position 2e. The disordering of Li to the 4k position decrease the number of coordinating BH_4 groups from four to two, but with Li-B distances of 2.54(1) Å (Figure 8.4 bottom), closer to this distance observed in $LiBH_4$ (2.48-2.54 Å) [280, 281]. Whether the disordering of Li along the c-axis of the tetragonal structure is of dynamic or static nature stays to be investigated. However, it indicates a high mobility of the Li atoms in this compound which is possibly as good ionic conductor as it was recently recognized for the HT-phase of $LiBH_4$ [282]

As shown in Figure 8.4, the Sc^{3+} is surrounded by four $[BH_4]^-$ tetrahedra in deformed tetrahedral coordinations (see table 8.2 for selected bond distances and angles). A similar tetrahedral coordination of a cation by four $[BH_4]^-$ anions is known for Li^+ in ambient pressure phases of $LiBH_4$ [280, 281], Zr^{4+} in zirconium borohydride at 113 K [283], and it was recently observed also for Mg^{2+} in magnesium borohydride [284]. As for the Li^+ coordination (discussed above), it looks unusual (Figure 8.4 bottom) compared to $LiBH_4$, and reflects the high mobility of the Li^+ cations in the structure. On the other side each $[BH_4]^-$ is approximately linearly coordinated by one Sc^{3+} and one Li^+ with half occupancy (Sc-B-Li = $148.8(5)^o$). The coordination of cations by hydrogen atoms cannot be determined unambiguously due to not enough accurate determination of the $[BH_4]^-$ tetrahedra orientation and lithium atom localization. However, from the analysis of the cation coordination in similar borohydrides like $LiBH_4$ [3,34] and $Mg(BH_4)_2$ [285] and from the space filling consideration we conclude that each Sc^{3+} is coordinated approximately by triangular faces of four BH_4 tetrahedra (Sc-H distances 1.92(5), 2.08(3) and 2.55(8) Å), thus resulting in an 8 + 4, relatively irregular hydrogen coordination. The comparison of the DFT structure with the refined crystal structure (see table 8.2) shows that the distortion from a high symmetry D3 complex seen by the DFT calculation is pushed further in the crystal structure.

As for Li^+ coordination by hydrogen it can be understood either like by vertices or by edges (Li-H distances 1.98(8) and 2.5(1) Å) of two BH_4 tetrahedra. A single B-H distance of 1.08(2) Å was refined for rigid BH_4 tetrahedron. It compares well with those in related borohydrides such as $LiBH_4$ (1.04-1.28 Å). [280, 281]

The Raman spectra suggest that the borohydride groups are not subject to orientation disorder, as the band width remains rather sharp (FWHM ca 16 cm^{-1} for the B-H bending modes). Raman spectra of the alkali borohydrides have shown significant broadening of the B-H bending modes in the

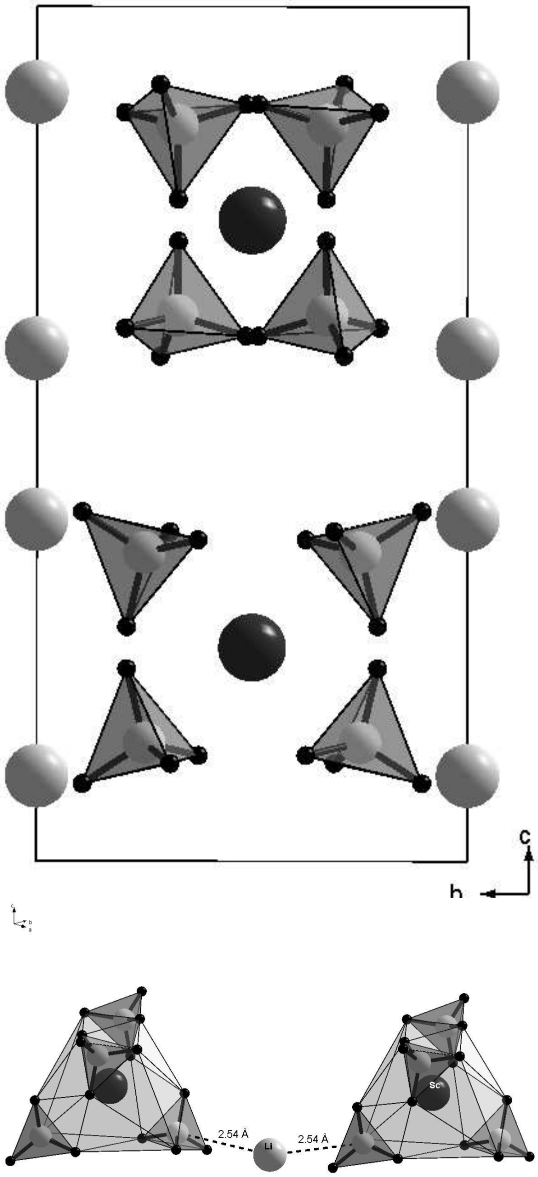


Figure 8.4: Crystal structure of $LiSc(BH_4)_4$ viewed along the a-axis (top) and approximately along the [1-10] direction showing the coordination of Li and Sc by BH_4 tetrahedra (bottom)

disordered phase. [285, 286]

Finally, we have performed DFT calculations on a model complex comprising two $Sc(BH_4)_4$ units separated by 4 Li atoms to simulate a truncated tetragonal unit cell. Initially, the Li atoms were placed on the (0 0 0.6) between the two Sc complexes. During the geometry optimization, two diagonally opposed Li ions moved towards the upper Sc complex, while the two others moved to the lower complex in agreement with the preferred Li-B distances of 2.54 Å (Figure 8.4 bottom).

Although this model calculation is biased by the symmetry constraints and the absence of charge compensation outside of the complex (resulting in several imaginary vibrational frequencies), it confirms the trend of the Li ions to form bonds with the BH_4 units rather than occupy an average central position. The resulting geometry and IR spectra are given in the supplementary informations. Periodic DFT calculations on this compound are planned, to account for the effect of the interactions of the studied system with its neighbors present in the solid but absent in the considered theoretical model.

8.4 Conclusions

The structure of $LiSc(BH_4)_4$ has been studied using a combination of vibrational spectra, DFT calculations and high resolution synchrotron powder diffraction. The local symmetry of the isolated $Sc(BH_4)^{4-}$ complex, as obtained from the DFT calculations, was confirmed by the crystal structure analysis. The combined experimental and theoretical results show consistently that each Sc^{3+} is tetrahedrally surrounded by triangular faces of four BH^{4-} tetrahedra. The crystal structure data show in addition disorder for the Li ions, which in turn leads to reasonable Li - B distances. Preliminary DFT calculations on a model complex comprising two $Sc(BH_4)_4$ units separated by 4 Li atoms confirm the trend to reduce individual Li - B distances in contrast to the average high symmetry central position.

The title compound is apparently the same as the material obtained by Nakamori et al through the balling milling of $ScCl_3$ and $LiBH_4$ in 1:3 molar ratio which they misformulated as $Sc(BH_4)_3$. [287] Their observation of the onset of hydrogen desorption for the compound around 450 K and complete release of hydrogen around 529 K [287] is noteworthy in view of its very high, 14.4, hydrogen weight percentage. Another potential practical application of this material is suggested by our observation of the disordering of Li^+ in the crystal structure. This finding points to the possibility that this material may be a good conductor of lithium ions.

The ionic radius of Al^{3+} is too small to present a similar BH^{4-} coordination, as this would lead to much closer hydride contacts which destabilize the potential $[Al(BH_4)_4]^-$ ion. However, the $[Al(BH_4)_4]^-$ anion is reported to exist [288]. Thus either the anion is misformulated or the coordinative interactions of the complex ion is very different those we have elucidated for $LiSc(BH_4)_4$.

ACKNOWLEDGMENT Work supported by the Swiss National Science Foundation and the Office of Hydrogen Fuel Cells and Infrastructure Technology of the U.S. Department of Energy. We thank D. Lovy for his help in digitizing the literature spectra from the published figures using his program Windig. The authors T.A.W. and J.W.K. acknowledge the Swiss National Supercomputing Centre for the CPU time.

Appendix: cartesian coordinates of model $Li_4(Sc(BH_4)_4)^{2+}$ complex, calculated ADF vibrational frequencies and IR intensities for the isolated complex, calculated IR spectrum of model $Li_4(Sc(BH_4)_4)^{2+}$ complex and Rietveld plot for synchrotron powder diffraction data.

Calculated coordinates of the $Sc(BH_4)_4^-$ complex (in Å)

Atom	X	Y	Z
Sc	3.038010	3.038008	9.025393
B	4.457403	4.317227	7.687211
H	4.793417	4.266899	8.884974
H	5.311018	4.851632	7.016605
H	3.344350	4.867042	7.644379
H	1.791596	2.934188	7.313645
B	1.618608	1.758628	7.687287
H	1.282586	1.809555	8.885009
H	0.764610	1.224620	7.016862
H	2.731418	1.208377	7.644321
H	4.284781	3.141493	7.314003
B	1.618499	4.317310	10.363485
H	1.282508	4.266436	9.165765
H	0.764568	4.851402	11.033931
H	2.731542	4.867558	10.406402
H	4.284719	2.934348	10.736865
B	4.457454	1.758683	10.363527
H	4.793474	1.809196	9.165829
H	5.311116	1.224292	11.034080

H	3.344204	1.208761	10.406357
H	1.791517	3.141745	10.737071

Cartesian coordinates of theoretical model $Li_4(Sc(BH_4)_4)^{2+}$ complex.

B	1.529705	1.332592	7.689584
H	1.754045	1.306577	8.901148
H	2.593091	1.298742	7.078646
H	0.762886	2.234626	7.381507
Sc	0.000000	0.000000	9.219440
H	-1.035813	-0.269646	7.317769
B	-1.529705	-1.332592	7.689584
H	-1.754045	-1.306577	8.901148
H	-2.593091	-1.298742	7.078646
H	-0.762886	-2.234626	7.381507
H	1.035813	0.269646	7.317769
B	-1.536789	1.194208	10.354034
H	-1.786983	1.081499	9.125380
H	-2.426974	1.690430	10.979572
H	-0.441608	1.807550	10.382510
H	1.279901	-0.038083	10.756853
B	1.536789	-1.194208	10.354034
H	1.786983	-1.081499	9.125380
H	2.426974	-1.690430	10.979572
H	0.441608	-1.807550	10.382510
H	-1.279901	0.038083	10.756853
B	1.522038	1.142729	1.698457
H	1.799189	1.037536	2.913386
H	2.401089	1.630277	1.048236
H	0.431444	1.769684	1.671620
Sc	0.000000	0.000000	2.895030
H	-1.230952	0.008983	1.302331
B	-1.522038	-1.142729	1.698457
H	-1.799189	-1.037536	2.913386
H	-2.401089	-1.630277	1.048236
H	-0.431444	-1.769684	1.671620
H	1.230952	-0.008983	1.302331
B	-1.493545	1.339180	4.518702
H	-1.777146	1.362490	3.318372
H	-2.585458	1.466366	5.076122
H	-0.636673	2.171302	4.794559
H	1.120380	-0.208513	4.837110
B	1.493545	-1.339180	4.518702
H	1.777146	-1.362490	3.318372
H	2.585458	-1.466366	5.076122
H	0.636673	-2.171302	4.794559
H	-1.120380	0.208513	4.837110
Li	-2.836992	-3.078395	7.218469
Li	-2.656521	2.568991	6.373243
Li	2.656521	-2.568991	6.373243
Li	2.836992	3.078395	7.218469

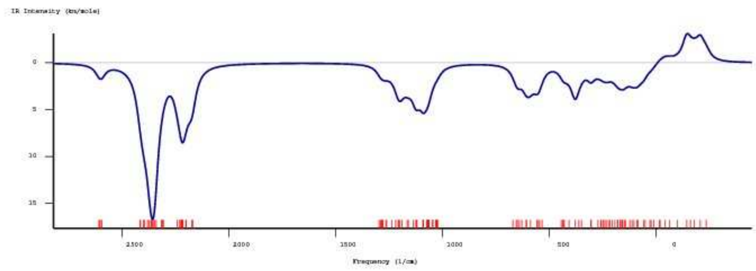


Figure 8.5: Calculated IR spectrum of model Li4(Sc(BH4)4)2+ complex. The imaginary frequencies are plotted as negative frequencies.

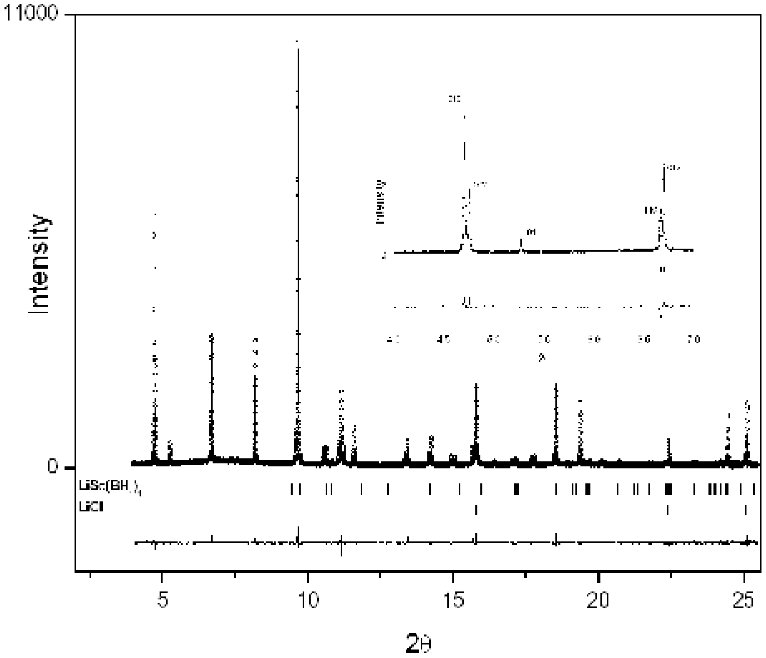


Figure 8.6: Rietveld plot of high resolution synchrotron data of LiSc(BH$_4$)$_4$, = 0.5000 Å , 2 = 1.1, R$_{wp}$= 0.10, R$_{Bragg}$ = 0.056 for LiSc(BH$_4$)$_4$. The position of reflections of LiSc(BH4)4 and of LiCl are shown by vertical bars. The splitting of cubic reflections and the superstructure reflection (011) of LiSc(BH$_4$)$_4$ are shown in the inset.

Table 8.3: Calculated ADF vibrational frequencies and IR intensities for the isolated complex

Frequency (cm^{-1})	Dipole Strength ($10^{-}40$ esu^2)	Absorption Intensity (km/mole)
158.686915	0.505051	0.020089
179.058886	362.814150	16.283890
191.731635	1.782952	0.085686
201.985284	67.406649	3.412719
205.772673	338.598971	17.464309
209.713981	780.822615	41.044771
220.592754	1061.708426	58.704930
264.433359	62.734468	4.158154
272.258028	371.402531	25.345661
277.234549	7.045440	0.489591
286.054231	250.082215	17.931196
327.234373	125.855903	10.323111
338.305688	0.698887	0.059264
438.926559	1099.744809	120.993448
459.309669	605.337970	69.691783
466.926066	0.249581	0.029210
469.984082	927.486956	109.261971
481.020760	1260.045493	151.924605
500.493830	276.789293	34.723707
506.319156	14.760058	1.873227
529.961616	0.792026	0.105211
1033.300282	118.627934	30.724978
1034.298588	14.721578	3.816615
1037.451086	1.910265	0.496752
1043.836195	1.657207	0.433598
1059.369465	51.393886	13.646993
1061.381041	0.105682	0.028116
1069.560730	155.266292	41.625619
1074.860886	57.410768	15.467627
1099.102523	272.187462	74.986726
1103.314170	195.417982	54.043291
1111.677402	565.974559	157.708012
1131.956995	0.072441	0.020554
1189.675481	173.311506	51.681365
1202.836232	39.889600	12.026638
1212.978220	178.976698	54.416119
1226.330746	118.610496	36.459341
1231.818666	24.919104	7.694090
1234.295336	0.555864	0.171975
1241.636157	30.974568	9.640014
1248.287624	1.097482	0.343392
2243.558029	101.016042	56.807506
2246.835189	86.738589	48.849671
2249.676797	199.708234	112.614463
2252.647803	70.776095	39.962989
2259.068633	0.208244	0.117918
2266.088555	39.320044	22.334125
2269.021234	107.683870	61.244527
2273.049404	0.864114	0.492332
2286.349697	148.802380	85.276743
2287.597084	192.138282	110.172069
2288.874164	4.684833	2.687782
2294.774759	235.324859	135.358627
2484.974464	362.294651	225.663813
2488.026081	495.939959	309.287309
2491.865296	228.297542	142.594859
2497.888040	5.409636	3.387031

Part VI

Conclusions

This thesis concerns new developments within the orbital-free embedding theory, improvement of the existing functionals and design of the new algorithms facilitating multi-level simulations in condensed matter.

The important component of the orbital-free embedding formalism is the non-additive kinetic energy functional, $\mathrm{T}_s^{nad}[\rho_A, \rho_B]$. In common practice it is calculated from known approximants to the non-interacting kinetic energy $\mathrm{T}_s[\rho]$. Dedicated studies [13] revealed, however, that there is no correlation between the accuracy of $\mathrm{T}_s^{nad}[\rho_A, \rho_B]$ and related potential with the accuracy of the parent approximant to $\mathrm{T}_s[\rho]$. Those conclusions stimulated a completely new, bottom-up strategy to develop approximants, where the direct quantity of interest is term $v_t[\rho_A, \rho_B]$ - the non-additive kinetic energy potential. This route was succesfully applied in Chapter 4. Moreover, by studying the asymptotic behaviour of $v_t[\rho_A, \rho_B]$ on large intermolecular distances, the new exact relation for this potential was derived. The violation of this property by conventional approximations to $v_t[\rho_A, \rho_B]$ is the reason behind the artificial transfer of the density from the embedded system to its environment on the intermediate intermolecular distances. In Chapter 4, the asymptotic relation was build-in the construction of the new approximant to the embedding potential, which leaded to the completely new class of *non-decomposable* approximants to $v_t[\rho_A, \rho_B]$, i.e. where the parent approximation to $\mathrm{T}_s[\rho]$ is impossible to recover. The numerical results clearly indicate that the proposed functional, dubbed as NDSD (Non-Decomposable approximant using first- and Second Derivatives of ρ), improve the studied observables, in some cases even by more than 50% when compared to reference data. Moreover, for the molecular systems where the artificial transfer of the density occurs, the new approximant partially cures this problem by improving the description of the unoccupied orbitals.

The author believes that the strategy to construct relevant density functionals by enforcing known exact properties is very promising. The numerical results in Chapter 4 confirm this statement. The NDSD functional can be treated as a successor of the widely used GGA97 approximation to $v_t[\rho_A, \rho_B]$, however, it still does not bring a major improvements in the practical applications. By including the exact asymptotic condition in construction of this functional, one important reason of inaccuracies was eliminated, nevertheless others still remain to be discovered. Although NDSD was developed as an non-empirical approximation, the practical implementation relies on arbitrary chosen parameters in the switching function *f*. The development of the non-empirical form of function f might lead, therefore, to further improvement of results.

The assessment of the accuracy conventional decomposable approximants to $T_s^{nad}[\rho_A, \rho_B]$ and $v_t[\rho_A, \rho_B]$ was one of the concerns in Chapter 6, where the first application of subsystem formulation of density functional theory to find the equilibrium geometries of noncovalently bound intermolecular complexes was presented. It relied on the efficient algorithm to perform geometry minimization, where the Born-Oppenheimer energy minimum is reached only at the end of the optimization procedure. Such approach brought a considerable savings in the time of calculations. Moreover, modular implementation of the method in computer code allowed for the optimization with various degrees of freedom being kept fixed.

Turning back to the approximants to $T_s^{nad}[\rho_A, \rho_B]$ and $v_t[\rho_A, \rho_B]$, the optimization procedure was tested on LDA and GGA97 approximants. In all of the studied cases, except $\pi - \pi$ stacked complexes, the former approximant leaded to superior results over the latter. The reason for this might be the fortunate error cancellation between local density approximation for kinetic- and exchange-correlation energy. The equilibrium geometries obtained with subsystem formulation of DFT are also close to the reference data taken from high-accuracy *ab-initio* calculations. Typically the deviations in the intermolecular distances are smaller than 0.1 Å in the case of LDA. Additionally, it has been observed that in some cases subsystem DFT leads to better equilibrium structures than corresponding Kohn-Sham calculations. Again, the aforementioned error cancellation between functionals might be the reason for such outcome.

The numerical results presented in Chapter 6 show that subsystem formulation of density functional theory is not only a tool that can be used to predict the properties of the embedded molecules. The calculated interaction energies between two subsystems and their equilibrium geometries match closely the reference data, providing, therefore, a complete description of the molecular interface. This leads to the conclusion, that taking into account the flexibility in the choice of degrees of freedom to be optimized within subsystem DFT, application of studied in this thesis theory might be especially beneficial in the case of the problems where, for instance, energetics, geometry and electronic structure of the chromphore embedded in large, rigid environment is subject of interest.

The orbital-free embedding potential depends on the density of the environment, ρ_B. In principle, the choice of this density is arbitrary - for example it can be taken from any quantum-chemical method working at any level of description. For this reasons, the embedding potential if perfectly suited for multi-level simulations combining various methods of treatment of embedded molecule and its environment. The development of algorithm facilitating applications of embedding potential for this purpose was subject of Chapter 5, whereas the example of using environment density taken from statistical mechanical theory of the liquids to calculate the statistically averaged embedding potential was presented in Chapter 7.

Linearizing orbital-free embedding potential (Chapter 5) by expanding it in Taylor series around some arbitrary chosen ρ_A^0 and truncating the series on first non-disappearing term, provides an approximation which can be very useful in practical applications. Avoiding the reevaluation of the embedding potential during every self-consistent field iteration reduces considerably the time of calculations. Moreover, the linearized embedding potential can be easily interfaced between different computer codes. This facilitates the implementation of possible multi-level modeling schemes, for example embedding of mutlideterminantal wave function in orbital-free environment. Such approaches were already successfully tested and used in practice [37, 38].

As in the case of every approximation, linearized form of the embedding potential introduces an error in the calculations. This error will depend on the choice of the reference density ρ_A^0 - the closer it will be to the ground-state density of the embedded system, ρ_A, the smaller discrepancy is to be expected. Dedicated analysis showed that optimal choice for ρ_A^0 is the density of isolated subsystem A. Among the studied weakly bound complexes, such choice of the linearized potential is not adequate only for the molecules where large charge-transfer occurs. To control the linearization errors, a special procedure, dubbed as *splitSCF*, was proposed where the self-consistent calculations are divided in the two loops - inner where the linearized embedding potential is used, and outer, where this potential is updated. It was shown that after three iterations of the external loop, the linearization error disappears and the results are equivalent to conventional calculations utilizing embedding potential. In the case of intermolecular complexes characterized by charge transfer, four iterations of the outer loop are required. The *splitSCF* procedure is not recommend to be applied in the large-scale multi-level simulations, as the linearization approximation with adequately chosen ρ_A^0 is very good and efficient approach. It is, however, very useful to verify the adequacy of the linearization approximation, especially in the case where large interaction induced deformation of the density occurs.

The benefits from using linearization approximation were also presented in Chapter 6. It was applied to obtain the equilibrium geometries for the chosen compounds, leading to the results which are in agreement with the benchmark data, but computed at reduced cost.

In all of the previous applications employing orbital-free embedding potential all the molecules constituting environment were included explicitly. Replacing the explicit description by "dressed up" in electron density site distributions obtained from statistical mechanical theory of the liquids reduces the complexity of computations and allows one to consider large environments of flexible structure. Additionally, opposed to other approaches describing solvent in an implicit way, applied in Chapter 7 3D-RISM method retains chemical specificity of molecular interactions (i.e. hydrogen bonds) which has a crucial importance in the applications where the interaction induced shifts of absorption bands are direct subject of interest.

The "dressed up" site distributions provide statistically averaged solvent density, which is used to obtain the statistically averaged orbital-free embedding potential. Such potential, as demonstrated in Chapter 7, allows one to replace costly calculations of the excitation energies on ensemble of solvent configurations by only one single evaluation. This is the advantage of multiscale models. To validate the adequacy of OFE/3D-RISM-KH combination, the results where compared with the previous atomistic simulations and experimental studies on acetone and aminocoumarin C151. The outcome was encouraging, for both of the considered chromophores the numerical results were in remarkable agreement with reference data. This allows one to draw the general conclusion, that despite of the fact that application of average potential gives only the positions of the maxima of the spectrum and not its shape, it recovers its most characteristic features. The reasons for the additional discrepancies might be buried either in 3D-RISM part (inadequacy of chosen force-field parameters or closure) or in OFE part (for example NDRE approximation, inaccurate functionals). Despite of this, the combination of orbital-free embedding with 3D-RISM method benefits from the properties of the both partners. The 3D-RISM-KH method provides statistical description of the solvent that can be situated between explicit and implicit models and can be evaluated without considerable computational effort. The orbital-free embedding formalism on the other hand, relies on the *exact* embedding potential (in the limit of exact functionals), that is an universal quantity defined for any density of the environment. Its analytical form constitutes exact electrostatic and non-electrostatic components, where the latter provide Pauli repulsion force necessary to avoid artificial transfer of the density from the embedded system to its environment. Dedicated analysis in Chapter 7 showed that those terms typically lead to better agreement of calculated spectral shifts with the experimental data.

Considering application of quantum-chemical tools to aid the experimental studies, in Chapter 8 the Kohn-Sham density functional theory was used to reinforce the interpretation of powder diffraction and vibrational spectroscopy findings on the structure of $\mathrm{LiSc(BH_4)_4}$. The equilibrium geometry of $[\mathrm{Sc(BH_4)_4}]^-$ complex and the corresponding IR spectrum were estimated. Theoretical results were in very good agreement with the experimental reports, and confirmed the structure of the complex anticipated from powder diffraction. The preliminary computational studies on $\mathrm{Li_4(Sc(BH_4)_4)^{2+}}$ where also reported and they provide excellent starting point for furtherer investigations involving orbital-free embedding potential.

As a final remark, this thesis concerns applications and new developments in orbital-free embedding formalism. It was presented that the intrinsic properties of the orbital-free embedding potential makes it ideally suited for large-scale multi-level simulations of chemical interactions. The discrepancies in the results are to be attributed to the deficiencies of the existing density functionals - both exchange correlation energy and non-additive kinetic energy. The latter is a crucial component of studied formalism, and for this reasons this work addresses its exact properties and introduces a new class of approximants that inherit all the good features of its ancestors and comprises the new exact asymptotic relation. Moreover, the results discussed in this thesis open many possibilities for continuation of started studies. The new and previously unexplored approach to construct approximations to non-additive kinetic energy potential was proposed. In view of previous studies showing the lack of correlation between $\mathrm{T}_s[\rho]$, $\mathrm{T}_s^{nad}[\rho_A, \rho_B]$ and $v_t[\rho_A, \rho_B]$ this route seems to be very promising. The efficient geometry optimization algorithm allows one to tackle problems where dependence of electronic structure and properties of the embedded system on the mutual orientation of both subsystems is subject of interest. The *splitSCF* procedure together with linearization approximation are ready to be used to interface the orbital-free embedding potential with any quantum-chemical code. On the other end, the multi-scale model combining orbital-free embedding formalism with statistical mechanical theory of the liquids proposed and validated in Chapter 7, already allows to study the properties of the large chromophores embedded in various solvents and offers many possible applications to come.

Bibliography

[1] Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964,** *136,* B864-B871.

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

[3] Levy, M. *Proc. Natl. Acad. Sci. U. S. A.* **1979,** *76,* 6062-6065.

[4] Wesolowski, T. A.; Warshel, A. *J. Phys. Chem.* **1993,** *97,* 8050-8053.

[5] Wang, X.; Andrews, L.; Infante, I.; Gagliardi, L. *J. Am. Chem. Soc.* **2008,** *130,* 1972.

[6] Kato, T. *Commun. Pure Appl. Math.* **1957,** *10,* 151-177.

[7] Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules;* Oxford, New York: , 1989.

[8] Gilbert, T. L. *Phys. Rev. B* **1975,** *12,* 2111.

[9] Thomas, L. H. *Proc. Cambridge Philos. Soc.* **1927,** *23,* 542.

[10] Fermi, E. *Z. Phys.* **1928,** *48,* 73.

[11] von Weizsäcker, C. F. *Z. Phys.* **1935,** *96,* 431.

[12] Ludena, E. V.; Karasiev, V. *Kinetic energy functionals: history, challeneges and prospects in "Reviews of Modern Qunatum Chemistry";* volume 1 of *Reviews of Modern Quantum Chemistry* World Scientific, Singapore: , 2002.

[13] Bernard, Y. A.; Dulak, M.; Kaminski, J. W.; Wesolowski, T. A. *J. Phys. A* **2008,** *41,* 0553902.

[14] Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. *Phys. Rev. B* **2004,** *69,* 075102.

[15] Perdew, J. P.; Schmidt, K. *Density Functional Theory and Its Application to Materials;* AIP, Melville, NY: , 2001.

[16] Dirac, P. A. M. *Proc. Cambridge Philos. Soc.* **1930,** *26,* 376-385.

[17] Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980,** *58,* 1200-1211.

[18] Ceperley, D. M.; Alder, B. J. *Phys. Rev. Lett.* **1980,** *45,* 566-569.

[19] Becke, A. D. *Phys. Rev. A* **1988,** *38,* 3098-3100.

[20] Perdew, J. P.; Yue, W. *Phys. Rev. B* **1986,** *33,* 8800-8802.

[21] Perdew, J. P. *Phys. Rev. B* **1986,** *33,* 8822-8824.

[22] Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988,** *37,* 785-789.

[23] Perdew, J. P. *Electronic Structure of Solids;* Akademie Verlag, Berlin: , 1991.

[24] Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1992,** *45,* 13244-13249.

[25] Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996,** *77,* 3865-3868.

[26] Tao, J. M.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003,** *91,* 146401.

[27] Wesolowski, T. A. *Hohenberg-Kohn-Sham Density Functional Theory: The formal basis for a family of succesful and still evolving computational methods for modelling interactions in complex chemical systems.;* Molecular Materials with Specific Interactions: Modeling and Design Springer Verlag: , 2007.

[28] Cortona, P. *Phys. Rev. B* **1991,** *44,* 8454-8458.

[29] Wesolowski, T. A. *One-electron equations for embedded electron density: challenge for theory and practical payoffs in multi-level modelling of soft condensed matter;* volume X of *Computational Chemistry: Reviews of Current Trends* World Scientific: , 2006.

[30] van Leeuwen, R. *Adv. Quantum Chem.* **2003,** *43,* 25-94.

[31] Wesolowski, T. A.; Weber, J. *Chem. Phys. Lett.* **1996,** *248,* 71-76.

[32] Cohen, A. J.; Mori-Sanchez, P.; Yang, W. T. *Science* **2008,** *321,* 792.

[33] Grimme, S. *Angew. Chem. Int. Ed.* **2006,** *45,* 4460.

[34] Cha, J.; Lim, S.; Choi, C. H.; Cha, N.; Park, N. *Phys. Rev. Lett.* **2009,** *103,* 216102.

[35] Gusarov, S.; Malmqvist, P.-A.; Lindh, R.; Ross, B. O. *Theor. Chem. Acc.* **2004,** *112,* 84.

[36] Wesolowski, T. A. *Phys. Rev. A* **2008,** *77,* 012504.

[37] Gomes, A. S. P.; Jacob, C. R.; Visscher, L. *Phys. Chem. Chem. Phys.* **2008,** *10,* 5353.

[38] Kluner, T.; Govind, N.; Wang, Y. A.; Carter, E. A. *J. Chem. Phys.* **2002,** *116,* 42-54.

[39] Pernal, K.; Wesolowski, T. A. *Int. J. Quantum Chem.* **2009,** *109,* 2520-2525.

[40] Buijse, M. A.; Baerends, E. J. *Mol. Phys.* **2002,** *100,* 401.

[41] Gritsenko, O.; Pernal, K.; Baerends, E. J. *J. Chem. Phys.* **2005,** *122,* 204102.

[42] Cances, E.; Pernal, K. *J. Chem. Phys.* **2008,** *128,* 134108.

[43] Lathiotakis, N. N.; Marques, A. L. *J. Chem. Phys.* **2008,** *128,* 184103.

[44] Marques, A. L.; Lathiotakis, N. N. *Phys. Rev. A* **2008,** *77,* 032509.

[45] Runge, E.; Gross, E. K. U. *Phys. Rev. Lett.* **1984,** *52,* 997.

[46] Ghosh, S. K.; Dhara, A. K. *Phys. Rev. A* **1988,** *38,* 1149.

[47] Willetts, A.; Rice, J. E.; Burland, D. M. *J. Chem. Phys.* **1992,** *97,* 7590-7599.

[48] Casida, M. E. *Time-dependent density-functional response theory for molecules;* Recent Advances in Density-Functional Methods. D. P. Chong, eds. World Scientific, Singapore, p. 155: , 1995.

[49] van Gisbergen, S. J. A.; Snijders, J. G.; Baerends, E. J. *Comp. Phys. Comm.* **1999,** *118,* 119-138.

[50] Bauernschmitt, R.; Ahlrichs, R. *Chem. Phys. Lett.* **1996,** *256,* 454.

[51] Casida, M. E.; Wesolowski, T. A. *Int. J. Quantum Chem.* **2004,** *96,* 577-588.

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

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

[54] Magyar, R. J.; Tretiak, S. *J. Chem. Theory and Comput.* **2007,** *3,* 976-987.

[55] Tawada, Y.; Tsuneda, T.; Yanagisawa, S.; Yanai, T.; Hirao, K. *J. Chem. Phys.* **2004,** *120,* 8425.

[56] Dreuw, A.; Head-Gordon, M. *J. Am. Chem. Soc.* **2004,** *126,* 4007-4016.

[57] van Gisbergen, S. J. A.; Schipper, P. R. T.; Gritsenko, O. V.; Baerends, E. J. *Phys. Rev. Lett.* **1999,** *83,* 694.

[58] Neugebauer, J. *J. Chem. Phys.* **2007,** *97,* 134116.

[59] Neugebauer, J. *J. Chem. Phys.* **2009,** *97,* 084104.

[60] Levy, M. *Phys. Rev. A* **1982,** *26,* 1200-1208.

[61] Wang, Y. A.; Carter, E. A. *Orbital-free kinetic energy functional theory in "Theoretical Methods in Condensed Phase Chemistry";* Theoretical Methods in Condensed Phase Chemistry Kluwer: , 2000.

[62] Karasiev, V. V.; Trickey, S. B.; Harris, F. E. *J. Computer-Aided Materials Design* **2006,** *13,* 111-129.

[63] Teller, E. *Rev. Mod. Phys.* **1962,** *34,* 627-631.

[64] Wang, B.; Stott, M. J.; von Barth, U. *Phys. Rev. A* **2001,** *63,* 052501.

[65] Chai, J. D.; Weeks, J. A. *J. Phys. Chem. B* **2004,** *108,* 6870-6876.

[66] Thakkar, A. J. *Phys. Rev. A* **1992,** *46,* 6920-6924.

[67] Tran, F.; Wesolowski, T. A. *Int. J. Quantum Chem.* **2002,** *89,* 441-446.

[68] Ayers, P. W.; Parr, R. G.; Nagy, A. *Int. J. Quantum Chem.* **2002,** *90,* 309-326.

[69] Perdew, J. P.; Constantin, L. A. *Phys. Rev. B* **2007,** *75,* 155109.

[70] Iyengar, S. S.; Ernzerhof, M.; Maximoff, S. N.; Scuseria, G. E. *Phys. Rev. A* **2001,** *63,* 052508.

[71] Wesolowski, T. A.; Chermette, H.; Weber, J. *J. Chem. Phys.* **1996,** *105,* 9182-9190.

[72] Wesolowski, T. A. *J. Chem. Phys.* **1997,** *106,* 8516-8526.

[73] Wesolowski, T. A.; Weber, J. *Int. J. Quantum Chem.* **1997,** *61,* 303-311.

[74] McRae, E. G. *J. Phys. Chem.* **1957,** *61,* 562.

[75] Tapia, O.; Goscinski, O. *Mol. Phys.* **1975,** *29,* 1653.

[76] Åquist, J.; Warshel, A. *Chem. Rev.* **1993,** *93,* 2523.

[77] Gao, A. *Reviews in computational chemistry;* volume 7 of *Reviews in computational chemistry* K.B. Lipkowitz and D.B. Boyd (VCH Publishers, New York): , 1996.

[78] Sauer, J.; Ugliengo, P.; Garrone, E.; Sounders, V. *Chem. Rev.* **1994,** *94,* 2095.

[79] Phillips, J. C.; Kleinman, L. *Phys. Rev.* **1959,** *116,* 287.

[80] Henderson, T. M. *J. Chem. Phys.* **2006,** *125,* 014105.

[81] Stefanovich, E. V.; Truong, T. N. *J. Chem. Phys.* **1996,** *104,* 2946-2955.

[82] Mei, W. N.; Boyer, L. L.; Mehl, M. J.; Ossowski, M. M.; Stokes, H. T. *Phys. Rev. B* **2000,** *61,* 11425-11431.

[83] Trail, J. R.; Bird, D. M. *Phys. Rev. B* **2000,** *62,* 16402-16411.

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

[85] Shimojo, F.; Kalia, R. K.; Nakano, A.; Vashishta, P. *Comput. Phys. Commun.* **2005,** *167,* 151-164.

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

[87] Olsson, M. H. M.; Hong, G. Y.; Warshel, A. *J. Am. Chem. Soc.* **2003,** *125,* 5025-5039.

[88] Choly, N.; Lu, G.; E, W.; Kaxiras, E. *Phys. Rev. B* **2005,** *71,* 094101.

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

[90] Neugebauer, J.; Baerends, E. J. *J. Phys. Chem. A* **2006,** 8786-8796.

[91] Hodak, M.; Lu, W.; Bernholc, J. *J. Chem. Phys* **2008,** *128,* 014101.

[92] Iannuzzi, M.; Kirchner, B.; Hutter, J. *Chem. Phys. Lett.* **2006,** *421,* 16-20.

[93] Wesolowski, T. A.; Tran, F. *J. Chem. Phys.* **2003,** *118,* 2072-2080.

[94] Dulak, M.; Kaminski, J. W.; Wesolowski, T. A. *J. Chem. Theory Comput.* **2007,** *3,* 735.

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

[96] Kirzhnits, D. A. *Sov. Phys. JETP* **1957,** *5,* 64.

[97] Lee, H.; Lee, C. T.; Parr, R. G. *Phys. Rev. A* **1991,** *44,* 768-771.

[98] Lembarki, A.; Chermette, H. *Phys. Rev. A* **1994,** *50,* 5328-5331.

[99] Fuentealba, P.; Reyes, O. *Chem. Phys. Lett.* **1995,** *232,* 31.

[100] Zupan, A.; Perdew, J. P.; Burke, K.; Causa, M. *Int. J. Quantum Chem.* **1997,** *61,* 835.

[101] Clementi, E.; Raimondi, D. L. *J. Chem. Phys.* **1963,** *38,* 2686.

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

[103] Dulak, M.; Wesolowski, T. A. *unpublished results* .

[104] Tran, F.; Wesolowski, T. A. *Chem. Phys. Lett.* **2002,** *360,* 209-216.

[105] Yang, W.; Parr, R.; Lee, C. *Phys. Rev. A* **1986,** *34,* 4586.

[106] Velde, G. T.; Bickelhaupt, F. M.; Baerends, E. J.; Guerra, C. F.; Gisbergen, S. J. A. V.; Snijders, J. G.; Ziegler, T. *J. Comput. Chem.* **2001,** *22,* 931-967.

[107] Fornberg, B. *Siam Rev.* **1998,** *40,* 685.

[108] Weber, V.; Daul, C.; Baltensperger, R. *Comp. Phys. Comm.* **2004,** *163,* 133.

[109] "MatLab 2006 http://www.mathworks.com/", 2006.

[110] Jacob, C. R.; Wesolowski, T. A.; Visscher, L. *J. Chem. Phys.* **2005,** *123,* 174104.

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

[112] Leopoldini, M.; Russo, N.; Toscano, M.; Dulak, M.; Wesolowski, T. A. *Chem. Eur. J.* **2006,** *12,* 2532-2541.

[113] Dulak, M.; Wesolowski, T. A. *Int. J. Quantum Chem.* **2005,** *101,* 543-549.

[114] "ADF2008.01, SCM, Theoretical Chemistry, Vrije Universiteit, Amsterdam, The Netherlands, http://www.scm.com, E. J. Baerends, J. Autschbach, A. Bèrces, F. M. Bickelhaupt, C. Bo, P. M. Boerrigter, L. Cavallo, D. P. Chong, L. Deng, R. M. Dickson, D. E. Ellis, M. van Faassen, L. Fan, T. H. Fischer, C. Fonseca Guerra, S. J. A. van Gisbergen, A. W. Gotz, J. A. Groeneveld, O. V. Gritsenko, M. GrÃ¼ning,

F. E. Harris, P. van den Hoek, C. R. Jacob, H. Jacobsen, L. Jensen, G. van Kessel, F. Kootstra, M. V. Krykunov, E. van Lenthe, D. A. McCormack, A. Michalak, J. Neugebauer, V. P. Nicu, V. P. Osinga, S. Patchkovskii, P.H. T. Philipsen, D. Post, C. C. Pye, W. Ravenek, J. I. Rodriguez, P. Ros, P. R. T. Schipper, G. Schreckenbach, J. G. Snijders, M. Sol'a, M. Swart, D. Swerhone, G. te Velde, P. Vernooijs, L. Versluis, L. Visscher, O. Visser, F. Wang, T. A. Wesolowski, E. M. van Wezenbeek, G. Wiesenekker, S. K. Wolff, T. K. Woo, A. L. Yakovlev, and T. Ziegler", 2008.

[115] Levy, M.; Perdew, J. P. *Phys. Rev. B* **1993,** *48,* 11638.

[116] Chan, G. K. L.; Handy, N. C. *Phys. Rev. A* **1999,** *59,* 2670-2679.

[117] Dulak, M.; Wesolowski, T. A. *J. Chem. Theory Comput.* **2006,** *2,* 1538-1543.

[118] Gorling, A.; Levy, M. *Phys. Rev. A* **1994,** *50,* 196-204.

[119] Garcia-Lastra, J. M.; Kaminski, J. W.; Wesolowski, T. A. *J. Chem. Phys.* **2008,** *129,* 074107.

[120] Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. *Phys. Rev. B* **1992,** *46,* 6671-6687.

[121] Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. *Phys. Rev. B* **1993,** *48,* 4978-4978.

[122] Kendall, R. A.; Dunning, T. H.; Harrison, R. J. *J. Chem. Phys.* **1992,** *96,* 6796-6806.

[123] "Basis sets were obtained from the Extensible Computational Chemistry Environment Basis Set Database, Version 02/25/04, as developed and distributed by the *Molecular Science Computing Facility, Environmental and Molecular Sciences Laboratory* which is part of the *Pacific Northwest Laboratory*, P.O. Box 999, Richland, Washington 99352, USA, and funded by the *U.S. Department of Energy.* The *Pacific Northwest Laboratory* is a multi-program laboratory operated by *Battelle Memorial Institute* for the *U.S. De-*

partment of Energy under contract DE-AC06-76RLO 1830. Contact *David Feller* or *Karen Schuchardt* for further information.", http://www.emsl.pnl.gov/forms/basisform.html, 2004.

[124] Dunning, T. H. *J. Chem. Phys.* **1989,** *90,* 1007-1023.

[125] Köster, A. M.; Calaminici, P.; Escalante, S.; Flores-Moreno, R.; Goursot, A.; Patchkovskii, S.; Reveles, J. U.; Salahub, D. R.; Vela, A.; Heine, T. "The deMon User's Guide, Version 1.0.3, 2003-2004", http://www.deMon-software.com/, 2004.

[126] Köster, A. M.; Flores-Moreno, R.; Geudtner, G.; Goursot, A.; Heine, T.; Reveles, J. U.; Vela, A.; Salahub, D. R. "deMon 2003, NRC, Canada", http://www.deMon-software.com/, 2003.

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

[128] Kevorkyants, R.; Dulak, M.; Wesolowski, T. A. *J. Chem. Phys.* **2006,** *124,* 024104.

[129] Wesolowski, T. A.; Ellinger, Y.; Weber, J. *J. Chem. Phys.* **1998,** *108,* 6078-6083.

[130] Tran, F.; Weber, J.; Wesolowski, T. A. *Helv. Chim. Acta* **2001,** *84,* 1489-1503.

[131] Dulak, M.; Wesolowski, T. A. *J. Mol. Mod.* **2007,** *13,* xx.

[132] Lieb, E. H.; Oxford, S. *Int. J. Quantum Chem.* **1981,** *19,* 427.

[133] Cortona, P.; Monteleone, A. V. *J. Phys.-Condes. Matter* **1996,** *8,* 8983-8994.

[134] Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005,** *1,* 415-432.

[135] Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2005,** *109,* 5656-5667.

[136] Zhao, Y.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2005,** *7,* 2701-2705.

[137] Kim, Y. S.; Gordon, R. G. *J. Chem. Phys.* **1974,** *61,* 1-16.

[138] Massey, H. S. W.; Sida, D. W. *Philos. Mag.* **1955,** *46,* 190-198.

[139] Gordon, R. G.; Kim, Y. S. *J. Chem. Phys.* **1972,** *56,* 3122 -3133.

[140] Dułak, M.; Wesołowski, T. A. Adaptive grid technique for computer simulations of condensed matter using orbital-free embedding formalism. In *Lecture Series on Computer and Computational Sciences Vol. 3*; Simos, T.; Maroulis, G., Eds.; VSP/Brill: , 2005.

[141] Lynch, B. J.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2003,** *107,* 1384-1388.

[142] Gill, P. M. W.; Johnson, B. G.; Pople, J. A. *Chem. Phys. Lett.* **1993,** *209,* 506-512.

[143] Liu, D. C.; Nocedal, J. *Math. Program.* **1989,** *45,* 503-528.

[144] Koster, A. M.; Reveles, J. U.; del Campo, J. M. *J. Chem. Phys.* **2004,** *121,* 3417-3424.

[145] Dułak, M. *Implementation and applications of the subsystem formulation of density functional theory,* Thesis, Université de Genève, 2006.

[146] Pulay, P. *Mol. Phys.* **1969,** *17,* 197.

[147] Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970,** *19,* 553.

[148] Xantheas, S. S. *J. Chem. Phys.* **1996,** *104,* 8821-8824.

[149] Reichardt, C. *Chem. Rev.* **1994,** *94,* 2319-2358.

[150] Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005,** *105,* 2999-3093.

[151] Tomasi, J. *Theor Chem Acc* **2004,** *112,* 184-203.

[152] Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996,** *100,* 19357-19363.

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

[154] Kovalenko, A. *Three-dimensional RISM theory for molecular liquids and solid-liquid interfaces in "Understanding Chemical Reactivity";* volume 24 of *Understanding Chemical Reactivity* Kluwer Acedemic Publishers: Dordrecht, The Netherlands: , 2003.

[155] Jensen, L.; van Duiknen, P. T.; Snijders, J. G. *J. Chem. Phys.* **2003,** *118,* 514-521.

[156] Jensen, L.; van Duiknen, P. T.; Snijders, J. G. *J. Chem. Phys.* **2003,** *119,* 3800-3809.

[157] Thole, B. T.; van Duijnen, P. T. *Theor. Chim. Acta* **1980,** *55,* 307.

[158] Lahav, D.; Kluner, T. *J. Phys. - Cond. Matt.* **2007,** *19,* 226001.

[159] Wesolowski, T.; Warshel, A. *J. Phys. Chem.* **1994,** *98,* 5183-5187.

[160] Galvan, I. F.; Sanchez, M. L.; Martin, M. E.; del Valle, F. J. O.; Aguilar, M. A. *Comp. Phys. Comm.* **2003,** *155,* 244.

[161] Coutinho, K.; Georg, H. C.; Fonseca, T. L.; Ludwig, V.; Canuto, S. *Chem. Phys. Lett.* **2007,** *437,* 148.

[162] Kongsted, J.; Osted, A.; Mikkelsen, K. V.; Christiansen, O. *J. Chem. Phys.* **2003,** *118,* 1620.

[163] Ten-no, S.; Hirata, F.; Kato, S. *Chem. Phys. Lett.* **1993,** *214,* 391.

[164] Dulak, M.; Kaminski, J. W.; Wesolowski, T. A. *Int. J. Quantum Chem.* **2009,** *109,* 1886.

[165] Aidas, K.; Kongsted, J.; ad K. V. Mikkelsen, A. O.; Christiansen, O. *J. Phys. Chem. A* **2005,** *109,* 8001-8010.

[166] Besley, N. A.; Oakley, M. T.; Cowan, A. J.; Hirst, J. D. *J. Am. Chem. Soc.* **2004,** *126,* 13502.

[167] Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985,** *55,* 2471.

[168] Dreuw, A.; Head-Gordon, M. *Chem. Rev.* **2005,** *105,* 4009.

[169] Klamt, A. *J. Phys. Chem.* **1996,** *100,* 3349.

[170] Hernandes, M. Z.; Longo, R.; Coutinho, K.; Canuto, S. *Phys. Chem. Chem. Phys.* **2004,** *6,* 2088.

[171] Baranovski, V. *Opt. Spectrosc.* **2007,** *103,* 540.

[172] Fortrie, R.; Chermette, H. *Theor. Chem. Account.* **2008,** *120,* 363.

[173] Kovalenko, A.; Hirata, F. *J. Chem. Phys.* **1999,** *110,* 10095.

[174] Hansen, J. P.; McDonald, I. R. *Theory of Simple Liquids, 2nd edn.;* Academic, London: , 1986.

[175] Perkyns, J. S.; Pettitt, B. M. *Chem. Phys. Lett.* **1992,** *97,* 7656.

[176] Kovalenko, A.; Hirata, F. *Chem. Phys. Lett.* **2001,** *349,* 496-502.

[177] Kovalenko, A.; Hirata, F. *J. Theor. Comput. Chem.* **2002,** *1,* 381-406.

[178] Yoshida, K.; Yamaguchi, T.; Kovalenko, A.; Hirata, F. *J. Phys. Chem. B* **2002,** *106,* 5042-5049.

[179] Omelyan, I.; Kovalenko, A.; Hirata, F. *J. Theor. Comput. Chem.* **2003,** *2,* 193-203.

[180] Yamazaki, T.; Imai, T.; Hirata, F.; Kovalenko, A. *J. Phys. Chem. B* **2007,** *111,* 1206-1212.

[181] Yoshida, N.; Imai, T.; Phongphanphanee, S.; Kovalenko, A.; Hirata, F. *J. Phys. Chem. B* **2009,** *113,* 873-886.

[182] Imai, T.; Kovalenko, A.; Hirata, F.; ; Kidera, A. *Journal of Interdisciplinary Sciences: Computational Life Sciences* **2009,** *1,* 156-160.

[183] Imai, T.; Oda, K.; Kovalenko, A.; Hirata, F.; Kidera, A. *J. Am., Chem. Soc.* **2009,** *131,* 12430-12440.

[184] Yamazaki, T.; Kovalenko, A.; Murashov, V. V.; Patey, G. N. *J. Phys. Chem. B* **2010,** *114,* 613.

[185] Kovalenko, A.; Hirata, F. *J. Chem. Phys* **2000,** *112,* 10391-10402.

[186] Kovalenko, A.; Hirata, F. *J. Chem. Phys* **2000,** *112,* 10403-10417.

[187] Gusarov, S.; Ziegler, T.; Kovalenko, A. *J. Phys. Chem. A* **2006,** *110,* 6083-6090.

[188] Chandler, D.; Andersen, H. C. *J. Chem. Phys.* **1972,** *57,* 1930.

[189] Hirata, F.; Rossky, P. *Chem. Phys. Lett.* **1981,** *83,* 329.

[190] Hirata, F.; Petitt, B. M.; Rossky, P. *J. Chem. Phys.* **1982,** *77,* 509.

[191] Hirata, F.; Rossky, P.; Petitt, B. M. *J. Chem. Phys.* **1983,** *78,* 4133.

[192] Chandler, D.; McCoy, J.; Singer, S. *J. Chem. Phys.* **1986,** *85,* 5971-5976.

[193] Chandler, D.; McCoy, J.; Singer, S. *J. Chem. Phys.* **1986,** *85,* 5977-5982.

[194] Beglov, D.; Roux, B. *J. Phys. Chem. B* **1997,** *101,* 7821-7826.

[195] Chandler, D.; McCoy, J. D.; Singer, S. J. *J. Chem. Phys.* **1986,** *85,* 5971;5977.

[196] Singer, S. J.; Chandler, D. *Mol. Phys.* **1985,** *55,* 621.

[197] Casanova, D.; Gusarov, S.; Kovalenko, A.; Ziegler, T. *J. Chem. Theory Comput.* **2007,** *3,* 458-476.

[198] von Lilienfield, O. A.; Tuckerman, M. A. *J. Chem. Phys.* **2006,** *125,* 154104.

[199] Jacob, C. J.; Neugebauer, J.; Visscher, L. *J. Comput. Chem.* **2008,** *29,* 1011.

[200] Lastra, J. M. G.; Kaminski, J. W.; Wesolowski, T. A. *J. Chem. Phys.* **2008,** *129,* 074107.

[201] Swart, M.; van Duijnen, P. T.; Snijders, J. G. *J. Comp. Chem.* **2001,** *22,* 79-88.

[202] Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996,** *118,* 11225-11236.

[203] Jorgensen, W. L.; Madure, J. D.; Swenson, C. J. *J. Am. Chem. Soc.* **1984,** *106,* 6638-6646.

[204] Schipper, P.; Gritsenko, O.; van Gisbergen, S.; Baerends, E. *J. Chem. Phys.* **2000,** *112,* 1344.

[205] Gritsenko, O. V.; Schipper, P. R. T.; Baerends, E. J. *Chem. Phys. Lett.* **1999,** *302,* 199-207.

[206] Gustavsson, T.; Cassara, L.; Gulbinas, V.; Gurzadyan, G.; Pommeret, J.-C.; Sorguis, M.; van der Meulen, P. *J. Phys. Chem. A* **1998,** *102,* 4229-4254.

[207] Moylan, C. R. *J. Phys. Chem.* **1994,** *98,* 13513-13516.

[208] II, G. J.; Jackson, W. R.; Choi, C.; Bergmark, W. R. *J. Phys. Chem.* **1985,** *89,* 294-300.

[209] M.Sulpizi,; Carloni, P.; Hutter, J.; Rothlisberger, U. *Phys. Chem. Chem. Phys.* **2003,** *5,* 4798-4805.

[210] Cave, R. J.; Burke, K.; Jr, E. W. C. *J. Phys. Chem. A* **2002,** *106,* 9294-9305.

[211] Klamt, A.; Shuurmann, G. *J. Chem. Soc. Trans.* **1993,** *2,* 799.

[212] Pavone, M.; Crescenzi, O.; Morelli, G.; Rega, N.; Barone, V. *Theor. Chem. Acc.* **2006,** *116,* 1456-461.

[213] Bernasconi, L.; Sprik, M.; Hutter, J. *Chem. Phys. Lett.* **2004,** *394,* 141.

[214] Aquilante, F.; Cossi, M.; Crescenzi, O.; Scalmani, G.; Barone, V. *Mol. Phys.* **2003,** *101,* 1945-1953.

[215] Rohrig, U. F.; Frank, I.; Hutter, J.; Laio, A.; VandeVondele, J.; Rothlisberger, U. *PhysChemPhys* **2003,** *4,* 1177.

[216] Coutinho, K.; Canuto, S. *THEOCHEM* **2003,** *632,* 235-246.

[217] Coutinho, K.; Saavedra, N.; Canuto, S. *THEOCHEM* **1999,** *466,* 69-75.

[218] Grozema, F. C.; van Duijnen, P. T. *J. Phys. Chem. A* **1998,** *102,* 7984-7989.

[219] de Vries, A. H.; van Duijnen, P. T. *Int. J. Quantum Chem.* **1996,** *57,* 1067.

[220] Ten-no, S.; Hirata, F.; Kato, S. *J. Chem. Phys.* **1994,** *100,* 7443-7453.

[221] Fox, T.; Rosch, N. *Chem. Phys. Lett.* **1992,** *191,* 33.

[222] Porter, C. W.; Iddings, C. *J. Am. Chem. Soc.* **1926,** *48,* 40.

[223] Bayliss, N. S.; McRae, E. G. *J. Chem. Phys.* **1954,** *58,* 1006.

[224] Bayliss, N. S.; Wills-Johnson, G. *Spectrochim. Acta, Part A* **1968,** *24,* 551.

[225] Hayes, W. P.; Timmons, C. J. *Spectrochim. Acta* **1965,** *21,* 529.

[226] Balasubramanian, A.; Rao, C. N. R. *Spectrochim. Acta* **1962,** *18,* 1337.

[227] Renge, I. *J. Phys. Chem. A* **2009,** *113,* 10678.

[228] Suppan, P. *J. Photochem. Photobiol. A* **1990,** *50,* 293.

[229] Aidas, K.; Mogelhoj, A.; Nilsson, E. J. K.; Johnson, M. S.; Mikkelsen, K. V.; Christiansen, O.; Soderhjelm, P.; Kongsted, J. *J. Chem. Phys.* **2008,** *128,* 194503.

[230] Becker, R. S.; Inuzuka, K.; King, J. *J. Chem. Phys.* **1970,** *52,* 5164.

[231] Moskvin, A. F.; Yablonski, O. P.; Bodnar, I. F. *Theor. Exp. Chem.* **1966,** *22,* 469.

[232] Inuzuka, K. *Bull. Chem. Soc. Jpn.* **1961,** *34,* 6.

[233] Inuzuka, K. *Bull. Chem. Soc. Jpn.* **1960,** *33,* 678.

[234] Forbes, W. F.; Shilton, R. *J. Am. Chem. Soc.* **1959,** *81,* 786.

[235] Mackinney, G.; Temmer, O. *J. Am. Chem. Soc.* **1948,** *70,* 3286.

[236] Walsh, A. D. *Trans. Faraday. Soc.* **1945,** *41,* 498.

[237] Buswell, A. M.; Dunlop, E. C.; Rodebush, W. H.; Swartz, J. B. *J. Am. Chem. Soc.* **1940,** *62,* 325.

[238] Blacet, F. E.; Young, W. G.; Roof, J. G. *J. Am. Chem. Soc.* **1937,** *59,* 608.

[239] Eastwood, E.; Snow, C. P. *Proc. R. Soc. London, Ser. A* **1935,** *149,* 446.

[240] Luthy, A. *Z. Phys. Chem.* **1923,** *107,* 284.

[241] Losa, A. M.; Fdez-Galvan, I.; Aguilar, M. A.; Martin, M. E. *J. Phys. Chem. B* **2007,** *111,* 9864-9870.

[242] Brancato, G.; Rega, N.; Barone, V. *J. Chem. Phys.* **2006,** *125,* 164515.

[243] Georg, H. C.; Coutinho, K.; Canuto, S. *J. Chem. Phys.* **2005,** *123,* 124307.

[244] Martin, M. E.; Losa, A. M.; Fdez-Galvan, I.; Aguilar, M. A. *J. Chem. Phys.* **2004,** *121,* 3710.

[245] Monte, S. A.; Muller, T.; Dallos, M.; Lischka, H.; Diedenhofen, M.; Klamt, A. *Theor. Chem. Acc.* **2004,** *111,* 78.

[246] Aquilante, F.; Barone, V.; Roos, B. O. *J. Chem. Phys.* **2003,** *119,* 12323-12334.

[247] Dilling, W. L. *J. Org. Chem.* **1966,** *31,* 1045.

[248] Georg, H. C.; Coutinho, K.; Canuto, S. *J. Chem. Phys.* **2007,** *126,* 034507.

[249] Fattebert, J.-L.; Law, R. J.; Bennion, B.; Lau, E. Y.; Schwegler, E.; Lightstone, F. C. *J. Chem. Theory and Comput.* **2009,** *5,* 2257.

[250] Laio, A.; VandeVondele, J.; Rothlisberger, U. *J. Chem. Phys.* **2002,** *116,* 6941-6947.

[251] Sato, H.; Hirata, F.; Kato, S. *J. Chem. Phys.* **1996,** *105,* 1546.

[252] Satya[al, S.; Petrovic, J.; Thomas, G. *Sci. Am.* **2007,** *296,* 8.

[253] Zuttel, A.; Wenger, P.; Rentsch, S.; Sudan, P.; Mauron, P.; Emmenegger, C. *J. Power Sources* **2003,** *118,* 1.

[254] Frankcombe, T. J.; Kroes, G. J.; Zuttel, A. *Chem. Phys. Lett.* **2005,** *405,* 73.

[255] Sato, T.; Miwa, K.; Nahamori, Y.; Ohoyama, K.; Li, H.; Noritahe, T.; Aohi, M.; Towata, S.; Orimo, S. *Phys. Rev. B* **2008,** *77,* 104114.

[256] Kang, J. K.; Kim, S. Y.; Han, Y. S.; Muller, R. P.; Goddard, W. A. *Appl. Phys. Lett.* **2005,** *87,* 111904.

[257] Au, M.; Jurgensen, A. *J. Phys. Chem. B* **2006,** *110,* 7062.

[258] Kostka, J.; Lohstroh, W.; Fichtner, M.; Hahn, H. *J. Phys. Chem. C* **2007,** *111,* 14026.

[259] Orimo, S.; Y.Nakamori,; Ohba, N.; Miwa, K.; Aoki, M.; Towata, S.; Zuttel, A. *Appl. Phys. Lett.* **2006,** *89,* 21920.

[260] Vajo, J. J.; Skeith, S. L.; Mertens, F. *J. Phys. Chem.* **2005,** *109,* 3719.

[261] Barhordarian, G.; Klassen, T.; Dornheim, M.; Bormann, R. R. *Solid State Commun.* **2006,** *139,* 516.

[262] Yu, X.; Grant, D. M.; Walker, G. S. *Chem. Commun.* **2006,** *37,* 3906.

[263] Alapati, S.; Johnson, J. K.; Sholl, D. S. *J. Phys. Chem. B* **2006,** *110,* 8769.

[264] Johnson, S. R.; Anderson, P. A.; Edwards, P. P.; Gameson, I.; Prendergast, J. W.; Al-Mamouri, M.; Book, D.; Harris, I. R.; Speight, J. D.; Walton, A. *Chem. Commun* **2005,** 2823.

[265] Chater, P. A.; David, W. I. F.; Johnson, S. R.; Edwards, P. P.; Anderson, P. A. **2006,** *23,* 2439.

[266] Pinkerton, F. E.; Meisner, G. P.; Meyer, M. S.; Balogh, M. P.; Kundart, M. D. *J. Phys. Chem. B* **2005,** *109,* 6.

[267] Nakamori, Y.; Ninomiya, A.; Kitahara, G.; Aoki, M.; Noritake, T.; Miwa, K.; Kojima, Y.; Orimo, S. *J. Power Sources* **2006,** *155,* 447.

[268] Jeon, E.; Cho, Y. W. *J. Alloys Compd.* **2006,** *422,* 273.

[269] Jensen, C. M.; Sulic, M.; Kuba, M.; Brown, C.; Langley, W.; Dalton, T.; Culnane, L.; Severa, G.; Eliseo, J.; Ayabe, R.; Zhang, S. *Proc. 2002 U.S. DOE Hydrogen and Fuel Cells Annual Program/Lab R&D Review May 14-18, 2007, Arlington, VA* .

[270] Hwang, S. J.; Jr, R. C. B.; Reiter, J. W.; Rijssenbeek, J.; Soloveichik, G. L.; Zhao, J. C.; Kabbour, H.; Ahn, C. C. *J. Phys. Chem. C* **2008,** *112,* 3164.

[271] "ADF 2007.01, SCM, Theoretical Chemistry, Vrije Universiteit, Amsterdam, The Netherlands, http://www.scm.com", .

[272] Boultif, A.; Louer, D. *J. Appl. Cryst* **2004,** *37,* 724.

[273] Favre-Nicolin, V.; Cerny, R. *J. Appl. Cryst.* **2002,** *35,* 734.

[274] Coelho, A. A. "TOPAS-Academic (2004) http://members.optusnet.com.au/ alancoelho", 2004.

[275] Nakamori, Y.; Li, H.-W.; Kikuchi, M.; Aoki, M.; Miwa, K.; Towata, S.; Orimo, S. *J. Alloys Comp.* **2007,** *296,* 446.

[276] Jensen, J. O. *Spectrochimica Acta A* **2003,** *59,* 1565.

[277] Marks, T. J.; Kolb, J. R. *Chem. Rev.* **1977,** *77,* 263.

[278] Smith, B. E.; Schurvell, H. F.; James, B. D. *J. Chem. Soc. Dalton* **1977,** 710.

[279] Carbonniere, P.; Hagemann, H. *J. Phys. Chem. A* **2006,** *110,* 9927.

[280] Soulie, J. P.; Renaudin, G.; Cerny, R.; Yvon, K. *J. Alloys Compd.* **2002,** *346,* 200.

[281] Filinchuk, Y.; Chernyshov, D.; Cerny, R. *J. Phys. Chem. C* **2008,** *112,* 10579.

[282] Matsuo, M.; Nakamori, Y.; Orimo, S.; Maekawa, H.; Takamura, H. *Appl. Phys. Letters* **2007,** *91,* 224103.

[283] Bird, P. H.; Churchill, M. R. *Chem. Commun.* **1967,** 403.

[284] Cerny, R.; Filinchuk, Y.; Hagemann, H.; Yvon, K. *Angw. Chem. Int. Ed.* **2007,** *46,* 5765.

[285] Gomes, S.; Hagemann, H.; Yvon, K. *J. Alloys Comp.* **2002,** *346,* 206.

[286] Hagemann, H.; Gomes, S.; Renaudin, G.; Yvon, K. *J. Alloys Comp.* **2004,** *363,* 129.

[287] Nakamori, Y.; Miwa, K.; Ninomiya, A.; Li, H.; Ohba, N.; Towata, S.; Zuttel, A.; Orimo, S. *Phys. Rev. B* **2006,** *64,* 45126.

[288] Titov, L. V.; Eremin, E. R. *Russ. Chem. Bull.* **1975,** *24,* 1095.