A fragment-based embedded approach for periodic systems and enhanced weak-pair treatment of molecular LCCSD, both based on direct-local-ring-CCD

Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften (Dr. rer. nat.)
der Fakultät
- Chemie und Pharmazie -
der Universität Regensburg

vorgelegt von

Oliver Masur

aus Ingolstadt

2016

Promotionsgesuch eingereicht am:	11.07.2016
Tag des Kolloquiums:	26.8.2016
Diese Arbeit wurde angeleitet von:	Prof. Dr. Martin Schütz
	Dr. Denis Usvyat

Prüfungsausschuss:	
Vorsitzender:	Prof. Dr. Martin Schütz
Erstgutachter:	Dr. Denis Usvyat
Zweitgutachter:	Dr. Lorenzo Maschio
Drittprüfer:	Prof. Dr. Arno Pfitzner

The results of this thesis have already been published or are submitted for publishing:

Chapter 2

O. Masur, D. Usvyat and M. Schütz

"Efficient and accurate treatment of weak pairs in local CCSD(T) calculations"

Journal of Chemical Physics

139, 164116 (2013), doi: 10.1063/1.4826534

Chapter 2

M. Schütz, O. Masur and D. Usvyat

"Efficient and accurate treatment of weak pairs in local CCSD(T) calculations: II. Beyond the ring approximation"

Journal of Chemical Physics

140, 244107 (2014), doi:10.1063/1.4884156

Chapter 3

O. Masur, M. Schütz, L. Maschio and D. Usvyat

"Fragment-based direct-local-ring-coupled-cluster doubles treatment embedded in the periodic Hartree-Fock solution"

Journal of Chemical Theory and Computation

12, 5145 (2016), doi: 10.1021/acs.jctc.6b00651

Acknowledgements

At this point i would like to extend my gratitude to:

First and foremost, Prof. Dr. Martin Schütz for the interesting and challenging task, his scientific insight, intense aid and his constructive criticism,

PD Dr. Denis Usvyat for his aid, constructive criticism and the willingness to clear even the smallest scientific inexperience, as well as being the first referee for my thesis,

Dr. Lorenzo Maschio for his hospitality in Turin and the short but good collaboration, and taking the time to be the second referee for my thesis,

My former colleagues Thomas Merz, Stephan Loibl, Katrin Freundorfer, Marco Lorenz, as well as my current colleagues Gero Waelz, David David, Martin Christlmaier, Matthias Hinreiner for the good working atmosphere,

Klaus Ziereis for his help with technical difficulties,

My former fellow students Eva, Eva, Katrin, Manu, Michi, Sanne, Tobi for good times outside the university,

As well as Doro, Eva, Fibbo, Floh, Franzi, Frosti, Jakob, Julia, Manu, Marco, Markus, Michi, Michi, Nicole, Schmitti, Susi, Tati, Toby for keeping me sane and making me enjoy life,

Especially Christina for her moral support, and simply being remarkable,

And finally my parents Sigrid and Georg and my sister Sarah for help and support in every way.

Contents

Chapter 1

Introduction

One of the most interesting and important fields studied by natural sciences is the creation and application of materials. This is a very broad field reaching from finding materials to efficiently store hydrogen to power an engine, to doping semiconductors in different ways for different electrical behaviour. Accomplishments in this field immensely effect our daily lifes. Taking just the latter of aforementioned examples, doped semiconductors, are used in photovoltaic cells, which produce electricity for a growing percentage of households in a environmentally beneficial way. Semiconducting materials are also used in another billion-dollar industry, light emitting devices (LEDs), which are used in most of the computer screens we use daily. A further example is adsorption on surfaces, be it intentional, e.g. hydrogen adsorption on palladium for storage, or unintentional, the consequences of environmental conditions for the exterior of our cars.
All these and a lot of other processes and properties of materials have to be understood and then even predicted. In order too understand, predict and advance in material science one has to engage in quantum chemical investigations. Depending on the investigated system two general types of quantum chemical programs will be consulted, molecular and solid state. Programs for solid state investigations are in general far behind in accuracy compared to their molecular counterparts. History tells us that it

takes more than a decade to convert molecular techniques to the solid state, due to the more complex theory, instrinsically more complex algorithms, and much higher computational cost.
The basis for modern quantum chemical methods, the Hartree Fock method, was first published for molecules in 1930.[1] It took about four decades to convert it to the solid state, in 1973, and almost another decade until Pisani and Dovesi made it generally available in 1980.[2;3]
The next step, local Møller-Plesset perturbation theory of second order (LMP2), was first proposed for molecules by Pulay and Saebø in 1986.[4] Here also almost two decades past until it reached the periodic case in 2005, also by Pisani.[5] Density fitting is a very good approximation to make computations more cost effective while keeping accuracy. Its theory was first introduced in 1973 and implemented for LMP2 for solids in 2007.[6–8]
The theory of the Random Phase Approximation (RPA) was introduced by Bohm and Pines in the 1950s.[9;10] In 1968 it has been shown that RPA is not accurate for short-range correlation in the case of a homogenous electron gas, which was probably a reason why it was dismissed when considering real systems for a long time.[10;11]
In 2001 RPA was recast and tested for molecules by Furche.[12] Seven years later, from 2008, it was used for periodic systems and surface adsorption studies, foremost by Kresse.[13–16] In course of the recent years several methods within the RPA framework have been published, like second-order screened exchange (SOSEX).[17] This is also due to an alternative formulation for RPA, as an approximation to the coupled cluster doubles (CCD) theory, called direct ring CCD, which is also used in this work.[18–20]
Within the framework of this dissertation, density fitted (df-) local direct ring CCD (d-rCDD) was implemented for molecular computations, expanded and used as a base for new weak pair treatments for density fitted local coupled cluster singles doubles (df-LCCSD) namely rCCD3, rCCD. These methods were developed and tested on a set of inter-molecular complexes compiled from the S66 and S22 benchmark data sets.[21–23] Ad-

ditional weak pair treatments, namely CCD-R^{-6}, CCD[S]-R^{-6}, were tested on the benchmark data sets.[22–24]
To expand the methods available for treating solid state systems, a direct ring CCD algorithm embedded in a periodic mean field solution was developed and tested. This method utilizes a fragmentation approach, instead of a purely periodic approach, thereby establishing a starting point for further development in the quantum mechanical hierarchy of methods for solid states, in order to catch up with the molecular techniques.[25]
An introduction to the direct ring CCD theory is given in the following section (1.1). Two approximations where exploited in the context of this thesis to reduce the computational effort, namely density fitting and local approximations, which are explained in sections 1.2.1 and 1.2.2. Afterwards (section 1.3) describes the diagrammatic techniques to obtain practical equations from the common coupled cluster expressions. Then, in section 1.4, the structure of the thesis is given.

1.1 Direct ring CCD

This section shows how direct RPA is derived from CCD theory, as seen also in refs.[18–20]. Analytic proof of equivalence is given in ref.[18]. Due to the equality and treatment through CC theory, the name direct ring CCD (d-rCCD) is used from now on. In the following and in this work all together, unless stated otherwise,

- indices $i, j, k, l, ..$ denote localized molecular orbitals and
- indices $r, s, t, u, ...$ denote localized virtual orbitals, in this case projected atomic orbitals (PAOs).

Starting from the simplest CC method, which includes only double excitations, namely CCD, where the equations to be solved for the amplitudes is given by

$$R_{ij}^{rs} = 0 =< \tilde{\Phi}_{ij}^{rs} | e^{-\hat{T}} \hat{H} e^{\hat{T}} | \Phi_0 > \tag{1.1}$$

Where $|\Phi_0>$ is the Hartree-Fock reference wave function, and $<\tilde{\Phi}_{ij}^{ab}|$ the contra-variant configuration state function,

$$<\tilde{\Phi}_{ij}^{rs}| = \frac{1}{6} <\Phi_0|(2\hat{E}_{ri}^{\dagger}\hat{E}_{sj}^{\dagger} + \hat{E}_{rj}^{\dagger}\hat{E}_{si}^{\dagger}) \tag{1.2}$$

used for projection. The $\hat{E}_{ri}$ denote spin-conserving one-particle excitation operators. The normal ordered Hamiltonian is represented by

$$\hat{H} = \hat{F} + \hat{V}, \tag{1.3}$$

which consists of Fock operator $\hat{F}$ and fluctuation potential $\hat{V}$. As stated above only doubles are included, thus $\hat{T} = \hat{T}_2$, with

$$\hat{T}_2 = \frac{1}{2}\sum_{ij}\sum_{rs} t_{rs}^{ij}\hat{E}_{ri}\hat{E}_{sj}, \tag{1.4}$$

The doubles amplitudes are represented by t_{rs}^{ij}. The CCD equations in spin orbital canonical basis to determine t_{rs}^{ij} are[18]

$$\begin{aligned} R_{ij}^{rs} = 0 = (ir||js) + (e_r + e_s - e_i - e_j)t_{ij}^{rs} + \frac{1}{2}t_{kl}^{rs}(ik||jl) \\ + \frac{1}{2}t_{ij}^{tu}(rt||su) + \frac{1}{4}t_{kl}^{rs}(kt||lu)t_{ij}^{tu} - \frac{1}{2}P_{rs}t_{ij}^{ts}(kt||lu)t_{kl}^{ru} \\ - \frac{1}{2}P_{ij}t_{kj}^{rs}(kt||lu)t_{il}^{tu} + P_{ij}P_{rs}t_{jk}^{st}((ir||tk) + \frac{1}{2}(kt||lu)t_{il}^{ru}). \end{aligned} \tag{1.5}$$

Repeated indices (k, l, t, u) are to be summed. And P_{ij}, P_{rs} are permutation operators,

$$P_{ij}g_{ik}g_{jl} = g_{ik}g_{jl} - g_{jk}g_{il}. \tag{1.6}$$

And $(ir||js)$ are anti-symmetrized two electron integrals,

$$\begin{aligned} (ir||js) = (ir|js) - (js|ir) = \\ = \int \Phi_m^*(\mathbf{r}_1)\Phi_r(\mathbf{r}_1)r_{12}^{-1}(1 - P_{12})\Phi_r^*(\mathbf{r}_2)\Phi_s(\mathbf{r}_2)\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2 \end{aligned} \tag{1.7}$$

$$\tag{1.8}$$

Now keeping only particle-hole ring contractions yields

$$\begin{aligned} 0 = (ir||js) + (e_r + e_s - e_i - e_j)t_{ij}^{rs} \\ + (ir||tk)t_{kj}^{ts} + t_{ik}^{rt}(js||tk) + t_{ik}^{rt}(kt||lu)t_{lj}^{us}. \end{aligned} \tag{1.9}$$

This is what is called ring CCD (rCCD), which is equivalent to RPA. By removing the exchange integrals, setting $(ir||js)$ to $(ir|js)$, gives us direct RPA and its equivalent direct rCCD.

1.2 Approximations

Though *ab inito* methods can be improved steadily and methodically, compared to density functional theory (DFT), their computational cost is rather high. CCSD's, e.g., computational effort scales $O(N^6)$, where N represents molecular size. Thus DFT might be the only choice for large molecular systems, though it is sometimes unreliable and fails for certain properties, e.g. charge transfer states or excitation of extended π systems.[26] In order to reduce the cost two approximations were used, density fitting and local approximations.

1.2.1 Density fitting

First the density fitting technique is used to reduce the cost of evaluation of the four-index electron repulsion integrals,

$$\begin{aligned} (mn|rs) &= \int \Phi_m^*(\mathbf{r}_1)\Phi_n(\mathbf{r}_1)r_{12}^{-1}\Phi_r^*(\mathbf{r}_2)\Phi_s(\mathbf{r}_2)\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2 = \\ &= \int \rho_{mn}(\mathbf{r}_1)r_{12}^{-1}\rho_{rs}(\mathbf{r}_2)\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2. \end{aligned} \tag{1.10}$$

The indices m, n, r, s denote general molecular orbitals. The orbital product density ρ_{mn}, is replaced by the approximated density $\tilde{\rho}_{mn}$, which is expanded in an auxiliary basis set $\{\chi_P\}$.

$$\rho_{mn}(\mathbf{r}) \approx \tilde{\rho}_{mn}(\mathbf{r}) = \sum_P d_P^{mn}\chi_P(\mathbf{r}). \tag{1.11}$$

with d_P^{mn} being the fitting coefficients.
In order to obtain the coefficients usually an error functional D is minimized. The basic approach by Baerends, Ellis, and Roos[7] uses a conventional least square fit method, thus minimizes

$$D = \int (\rho_{mn}(\mathbf{r}_1) - \tilde{\rho}_{mn}(\mathbf{r}_1))^2 \mathrm{d}\mathbf{r}_1. \tag{1.12}$$

with respect to the fitting coefficients. The proposition of Dunlap[27], minizing

$$D = \int (\rho_{mn}(\mathbf{r}_1) - \tilde{\rho}_{mn}(\mathbf{r}_1))(\rho_{mn}(\mathbf{r}_2) - \tilde{\rho}_{mn}(\mathbf{r}_2)) r_{12}^{-1} \mathrm{d}\mathbf{r}_1 \mathrm{d}\mathbf{r}_2 \tag{1.13}$$

is generally used though. Minimizing eq.(1.13) reveals

$$d_P^{mn} = \sum_Q J_{PQ}^{-1}(Q|\rho_{mn}). \tag{1.14}$$

Where J_{PQ}^{-1} is the inverse of the coulomb metric, $(P|Q)$. This results in approximating the four-index integrals in the following manner:

$$(mn|rs) \approx \sum_P (mn|P) d_P^{rs} = \sum_{PQ} (\rho_{mn}|Q)(Q|P)^{-1}(P|\rho_{rs}). \tag{1.15}$$

The density fitting approximation, however, does not reduce the scaling with respect to molecular size N, when the four-index integrals are assembled.

1.2.2 Local approximation

A more significant reduction of computational cost is achieved by using local methods, which will be briefly introduced in this section. The canonical orbitals resulting from the initial Hartree Fock calculation are delocalized. But in order to utilize the short-range nature of the dynamic electron correlation in non-metallic systems, one has to move to spatially localised orbitals.
The occupied space can be spanned by localized molecular orbitals (LMOs).[28;29]

Molecular orbitals (MOs) are obtained by expansion from a non-orthogonal basis of atomic orbitals (AOs) $\{\chi_\mu\}$, labeled with greek letters,

$$|\phi_{\bar{i}}\rangle = \sum_{\mu}^{N_{AO}} \chi_\mu C_{\mu\bar{i}}$$
$$|\phi_i\rangle = \sum_{\mu}^{N_{AO}} \chi_\mu L_{\mu i}. \tag{1.16}$$

The coefficient matrix **L** is generated by unitary transformation **W** of the occupied canonical coefficients,

$$L_{\mu i} = \sum_{\bar{i}}^{n_{occ}} C_{\mu\bar{i}} W_{\bar{i}i}. \tag{1.17}$$

Where $\bar{i}$ denotes a canonical occupied orbital, and i denotes a LMO.
The two main choices to obtain the matrix **W** are the Pipek-Mezey procedure, and the Boys procedure.[30;31] The localisation criterion for Pipek-Mezey focuses on minimizing the number of atoms the LMO is located on, while Boys minimizes the LMO's spatial spread. In periodic systems, however, localising occupied orbitals to Wannier functions (WFs), takes more effort.[32;33]
The localisation of virtual orbitals is problematic because of their diffuse nature. To span the virtual space, e.g., one can use projected atomic orbitals (PAOs).[28;29] They are generated, as the name implies, by projecting AOs onto the virtual space,

$$|\phi_r\rangle = \left(1 - \sum_{i}^{n_{occ}} |\phi_i\rangle\langle\phi_i|\right)|\chi_r\rangle = \sum_{\mu}^{N_{AO}} |\chi_\mu\rangle P_{\mu r}. \tag{1.18}$$

The projector matrix **P** consists of,

$$\mathbf{P} = \mathbf{1} - \mathbf{L}\mathbf{L}^\dagger\mathbf{S}^{\mathbf{AO}}. \tag{1.19}$$

Where $\mathbf{S}^{\mathbf{AO}}$ is the overlap matrix between AOs. Additionally the metric of the PAOs **S** is constructed by,

$$\mathbf{S} = \mathbf{P}^\dagger\mathbf{S}^{\mathbf{AO}}\mathbf{P}. \tag{1.20}$$

PAOs are not mutually orthogonal, since the AOs are not. However PAOs are by construction orthogonal to the LMOs.
Now in this LMO/PAO basis local approximations can be introduced.[34–36] In local CCD based methods double amplitudes are restricted to pairs of LMOs ij that can only be excited to an LMO pair specific subspace of PAOs. The pair list consists of LMO combinations where the inter-orbital distance R does not exceed a certain distance criterion. This is motivated by a contribution to the pair energy with a polynomial decay of R^{-6} with respect to the inter-orbital distance R. The virtual space is truncated to pair domains $[ij]$, which consist of the unified orbital $[i], [j]$ domains of the corresponding pair LMOs i, j. The contribution to the pair energy decays exponentially with respect to the distance between related local occupied orbital and virtual orbital. Orbital domains $[i]$ constructed by the Boughton Pulay procedure comprise of the PAOs arising from AOs, which significantly contribute to the corresponding LMO i.[37] The earlier mentioned inter-orbital distance for LMO pairs, is the shortest distance between atoms contained in the orbital domains of the respective orbitals.

1.3 Diagrams for CC

When dealing with theories in quantum chemistry often a rather unmanageable amount of terms for an equation can arise. For CC theory based methods powerful diagrammatic techniques have been developed to construct algebraic expressions for the energy and residual equations.[38] The starting point are the common CC expressions based on the normal ordered second quantized operators. The normal ordered second quantized Hamiltonian consists of the Fock operator **F** and the fluctuation potential **V**, represented by dashed horizontal lines. While the singles and doubles excitation operators are represented by continuous horizontal lines.
Every operator has one or more vertices. One vertex is coupled with one incoming and one outgoing line, together they stand for the action of the

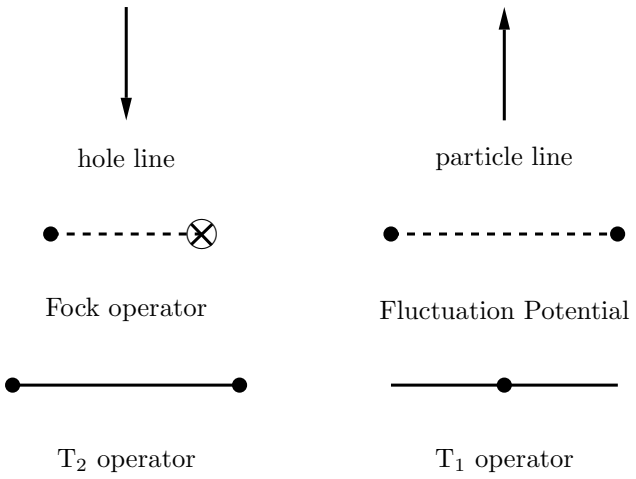

Figure 1.1: Representation of diagrammatic tools. Every filled dot stands for a vertex, which can interact with one hole and one particle line.

operator on one electron. These vertical lines represent hole and particle states, meaning occupied and virtual orbitals respectively. Fig.1.1 shows the diagrammatic representations of these entities. Every filled circle represents a vertex, a line can interact with. Therefore one-electron operators, like Fock and single excitation operator, have one vertex and two-electron operators, like fluctuation and double excitation operator, have two.
The *bra* side of the corresponding expression is represented by the top of the diagram and the *ket* side by the bottom.
Expanding the normal ordered second quantized similiraty transformed Hamiltonian according to the Baker-Campbell-Housedorff formula

$$\hat{H}^T = exp(-\hat{T})\hat{H}exp(\hat{T}) = \hat{H} + [\hat{H},\hat{T}] + \frac{1}{2}[[\hat{H},\hat{T}],\hat{T}] + \frac{1}{6}[[[\hat{H},\hat{T}],\hat{T}],\hat{T}] + ... \tag{1.21}$$

it can be shown that only the terms up until the fourth commutator contribute and only diagrams in which all operators are connected fully through vertical lines contribute.
The rules to translate such diagrams into their algebraic expression are as follows:

1. Hole lines represent occupied $(i, j, k, l, ...)$ and particle lines virtual

$(r, s, t, u, ...)$ orbitals.

2. Each horizontal line, i.e., operator contributes an integral or an amplitude to the final expression.
   The outgoing (*out*) and incoming (*in*) lines on vertices of one operator are evaluated. The indices denote the vertex, increasing from left to right. The scheme for the operators is thus as follows:

   - Singles and doubles excitation operators: t^{in}_{out},$t^{in_1 in_2}_{out_1 out_2}$
   - Fock operator: $(out|\hat{F}|in)$,
     e.g. yields: $\langle r|\hat{F}|s\rangle = F_{rs}$
   - Fluctuation operator: $(out_1 in_1|out_2 in_2)$,
     e.g. yields: $(ri|st)$.
     The chemical notation for two-electron integrals is used here.

3. Summation runs over all internal lines, lines connecting two vertices.

4. The sign of the diagram is $(-1)^{h+l}$, where h and l are the number of hole lines and loops, respectively.

5. Every internal loop contributes with a factor of two.

6. Equivalent vertex pairs contribute with a factor of $\frac{1}{2}$.
   Connected vertices on two operators are a vertex pair. Another vertex pair on the same operators connected in the same way yield the factor.

7. External particle lines, starting from a vertex of an excitation operator, contribute an element of the PAO overlap matrix $\mathbf{S}$.
   An external particle line directly connects the *bra* and the *ket* without and operator in between. The reason for this rule are the non-mutually orthogonal PAOs.

1.4 Structure of the thesis

This thesis is composed in the following way.
After this short introduction to the basic concepts, the new formalism for weak pair treatment within density fitted LCCSD calculations is presented, and tested for accuracy and efficiency through test systems of the S66 and S22 benchmark data sets.
Chap.3 presents the implementation and testing for the new embedded direct-LrCCD method for periodic systems.
Finally, the results of this thesis are summarized in chap.4.

Chapter 2

Efficient and accurate treatment of weak pairs in local CCSD(T) calculations

The content of this chapter has already been published in the Journal of Chemical Physics, refs.[21; 24].
The manuscript was revised for the context of this thesis, i.e. basic principles, which were discussed in chap.1 were shortened or omitted, while other aspects are discussed more detailed.
Prof. Dr. Martin Schütz implemented the LCCD-R^{-6} method as well as the LCCD[S]-R^{-6} method with its fourth order singles correction [S]. The implementation of the methods rCCD3 and rCCD was realised by the author. The testing, presented here, for all the methods was realized by the author.

2.1 Introduction

There is a wide range of *ab inito* methods to choose from in quantum chemistry. The cost and accuracy for a calculation of the electronic correlation

effects is different for every method. In the hierarchy of methods the lowest and therefore starting point is the Hartree-Fock (HF) method. The prime example for the highly accurate, but at the same time still affordable, at least for not very large systems, would be Couple Cluster Singles Doubles (CCSD), or even CCSD with an *a posteriori* perturbative triples correction (CCSD(T)). The latter has earned the expression "Gold Standard" in quantum chemistry, originated by T. H. Dunning.

When considering single reference cases, this method reaches chemical accuracy when using sufficiently large basis sets. However canonical CCSD(T)'s computational complexity is still very high. Its computational cost is scaling with $O(N^7)$ relative to the molecular size N. To be more precise the iterative CCSD calculation is scaling with $O(N^6)$, and the *a posteriori* perturbative triples correction (T) with $O(N^7)$.[39] This reduces the applicability for very extended molecular systems. During the last almost two decades, in order to overcome this high scaling disadvantage, *local* CCSD(T) methods have been developed.[34;36;40–45]

The prefix *local* implies the formulation of CCSD(T) in terms of the approximation described in chapter 1.2.2. The aforementioned pair approximation is now utilised by assigning the individual LMO pairs to distinct pair classes *strong*, *weak*, *distant* and *very distant* based on the interorbital distance R of the respective LMOs.

Each of these classes is then treated on a different level. The most contributing class, strong, exclusively undergoes the full Local Coupled Cluster (LCC) treatment. With decreasing contribution, accuracy and cost is lowered. Typically weak pairs are treated at the level of second-order Møller-Plesset perturbation theory (LMP2). The class of very distant pairs normally are neglected entirely. Later the subdivision of weak pairs into close pairs and the rest was introduced in the context of local triples.[34;46] The doubles amplitudes corresponding to this close pair subset, contribute together with those of the strong pairs, to the triples residual, and optionally, to the LCC (strong pair) doubles residuum.

The decay of the individual pair energies is of R^{-6} and thus needs to be treated more carefully than the local approximation for the virtual space with a stronger, exponential decay.
The weak pair treatment at LMP2 level is known to have its problems. In cases of π-stacked aromatic rings, for example, LMP2 doesn't provide a sufficiently accurate description for the interaction energy.[47;48] For intermolecular complexes with low polarizability LMP2 often underestimates the interaction energy, while overestimating it for such cluster with high polarizability. Furthermore, Axilrod-Teller, non-additive dispersion, contributions are not included at the LMP2 level. However an MP2-like treatment for most of the pairs is applied by many approximated CCSD methods.
With the above mentioned arguments this work is revisiting the capability of such treatment to provide in general sufficient accuracy. This work substitutes the LMP2 treatment of weak pairs with methods based on ring-CCD, which corresponds to RPA, with additional exchange terms such that the antisymmetry of the amplitudes in a spin-orbit formalism remains intact.
The two points the methods incorporate are:

- The included diagrams contribution to the pair energy decays with the same rate as LMP2, i.e. as R^{-6} with respect to the inter-orbital distance R.
- The order of a diagram arises from its leading contribution to the correlation energy within the Møller-Plesset partitioning.

The decay rate of a diagram arises from the slowest decaying, with respect to the inter-orbital distance, energy component originating from it. The resulting methods are distinguished as follows:

1. LrCCD3 - only diagrams with an energy contribution up to the third order are kept within the ring approximation

2. LrCCD - including some additional fourth order diagrams.
3. LCCD-R^{-6} - including all fourth order diagrams decaying as R^{-6}, thereby going beyond the ring approximation
4. LCCD[S]-R^{-6} - adding a correction for the fourth order singles contribution to the correlation energy, also decaying as R^{-6}.

LrCCD3 results in being superior to the LMP2 treatment, while keeping the increase in cost low.
The rather disappointing behavior of LrCCD, as seen due to absent additional fourth order diagrams, lead to LCCD-R^{-6}; which includes said diagrams. These diagrams also decay only with R^{-6}.
LCCD-R^{-6} goes beyond the ring approximation and provides similar accuracy to LrCCD3, but not, as seen later, due to fortuitous error cancellation. Also these diagrams can be efficiently evaluated by pre-contraction to a matrix reminiscent of a LMP2 density matrix and the latter to the Fock matrix during the orbital invariant LMP2 residual computation.
However to increase accuracy over the LrCCD3 method the inclusion of the fourth order singles contribution, which also decays with R^{-6}, to the close pair energy is vital. This can be evaluated very inexpensively via *a posteriori* perturbative estimate, by reusing the already calculated intermediates from above mentioned close pair feedback to the LCC singles residual. The accuracy of the resulting hybrid method, LCCSD|LCCD[S]-R^{-6}, is indeed on par with a much more expensive, full LCCSD computation.
The classification of the LCCSD|LCCD[S]-R^{-6} diagrams according to their decay rate and the leading order of their contribution to the energy, finally conforms to the following hierarchy:

1. strong pairs are treated at full LCCSD level.
2. close pairs are treated by all the diagrams up to fourth order decaying with R^{-6}.
   They contribute to the (4th order) triples and via 3rd and 4th order diagrams to the LCC residual.

3. weak/distant pairs are treated all the diagrams of 3rd or 2nd order decaying as R^{-6}.

4. very distant pairs are disregarded.

The leading contribution in the local triples treatment is of fourth order and decays as R^{-6}, thus the strong and close pairs have to be included. This is consistent with the usual L(T) treatment.
In this chapter the accuracy of these methods is investigated for a set of diverse inter-molecular complexes and clusters. In sec.2.2 the theory of these approximations is outlined. Their performance is tested against the previous treatment of weak pairs at the MP2 level for a set of test systems in sec.2.3.
Sec.2.4 summarizes and concludes this chapter.

2.2 Theory

These sections first provide the CCSD residual equations for the strong pairs. Afterwards a discussion about the inclusion of diagrams for the weak pair treatment in the different methods is given. In the following, the index notations are:
i, j, k, l for localized occupied molecular orbitals (LMOs) and
r, s, t, u for localized virtual orbitals, i.e. projected atomic orbitals (PAOs).
Mullikens notation is used for the integrals and Einstein convention is employed, meaning implicit summation over repeated indices. Boldfaced letters indicate the matrix form of the respective quantity.

2.2.1 CCSD strong pair residuum

For a closed-shell system the wave function for local CCSD is defined in the following exponential expression:

$$|\Psi_{CC}> = exp(\hat{T})|\Phi_0>, \tag{2.1}$$

where $\hat{T}$ is the cluster operator, restricted to single and double excitations:

$$\hat{T} = \sum_i \sum_{r\in[ii]} t_i^r \hat{E}_{ri} + \frac{1}{2} \sum_{ijrs\in[ij]} t_{ij}^{rs} \hat{E}_{ri}\hat{E}_{sj}. \tag{2.2}$$

In comparison to eq.(1.4), now singles excitations are included and the range of virtual orbitals is restricted to pair domains $[ij]$ as well as LMO pairs ij are restricted to strong pairs. The singles and doubles amplitudes t_i^r, t_{ij}^{rs} are the solution for LCCSD. They are obtained by solving the residual equations,

$$R_i^r =< \tilde{\Phi}_i^r|exp(-\hat{T})\hat{H}exp(\hat{T})|\Phi_0 >= 0;\ \forall\ i; r \in [ii], \tag{2.3}$$

$$R_{ij}^{rs} =< \tilde{\Phi}_{ij}^{rs}|exp(-\hat{T})\hat{H}exp(\hat{T})|\Phi_0 >= 0;\ \forall\ ij \in \{s\}; r,s \in [ij], \tag{2.4}$$

where local approximation restrictions were added compared to eq.(1.1) and with $\hat{H}$ being the normal ordered Hamiltonian, seen in eq.(1.3). The Hartree-Fock reference wave function $|\Phi_0>$ and the contra-variant configuration state functions $< \tilde{\Phi}_i^a|$,

$$< \tilde{\Phi}_i^r| = \frac{1}{2} < \Phi_0|\hat{E}_{ri}^{\dagger}, \tag{2.5}$$

and $< \tilde{\Phi}_{ij}^{ab}|$ (eq.(1.2)) are used for projection.
The distribution of pairs to the individual pair classes (strong - distant) is usually decided on the basis of the inter-orbital distance R between the respective LMOs, or their connectivity, i.e. the count of bonds between them.[35;36] An arrangement on the basis of distance was utilized in this work.

A subset of relevant atoms is specified for each LMO, realized by a Löwdin population analysis and truncating the ordered atom list beyond a certain population. The distance R_{ij} between the two LMOs i and j is equal to the distance between the two nearest atoms of the two atom subsets, belonging to LMO i and LMO j.

The strong pair class normally only consists of LMO pairs with mutual overlapping atom subsets, or in some cases the criteria is increased up to 3 bohr. Remaining pair classes have normally been treated at an LMP2 level. The residual for the LMP2 amplitudes is as follows:

$$R_{ij}^{rs} =< \tilde{\Phi}_{ij}^{rs}|\hat{V} + [\hat{F}, \hat{T}_2]|\Phi_0 >= 0; \ \forall \ ij \in \{\text{sw}\}; r, s \in [ij]. \tag{2.6}$$

Normally there is an initial calculation for all pair classes at LMP2 level. Afterwards the strong pairs are refined at the LCCSD level, i.e. solving eqs.(2.4). The amplitudes of the close pair class, as mentioned in the introduction, can optionally be included in solving these equations, making them couple with the strong pair amplitudes. This means the second sum over pairs ij in eq.(2.2) is then extended to run over strong and close pairs. The reverse coupling, of strong pairs with close pairs, was assumed to be weak, which is indeed valid for LMP2 equations. This is due to the fact that strong pair amplitudes only can affect the LMP2 residual (eq.(2.6)) through the term involving the Fock operator, whose matrix representation, in the local basis, is dominated by diagonal elements. Hence a further iteration beyond the initial LMP2 calculation is not necessary.

As can be seen in sec.2.3 the effect of close pairs in the strong pair residual is of great importance and highly advisable. In the end, when the LCCSD equations are converged, an *a posteriori* local triples correction L(T), or L(T0) can be calculated.[34;40;41] It needs to be noted that strong and close pair amplitudes enter the triples residual.

2.2.2 Weak pair residua

As already argued in the introduction for this chapter LMP2 treatment for weak pairs does not always yield a sufficient description of the correlation energy. In this section the proposed, computationally efficient treatment for weak pairs based on the rCCD approach are presented.
Refs.[18; 20] already considered several versions of rCCD. The introduction of these methods was mainly motivated by the rCCD equivalence to RPA. This equivalence however required that the antisymmetry of the doubles amplitudes in spin-orbital basis with respect to permutation of the occupied or virtual indices should be sacrificed. As the presented methods are an approximation to LCCSD, without explicit reference to RPA, there is no reason to drop the antisymmetry of the amplitudes.
The corresponding additional diagrams, which are absent in RPA, decay as slowly as the respective standard RPA-terms. There is no reason not to include them from the point of view of locality. At the same time they can easily be evaluated with negligible additional computational effort. The starting point, the rCCD residual equation, is given by

$$\begin{aligned} R_{ij}^{rs} =& < \tilde{\Phi}_{ij}^{rs} | \hat{V} + [\hat{H}, \hat{T}_2]_{\text{ring}} + [[\hat{H}, \hat{T}_2], \hat{T}_2]_{\text{ring}} | \Phi_0 >= 0; \\ & \forall\ ij \in \{\text{w}\}; r, s \in [ij]. \end{aligned} \tag{2.7}$$

Here the subscript "ring" implies the restriction of the corresponding commutators to ring diagrams, maintaining all corresponding permutators of eq.(1.5). For this ansatz and the resulting methods we provide the diagrams and corresponding explicit expressions, in spin-free closed-shell formalism, in the chapters below.

The direct-rCCD, rCCD3 and rCCD residua

The first approximation to rCCD is direct-rCCD. In contrast to direct-RPA, and generally RPA, the exchange terms originating from above mentioned anti-symmetrized index permutations in the residual are kept. In this

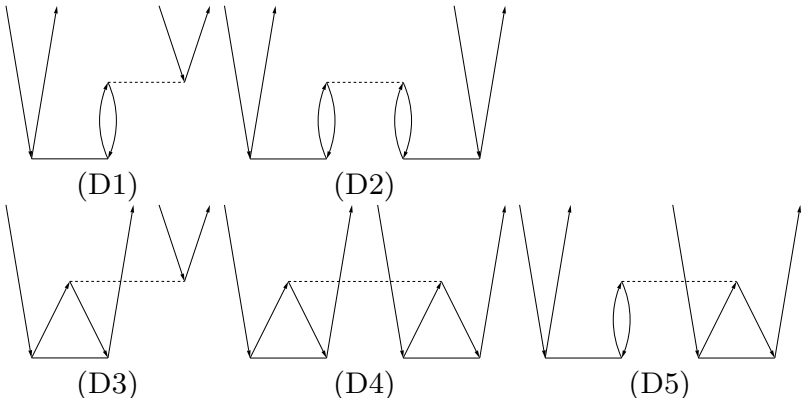

Figure 2.1: Goldstone diagrams of *direct*-rCCD beyond MP2 (D1,D2) including additional exchange terms as obtained by transposition of respective amplitude matrix (D3-D5). The corresponding algebraic expressions are given in tab.2.1. [1]

method, the exchange terms beyond those from above mentioned anti-symmetrized index permutations are disregarded. The remaining diagrams are shown in fig.2.1. In addition, the standard CCD energy expression is used to compute the weak pair energies, where the exchange terms are also included. In this regard, direct-rCCD is related to the second-order screened exchange [SOSEX] variant of direct-RPA.[49]

The influence of strong on close pairs is no longer restricted to just via the Fock term as in LMP2. As seen in eq.(2.7) the strong pairs amplitudes entering the equation also couple via terms involving the fluctuation potential. That's why the reverse coupling of the strong pair amplitudes on the rCCD residual is much stronger than on the LMP2 residual. In fact thus an initial rCCD calculation as in the case of LMP2 is entirely insufficient. Therefore a simultaneous solution in a common iterative procedure of eqs.(2.3,2.4,2.7) needs to employed. Note that the rCCD residual, eq.(2.7) is restricted to weak pairs, $ij \in w$, but the amplitudes that enter it are strong, close and weak pairs.

The LCCSD residual, eqs.(2.3,2.4), on the other hand, is entered by strong and close pairs.[34–36] The diagrams and their corresponding exchange diagrams, shown in fig.2.1 are implemented using the virtue of density fitting,

[1] Reprinted from ref.[21] with the permission of AIP Publishing.

described in sec.1.2.1 and can thus be evaluated very efficiently. The first term of diagram D1 in tab.2.1, also seen in eq.(1.9), can be written as

$$R_{ij}^{rs} += \sum_{P} A_{ri}^{P} d_{sj}^{P};\ \forall\ ij \in \{\mathrm{w}\}; r,s \in [ij], \quad \text{with}$$
$$A_{ri}^{P} = \sum_{r' \in \cup[i]} S_{rr'} \sum_{j} \sum_{s \in [ij]} T_{ik}^{r't}(kt|P);\ r \in \cup[i], \tag{2.8}$$

where integral over the charge density kt and the fitting function P form the three index electron repulsion integrals $(kt|P)$, d_{sj}^{P} is the fitting coefficient and $\cup[i]$ is the union of the pair domains $[ij]$ with common LMO i.
The motivation to incorporate the diagrams of direct-rCCD shown in fig.2.2 into the weak pair residual is that their contribution to the pair energy decays with the same rate as LMP2, i.e. as R^{-6} with respect to the inter-orbital distance R. Within the MP2 partitioning the leading order of the direct-rCCD diagrams to the correlation energy is higher. A contribution of the third order arises from the diagrams D1 and D3. The diagrams D2, D4 and D5 provide a fourth order contribution.
The investigation in sec.2.3 shows that the accuracy of d-LrCCD is rather low. The interaction energies are substantially underestimated. It was attributed to the fact that additional exchange diagrams of rCCD which are not included in direct-rCCD, also decaying not faster than R^{-6}, are missing. Unfortunately these two diagrams, D6 and D7, shown in fig.2.2, cannot be evaluated as easily as the other diagrams discussed above. The D6 diagram makes an additional contribution to the correlation energy of the third order and the D7-D9 ones of the fourth order.

The decay of the energy contributions of the D6-D8 diagrams with the distance R between the two remote subsystems A and B can be analyzed in fig.2.4. Generally a doubles amplitude T_{ij}^{rs}, as well as an integral $(ir|js)$ decays as R^{-3} with the distance between i and j or between r and s, and exponentially with the distance between i and r, or j and s. This comes from the strong orthogonality between the occupied and virtual manifold, that gives rise to the dipole-dipole interaction as the leading term in the multipole approximation of the integral. At the same time an integral $(ij|rs)$

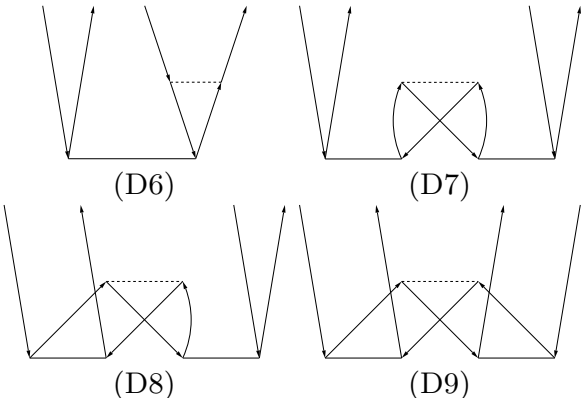

Figure 2.2: Goldstone diagrams of the two additional exchange rCCD terms, not generated from the *direct*-rCCD ones by transposition of respective amplitude matrix (first column), and the diagrams, obtained by transposition of D7 (second column). The algebraic expressions corresponding to the diagrams are given in tab.2.1. [2]

has an exponential decay with the distance between i and j and generally just as R^{-1} with the separation of the occupied and virtual orbitals.

The leading energy contribution arising from these open residual diagrams is obtained by bra-side closure, i.e. a contraction with an integral of the $(ir|js)$ type. This contraction leads to a further R^{-3} factor of decay. Considering this, one can conclude that the asymptotical decay of the energy contribution is indeed R^{-6}.

As seen in fig.2.4, D6, e.g. implies contraction of R^{-3} decaying amplitudes $T_{i_A k_B}^{r_A t_B}$ with R^{-3} decaying integrals $(i_A r_A|j_B s_B)$ which does give the R^{-6} decay with distance between subsystems A and B. It is to note, that the coupling integral $(k_B j_B|s_B t_B)$ is entirely localized on subsystem B, as this term decays exponentially with distance between orbitals k and j. The only diagram among those shown in fig.2.2 that does not have a slow decay is D9, whose contribution to the energy decays as R^{-12}.

Now we come to the important class of ladder diagrams, which are shown in fig.2.3. They do not appear in the ring-CC-approximation , but nevertheless are legitimate CC terms. Moreover, all three diagrams represent a contribution to the correlation energy already in the third order, which

[2]Reprinted from ref.[21] with the permission of AIP Publishing.

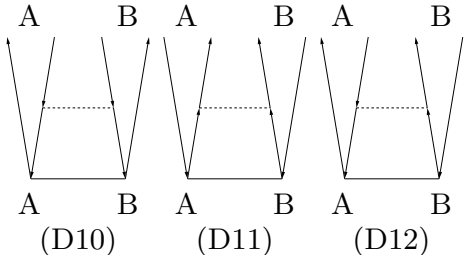

Figure 2.3: Goldstone ladder diagrams omitted in rCCD. They altogether decay as R^{-9} with the distance R between the two remote subsystems A and B containing the two LMOs of the close/weak pair (see text). [3]

decays rather slow; each as R^{-7}. This comes, as before, from a factor of R^{-3} for the amplitude, a factor of R^{-3} for the contraction for the energy contribution and a factor of R^{-1} for the integral since its orbital products are not chargeless.

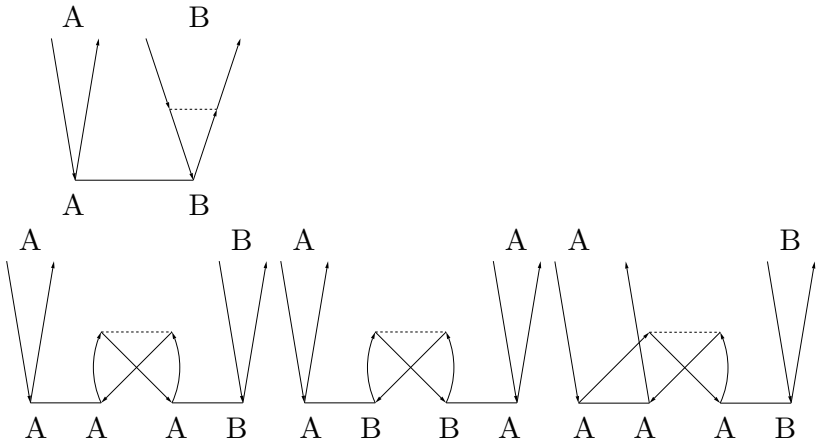

Figure 2.4: Contributions of the Goldstone diagrams D6, D7,and D8, which decay as R^{-6} with the distance R between the two "remote" subsystems A and B containing the two LMOs of the close/weak pair. [3]

The problem with these diagrams is that they cannot be evaluated so effortlessly, particularly diagram D11. D11's integrals consist of four virtual indices. These electron repulsion integrals constitute a very large integral set and generally this term forms the bottleneck of the CCSD method.

[3]Reprinted from ref.[21] with the permission of AIP Publishing.

Below is shown that in fact these ladder terms, when combined together, decay much faster (R^{-9}) and thus these diagrams can indeed be restricted to the strong pairs only. This is extremely essential for the efficiency of the whole local scheme.
To show this, consider the terms in a basis of localized orthonormal orbitals. Formally this always can be fulfilled, as PAOs can always be orthonormalized in the local basis and form, e.g., the appropriate pseudo canonical local basis. The LMOs are orthonormal anyways. A contribution to the residual decays as R^{-4} only for the diagonal terms where the orbital products in the integral consist of identical indices for both virtual and occupied orbitals. Every other term decays faster since it involves chargeless densities in the integral. Now combining the slowly decaying diagonal terms together one gets,

$$\sum_{ia\,jb} t_{ij}^{ab}(ii - rr|jj - ss) \tag{2.9}$$

where the coulomb-interacting densities are differences between squared virtual and squared occupied orbitals. Through the normality of the orbitals the charge of such orbital products differences vanishes and the integrals decays not slower than R^{-3}. The energy contribution altogether then consists of three factors of R^{-3} for amplitudes, integrals and the bra-contraction in the energy term, and hence results in an overall decay of R^{-9}.

As it turns out from the calculations, in order to improve upon direct-LrCCD the diagram D6 is of prime importance. Based on this discussion the methods

- rCCD3, which represents the third order approximation to rCCD, and
- the full rCCD

were implemented. rCCD3 contains the diagrams D1, D3, D6, in addition to the MP2 diagrams. The number of needed electron repulsion integrals

Table 2.1: Algebraic expressions for the Goldstone diagrams in figs.2.1 and 2.2 contributing to the weak pair residual R_{rs}^{ij}. For compactness Einstein's convention (implicit summation over repeated indices) is assumed. Electron repulsion integrals are given in Mulliken's notation. $S_{rr'}$ is an element of the metric of the (mutually not orthogonal) local virtuals. The respective weights of the diagrams are also given. [4]

(D1)	+2	$S_{rr'} T_{ik}^{r't}(kt\|sj) + (ri\|kt) T_{kj}^{ts'} S_{s's}$
(D2)	+4	$S_{rr'} T_{ik}^{r't}(kt\|lu) T_{lj}^{us'} S_{s's}$
(D3)	-1	$S_{rr'} T_{ik}^{tr'}(kt\|sj) + (ri\|kt) T_{jk}^{ts'} S_{s's}$
(D4)	+1	$S_{rr'} T_{ik}^{tr'}(kt\|lu) T_{jl}^{us'} S_{s's}$
(D5)	-2	$S_{rr'} \left(T_{ik}^{r't}(kt\|lu) T_{jl}^{us'} + T_{ik}^{tr'}(kt\|lu) T_{lj}^{us'} \right) S_{s's}$
(D6)	-1	$S_{rr'} T_{ik}^{r't}(kj\|st) + (rt\|ki) T_{jk}^{s't} S_{s's}$
(D7)	-2	$S_{rr'} T_{ik}^{r't}(lt\|ku) T_{lj}^{us'} S_{s's}$
(D8)	+1	$S_{rr'} \left(T_{ik}^{tr'}(lt\|ku) T_{lj}^{us'} + T_{ik}^{r't}(lt\|ku) T_{lj}^{s'u} \right) S_{s's}$
(D9)	-$\frac{1}{2}$	$S_{rr'} T_{ik}^{tr'}(lt\|ku) T_{lj}^{s'u} S_{s's}$

$(kj|st)$ for D6 is fortunately rather small. In fact, because the integrals decay exponentially with the distance between LMOs k and j, only strong pairs kj are of importance in the computation of diagram D6. The increase of computational cost when going from LMP2 to rCCD3 for weak pair treatment is thus relatively modest. However the improvement of LrCCD3 over LMP2, as shown in sec.2.3, is remarkable.

The diagrams D1-D5 allow for very computationally convenient factorization through the density fitting approximation. It is discussed later in sec.3.2.3; see eqs.(3.15, 3.16). Diagrams D6-D9 on the other hand cannot be factorized in this way. DF however is still used as a computationally convenient approximation to compute the integrals in these diagrams (see sec.1.2.1).

The technical evaluation of those diagrams however is rather difficult and

[4] Reprinted from ref.[21] with the permission of AIP Publishing.

they are therefore not included in the methods of this section. The discussion concerning these is part of the next section.

The CCD-R^{-6} residuum

As seen later in sec.2.3, the LrCCD method of the previous section is not an improvement over LrCCD3, despite involving more diagrams and at a higher cost. This was attributed to further fourth order diagrams that also decay as R^{-6}, but missing in the ring-CCD formalism, shown in fig.2.5.

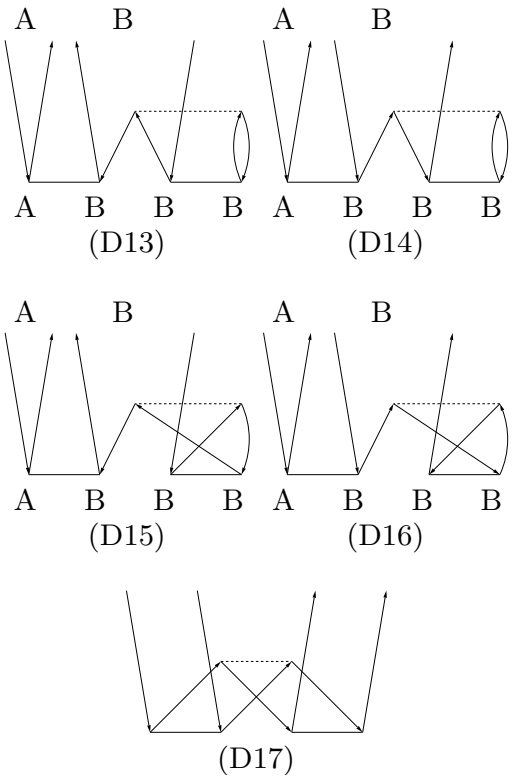

Figure 2.5: Goldstone diagrams of the CCD 4th-order terms, absent in the rCCD formalism. [5]

Diagram D17's contribution to the weak pair residual can certainly be neglected since its contribution to the correlaton energy decays as quickly as R^{-12}. On the other hand diagrams D13-D16 can have a contribution

[5]Reprinted from ref.[21] with the permission of AIP Publishing.

with a decay of R^{-6}. Fortunately D13-D16 can be evaluated very efficiently, using a similar approach to that of ref.[50] and ref.[51]. This can be done via pre-contraction to a matrix reminiscent of a LMP2 density matrix, and then accumulating it to the Fock matrix in the LMP2 residual terms.
For this the diagrams D13-D16 are combined in two pairs. The first pair is the diagrams D13 and D15, which yield together

$$\begin{aligned}\Delta R_{ij}^{rs} &\Leftarrow -\mathcal{P}S_{rr'}T_{ik}^{r's'}S_{s's}(kt|lu)\tilde{T}_{jl}^{tu} \\ &= -\mathcal{P}S_{rr'}T_{ik}^{r's'}\mathcal{D}_{kj}S_{s's} \\ &= -\mathbf{S}\left(\mathcal{D}_{ki}\mathbf{T}_{kj} + \mathbf{T}_{ik}\mathcal{D}_{kj}\right)\mathbf{S}, \end{aligned} \tag{2.10}$$

with

$$\mathcal{D}_{ki} = (kt|lu)\tilde{T}_{il}^{tu} \tag{2.11}$$

Here $\mathcal{P} = 1 + (ij)(rs)(r's')$ represents the permutation operator and $\mathbf{S}$ the PAO overlap matrix. The second pair, i.e. the diagrams D13 and D15, can be treated accordingly:

$$\begin{aligned}\Delta R_{rs}^{ij} &\Leftarrow -\mathcal{P}S_{rr'}T_{ij}^{r't}(kt|lu)\tilde{T}_{lk}^{us'}S_{s's} \\ &= -\mathcal{D}_{tr}T_{ij}^{ts'}S_{s's} - S_{rr'}T_{ij}^{r't}\mathcal{D}_{ts} \\ &= -\mathcal{D}^{\dagger}\mathbf{T}_{ij}\mathbf{S} - \mathbf{S}\mathbf{T}_{ij}\mathcal{D}. \end{aligned} \tag{2.12}$$

with

$$\mathcal{D}_{tr} = (kt|lu)\tilde{T}_{lk}^{ur'}S_{r'r}. \tag{2.13}$$

In each iteration the $\mathcal{D}$ intermediates are added to the internal and external Fock matrices before the evaluation of the LMP2 residual. [52] The LCCD-R^{-6} method described here then consists of the LrCCD method of the previous section with the addition of diagrams D13-D16. The LCCD-R^{-6} treatment for weak pairs is qualitatively superior to LrCCD3:
It provides similar accuracy, but not due to fortuitous error cancellation. In order to improve the quantitative accuracy, the fourth-order singles contribution to the close pair energy must be included, as is discussed in the next section.

2.2.3 The 4th order singles correction [S]

An analysis of the CCSD method reveals that it contains another R^{-6} decaying fourth order energy contribution, that is not contained in LCCD-R^{-6}. This is the singles contribution to the doubles residual as shown in fig.2.6. The sum of these two diagrams indeed decays with R^{-6}, and thus according to the protocol discussed in sec.2.1, should not be neglected in the close pair treatment.

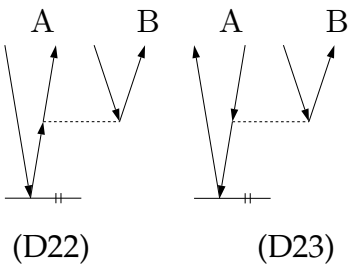

Figure 2.6: The singles contributions to the close-pair doubles residual, which contribute in the fourth-order to the correlation energy (not included in the LCCD-R^{-6} method). The sum of these diagrams decays as R^{-6}. [6]

The diagram D23 involves an integral with three external indices. Such integrals pose an expensive computational effort. To avoid the high computational cost of evaluation of such a term in each CCSD iteration, we propose an *a posteriori* perturbative estimate of their contribution to the correlation energy. This MP4-like energy correction is computed from the converged LCCSD|LCCD-R^{-6} singles and close pair doubles amplitudes, thus is defined along the same lines as the [T] correction in the CCSD[T] method, and hence denoted "[S]". For a computationally convenient scheme to evaluate these corrections, we redraw the corresponding diagrams upside down, which is shown in fig.2.7. The resulting algebraic

[6] Reprinted from ref.[24] with the permission of AIP Publishing.

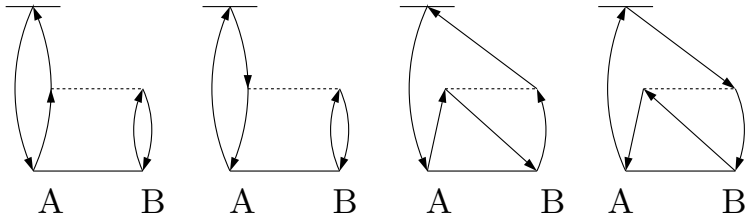

Figure 2.7: The diagrams of the *a posteriori* perturbative estimate of the fourth-order singles contribution to the weak pair correlation energy of the LCCD[S]-R^{-6} method. [7]

expressions are then,

$$\begin{aligned} E^{4s} &= 2t_i^r \left[\tilde{T}_{ik}^{st}(rs|kt) - S_{rr'} \tilde{T}_{kl}^{r's}(sl|ki) \right] \\ &= 2t_i^r \left[2K(T_{ik})^{rk} - K(T_{ki})^{rk} - S_{rr'} \tilde{T}_{kl}^{r's}(sl|ki) \right], \end{aligned} \tag{2.14}$$

with

$$K(T_{ij})^{rk} = T_{ij}^{st}(rs|kt) \tag{2.15}$$

Since this is a correction to the close pair energy and already it is of the fourth order, we restrict the doubles amplitudes to close pairs. The contraction of the close pair doubles with the three external integrals, namely the $K(T_{ij})^{rk}$ operators, given in eq.(31) in ref.[36], is already available in the existing LCCSD code. It is used for the feedback of the close pair doubles to the LCCSD singles residual, eq.(B27) of ref.[36]. Therefore the additional computational effort for the evaluation of E^{4s} is minuscule, as shown in sec.2.3.3. This non-iterative [S]-correction only represents the energy component originating from the singles contribution to the close pair doubles. The singles themselves are still fully iterated in the framework of the standard LCCSD treatment.

[7] Reprinted from ref.[24] with the permission of AIP Publishing.

2.3 Calculations

In this section the aforementioned methods were put to the test. In order to test the accuracy, the LCCSD|X and LCCSD(T)|X methods, where X stands for the different close/weak pair treatments from LMP2 until LCCD[S]-R^{-6}, were performed on a set of diverse inter-molecular complexes ranging from hydrogen bonded to van der Waals dominated systems. These complexes and their geometries are from the S22, S66 and JSCH-2005 benchmark data sets.[22;23;53] The computational cost of the different methodological schemes is compared in sec.2.3.3.
The basis aVTZ was used (cc-pVTZ on hydrogen atoms, aug-cc-pVTZ on all other atoms) for all calculations.[54;55]

2.3.1 Within the ring approximation

In this section the accuracy of methods within the ring approximation are discussed. For testing purposes in all the inter-molecular complex calculations the intra-monomer pairs were specified as strong pairs, and the inter-monomer pairs as close pairs (thus avoiding the close-weak pair distinction).
The test set consists of inter-molecular complexes of the benchmark set S66.[22] In tab.2.2 the interaction energies E_{int}, their overall correlation contribution E_c and the intra- and inter-molecular components E_{intra}/ E_{inter} for the tested dimers are given. Compared are the hybrid methods LCCSD|X and pure LMP2 with reference to the LCCSD calculation with all, intra- and inter-molecular, pairs taken as strong. In case of the hybrid methods, the following choices for the close pair treatment (X) were tested:

- LMP2*uc*, i.e. uncoupled (*uc*) LMP2, meaning close pair amplitudes do not contribute to the strong pair residual

- LMP2
- d-LrCCD*uc*
- d-LrCCD
- LrCCD3
- LrCCD

The analysis of the results reveals that for the hydrogen bonded systems the LMP2 interaction energies are matching the LCCSD reference quite well.
For the π-stacked systems on the other hand LMP2 exhibits a pronounced overestimation. By splitting the correlation contribution ΔE_c into its intra δE_{intra} and inter δE_{inter} parts, it can be seen that the overestimation comes significantly from an underestimation of the repulsive δE_{intra} part. In fact this is more significant for π-stacked systems like the pyridine dimer where it gives 0.57 kcal/mol vs. the reference LCCSD intra contribution of 2.39 kcal/mol. Another example is the benzene dimer with the LMP2 value of 0.29 kcal/mol against the reference 1.97 kcal/mol.
Moving to LCCSD with uncoupled LMP2 treatment for the close pairs, LCCSD|LMP2*uc*, this deficiency remains without noticeable improvement. Turning on the coupling, i.e. adding the feedback from the inter-molecular amplitudes onto the intra-molecular ones, however, effects the intra contribution immensely to the better. δE_{intra} is improved while δE_{inter} stays at the previous LMP2 level.
Looking again at the π-stacked pyridine dimer, LCCSD|LMP2's δE_{intra} compares now to the reference with a result of 2.86 kcal/mol, and the benzene dimer with a result of 2.40 kcal/mol. Though this is now an overestimation of 0.5 kcal/mol and 0.4 kcal/mol in the respective cases, this is favorably compensated partly by a too attractive δE_{inter}. Therefore, the LCCSD|LMP2 treatment is clearly favorable over LCCSD|LMP2*uc*, as well as over LMP2.

Table 2.2: The inter-molecular interaction energies and its components are compared on the level of LCCSD|X with X=LMP2 *uc* (for uncoupled), X=d-LrCCD *uc*, X=LMP2, X=*direct*-LrCCD, X=LrCCD3, X=LrCCD, to a full LCCSD calculation (all pairs strong) for a set of dimers. Correlation contributions ΔE_c to the respective inter-molecular interaction energies and their intra- and inter-molecular components δE_{intra}, and δE_{inter}, respectively, all in [kcal/mol]. The basis aVTZ is used. Also the RMS and maximum deviations relative to the CCSD calculation are given. The prefix "L" for local was omitted in the method names, for brevity. [8]

	CCSD	MP2	CCSD| MP2*uc*	CCSD| d-rCCD*uc*	CCSD| MP2	CCSD| d-rCCD	CCSD| rCCD3	CCSD| rCCD
				electrostatic dominated				
				peptide-peptide				
δE_{inter}	-5.663	-5.621	-5.621	-4.427	-5.620	-4.200	-5.793	-5.945
δE_{intra}	3.070	2.364	2.274	2.274	3.106	2.792	3.092	3.121
ΔE_c	-2.593	-3.257	-3.348	-2.153	-2.514	-1.408	-2.701	-2.824
E_{int}	-7.162	-7.825	-7.915	-6.721	-7.081	-5.976	-7.269	-7.392
				dispersion dominated				
				pyridin-pyridine (π-π)				
δE_{inter}	-7.707	-9.150	-9.150	-6.486	-9.148	-6.060	-8.346	-8.795
δE_{intra}	2.388	0.567	0.592	0.593	2.857	1.953	2.536	2.671
ΔE_c	-5.319	-8.583	-8.558	-5.893	-6.291	-4.108	-5.810	-6.124
E_{int}	-1.976	-5.241	-5.215	-2.551	-2.948	-0.765	-2.467	-2.781
				benzene-benzene(π-π)				
δE_{inter}	-7.035	-8.369	-8.369	-5.891	-8.368	-5.504	-7.672	-8.086
δE_{intra}	1.965	0.293	0.281	0.282	2.397	1.559	2.110	2.236
ΔE_c	-5.070	-8.076	-8.089	-5.609	-5.971	-3.946	-5.562	-5.850
E_{int}	-1.103	-4.110	-4.121	-1.643	-2.003	0.022	-1.594	-1.882
				guanine-cytosine (stacked)				
δE_{inter}	-18.237	-20.212	-20.212	-15.028	-20.212	-14.140	-18.727	-19.555
δE_{intra}	9.193	6.752	5.932	5.931	9.834	8.193	9.295	9.503
ΔE_c	-9.045	-13.460	-14.280	-9.098	-10.378	-5.947	-9.431	-10.053
E_{int}	-14.455	-18.875	-19.699	-14.514	-15.790	-11.360	-14.845	-15.465
				RMS deviation				
δE_{inter}		0.900	0.900	1.613	0.899	2.021	0.364	0.719
δE_{intra}		1.921	1.545	1.545	0.293	0.467	0.079	0.171
ΔE_c		2.591	2.380	0.389	0.613	1.558	0.285	0.549
E_{int}		2.194	2.381	0.389	0.613	1.557	0.286	0.550
				maximum deviation				
δE_{inter}		1.975	1.975	3.209	1.975	4.097	0.639	1.318
δE_{intra}		5.381	3.261	3.262	0.642	1.000	0.149	0.310
ΔE_c		5.397	5.235	0.825	1.333	3.097	0.492	1.008
E_{int}		4.420	5.244	0.825	1.335	3.095	0.491	1.010

Now we move to the next method: LCCSD|d-LrCCD.
Its main problem is a considerable underestimation of the inter correlation contribution δE_{inter}. Though the interaction energies resulting from LCCSD|d-LrCCD*uc* are in some cases not that far off, it is due to a fortunate error cancellation. Both contributions are too small: δE_{inter} not attractive and δE_{intra} not repulsive enough, which cancel one another in those cases. When going over to the coupled case the error compensation vanishes. This results in a severe underestimation of the interaction energy in all the investigated systems. For example in the stacked guanine-cytosine dimer, the underestimated inter contribution in the coupled case, -15.03 kcal/mol, remains nearly unchanged in the presence of coupling, -14.14 kcal/mol, vs. LCCSD reference with -18.24 kcal/mol. At the same time the intra contribution gets corrected from the uncoupled, 5.93 kcal/mol, to the coupled case with 8.20 kcal/mol, vs. LCCSD reference with 9.20 kcal/mol. The resulting interaction energy of the uncoupled case, -14.51 kcal/mol, gets corrected to -11.36 kcal/mol in the coupled case, thus leading to the described severe underestimation of the reference value of -14.455 kcal/mol.
Moving on from d-LrCCD by including the missing R^{-6} decaying diagram of the third order, D6, results in the LCCSD|LrCCD3 method. As the calculations indicate, taking account of diagram D6 is absolutely essential, as the results dramatically improve.
Comparing the interaction energies E_{int} of LCCSD|LrCCD3 to the LCCSD reference, the difference in almost all cases is not higher than 0.5 kcal/mol. Also in almost all cases the respective contributions of δE_{intra} and δE_{inter} are close to the reference values, with a difference of under 1 kcal/mol which points to less dependence on fortuitous error cancellation between the two parts than LCCSD|LMP2. This makes the LrCCD3 treatment obviously favored over all the other methods previously investigated here.
Now we consider the inclusion of the fourth order diagrams, that decay as R^{-6}, remaining within the ring approximation. The corresponding

[8]Reprinted from ref.[21] with the permission of AIP Publishing.

treatment, LCCSD|LrCCD, however worsens the results. LCCSD|LrCDD shows no improvement over LCCSD|LMP2. Both their RMS-deviations are about the same for every contribution. We note again, that LrCCD is constrained to the ring approximation and thus misses additional diagrams of the fourth order, decaying with R^{-6}. These shortages are addressed in the methods LCCD-R^{-6} and LCCD[S]-R^{-6} respectively, which are investigated in the next section.

Now we consider the *a posteriori* triples correction L(T). Its residual also includes strong and close pair doubles amplitudes, and thus is affected by the new treatment of the close pairs. Tab.2.3 compares the pure triples contribution ΔE_T as well as E_{int} of the reference LCCSD(T) against LCCSD(T)|LMP2 and LCCSD(T)|LrCCD3. The results for ΔE_T generated by LCCSD(T)|LrCCD3 are overall always in better agreement with the reference than LCCSD(T)|LMP2, and hence indicate a better approximation to the LCCSD close pair amplitudes. A minor exception are some of the electrostatic dominated systems, where the additional difference in the worst case, stacked guanine-cytosine, is nevertheless only 0.03 kcal/mol.

[9]Reprinted from ref.[21] with the permission of AIP Publishing.

Table 2.3: Inter-molecular interaction energies E_{int}, correlation contributions of the local triples correction, ΔE_{T} (all in [kcal/mol]). Also the RMS and maximum deviations relative to the CCSD(T) calculation are given. The prefix "L" for local was omitted in the method names, for brevity. [9]

	CCSD(T)	CCSD(T)/ MP2	CCSD(T)/ rCCD3	CCSD(T)	CCSD(T)/ MP2	CCSD(T)/ rCCD3	CCSD(T)	CCSD(T)/ MP2	CCSD(T)/ rCCD3
	dispersion dominated			electrostatic dominated			mixed		
	pyridin-pyridine (π-π)			water dimer			uracil-uracil (π-π)		
E_{int}	-3.065	-4.327	-3.706	-4.522	-4.390	-4.524	-8.252	-9.188	-8.636
ΔE_{T}	-1.089	-1.379	-1.239	-0.173	-0.152	-0.173	-1.568	-1.724	-1.659
	pyridin-pyridine (TS)			peptide-peptide			uracil-ethyne		
E_{int}	-2.857	-3.267	-3.149	-7.691	-7.599	-7.818	-3.167	-3.635	-3.484
ΔE_{T}	-0.603	-0.710	-0.674	-0.529	-0.518	-0.549	-0.627	-0.720	-0.693
	benzene-benzene(π-π)			uracil-$AcNH_2$			uracil-pyridine (π-π)		
E_{int}	-2.158	-3.323	-2.797	-17.731	-17.473	-17.831	-5.609	-6.810	-6.143
ΔE_{T}	-1.055	-1.320	-1.203	-0.618	-0.585	-0.639	-1.319	-1.567	-1.442
	benzene-benzene(TS)			pyridine-methanol			RMS deviation		
E_{int}	-2.220	-2.691	-2.566	-6.652	-6.621	-6.723		0.773	0.373
ΔE_{T}	-0.602	-0.722	-0.683	-0.426	-0.432	-0.442		0.161	0.087
	uracil-ethene			guanine -cytosine (Watson Crick)			maximum deviation		
E_{int}	-2.739	-3.108	-3.068	-28.550	-28.292	-28.773		1.663	0.651
ΔE_{T}	-0.683	-0.751	-0.752	-0.821	-0.795	-0.876		0.328	0.150
	uracil-neopentane								
E_{int}	-2.876	-3.029	-3.122						
ΔE_{T}	-0.771	-0.792	-0.818						
	guanine-cytosine (stacked)								
E_{int}	-16.401	-18.064	-16.925						
ΔE_{T}	-1.946	-2.274	-2.081						

Intra close pair treatment comparison

In this subsection we start distinguishing between close and weak pairs for the intra-molecular interactions to test the behaviour of the described techniques. This was done on the example of the linear poly-glycine peptide chain $(Gly)_4$, HO[C(O)CH2 NH]$_4$H. The LCCSD(T)|LrCCD3 method was compared to LCCSD(T)|LMP2. These results of these computations are shown in tab.2.4. The geometry is taken from the supplementary material of ref.[41].
The classification into strong, close and weak pairs is done via two distance criteria, R_c and R_w. The inter-orbital distance R classifies the pair.

- strong, $R < R_c$
  full CCSD(T) residual. The amplitudes in the strong pair residual are from strong and close pairs.

- close, $R_c \leq R < R_w$
  LrCCD3 and (T) residual. The amplitudes in the close pair residual are from strong, close and weak pairs.

- weak, $R_w \leq R$
  LrCCD3 residual. The amplitudes in the weak pair residual are from strong, close and weak pairs.

In tab.2.4 LCCSD correlation energies E_{CCSD} and L(T0) triples corrections E_{T} are compared for different values of R_c and R_w. The energies are obtained with LCCSD(T)|LMP2 (uncoupled and coupled) and LCCSD(T)|LrCCD3 (coupled). The table shows that the E_{CCSD} values of LCCSD(T)|LrCCD3 are much closer to the full LCCSD(T) reference than those of the related LCCSD(T)|LMP2 value. Looking at $(R_c, R_w) = (1, 3)$ bohr, the difference to the reference is -10 mH for LCCSD(T)|LMP2 and 2 mH for LCCSD(T)|LrCCD3. LCCSD(T)|LMP2's overestimation of E_{CCSD} and underestimation of E_{T} fortuitously cancel each other to a large extent, since for $R_w = 3$ bohr the local triples list, which includes strong and close pairs,

is still rather short.[34]

Setting R_w to 8 bohr, i.e. increasing the number of close pairs, brings both methods closer to the reference E_{CCSD}; -5 mH and -0.1 mH, respectively. With such a setting the underestimation of E_{T} is still at circa 1 mH. Evidently, with increasing number of close pairs E_{CCSD} converges quicker than E_{T}.

Consequently one may consider to decouple the construction of the orbital triples list from the criterion of coupling strong and close pairs in the amplitude equations. The average wall clock times per iteration is also shown in tab.2.4. (The calculations were executed by four Intel Xeon cores X5560@2.8 GHz.) This shows that the additional cost of LCCSD(T)|LrCCD3 treatment of close and weak pairs, compared to coupled LCCSD(T)|LMP2, is rather small. For R_c = 1 bohr it is only about 20% and it would be considerably less when shifting to more strong pairs.

This means that LCCSD(T)|LrCCD3 is a computationally economic method and can substitute LCCSD(T)|LMP2 without a great effort.

Table 2.4: CCSD correlation energies E_{CCSD} and corresponding triples corrections E_{T} (all in hartree) for the linear Gly_4 chain with different specifications for the pair approximation. LCCSD(T0)|X with X=LrCCD3, X=LMP2, and X=LMP2*uc* (for uncoupled) are compared to a full LCCSD(T0) calculation (all pairs strong). The number of strong, close, and weak pairs, as well as the number of orbital triples is also given. Furthermore, the average elapsed times per iteration T_{avg} (in seconds) is provided for each method. Convergence was always reached in nine iterations. [10]

R_c	R_w	number of strong|close|weak pairs	number of orbital triples	E_{T}	E_{CCSD}	T_{avg}
			LCCSD(T0)|MP2*uc*			
1	3	173 | 140 | 863	1197	-0.138429	-3.356354	211
1	8	173 | 439 | 564	3420	-0.148418	-3.356352	314
4	8	313 | 299 | 564	4458	-0.145835	-3.335908	590
8	18	612 | 387 | 177	13101	-0.145289	-3.330224	1652
9	28	636 | 540 | 0	18617	-0.145291	-3.330197	1840
			LCCSD(T0)|MP2			
1	3	173 | 140 | 863	1197	-0.135811	-3.340167	456
1	8	173 | 439 | 564	3420	-0.144770	-3.334735	624
4	8	313 | 299 | 564	4458	-0.144932	-3.330585	882
8	18	612 | 387 | 177	13101	-0.145274	-3.330125	2024
9	28	636 | 540 | 0	18617	-0.145281	-3.330130	2267
			LCCSD(T0)|rCCD3			
1	3	173 | 140 | 863	1197	-0.135279	-3.327975	552
1	8	173 | 439 | 564	3420	-0.144021	-3.330273	745
4	8	313 | 299 | 564	4458	-0.144921	-3.330367	980
8	18	612 | 387 | 177	13101	-0.145291	-3.330156	2149
9	28	636 | 540 | 0	18617	-0.145297	-3.330161	2310
			full LCCSD(T0)			
		1176 | 0 | 0	19552	-0.145292	-3.330191	4272

[10] Reprinted from ref.[21] with the permission of AIP Publishing.

2.3.2 Beyond the ring approximation

In this section the accuracy of methods beyond the ring approximation is discussed. In all the inter-molecular complex calculations he calculations were carried out analogous to the calculations in sec.2.3.1.
The test set consists of inter-molecular complexes of the benchmark sets S66, S22 and JSCH-2005.[22;23;53] The selected systems represent a mixture between large and small hydrogen-bonded, dispersion dominated, and mixed inter-molecular complexes. Compared are the hybrid methods LCCSD|X in reference to the LCCSD calculation with all, intra- and inter-molecular, pairs strong. In case of the hybrid methods, for the close pair treatment (X), now the following methods are taken,

- LMP2
- LrCCD3
- LrCCD

and the focus of this section,

- LCCD-R^{-6}
- LCCD[S]-R^{-6}

In fig.2.8 the comparison of these methods is shown on the level of deviation to the correlation contribution to the interaction energy ΔE_c from the full LCCSD reference. A similar plot is given in fig.2.9, but there with the addition of L(T) correction compared to the full LCCSD(T) as reference. The tables with the exact values used to make these pictures, as well as a table analogous to tab.2.2 for these methods, can be found in the supplementary data.
As pointed out at the end of the previous section LrCCD is constrained

to the ring approximation and thus does not include all fourth order diagrams with R^{-6} decay. Also already discussed in the previous section, LCCSD|LrCCD shows no improvement over LCCSD|LMP2. Now dropping the ring approximation and including the additional R^{-6} diagrams D13-D16 (fig.2.5), have a substantial effect. LCCSD|LCCD-R^{-6} makes the interaction energy less attractive, compared to LCCSD|LrCCD. Looking at some examples, ΔE_c becomes less attractive by 0.70 kcal/mol for the π-stacked benzene dimer, 0.75 kcal/mol for the π-stacked pyridine dimer, 1.02 kcal/mol for the guanine-cytosine dimer in the Watson-Crick configuration, and 1.50 kcal/mol for its stacked configuration. Compared to LCCSD|LrCCD the inclusion of the additional diagrams is a considerable improvement.

[11]Reprinted from ref.[24] with the permission of AIP Publishing.

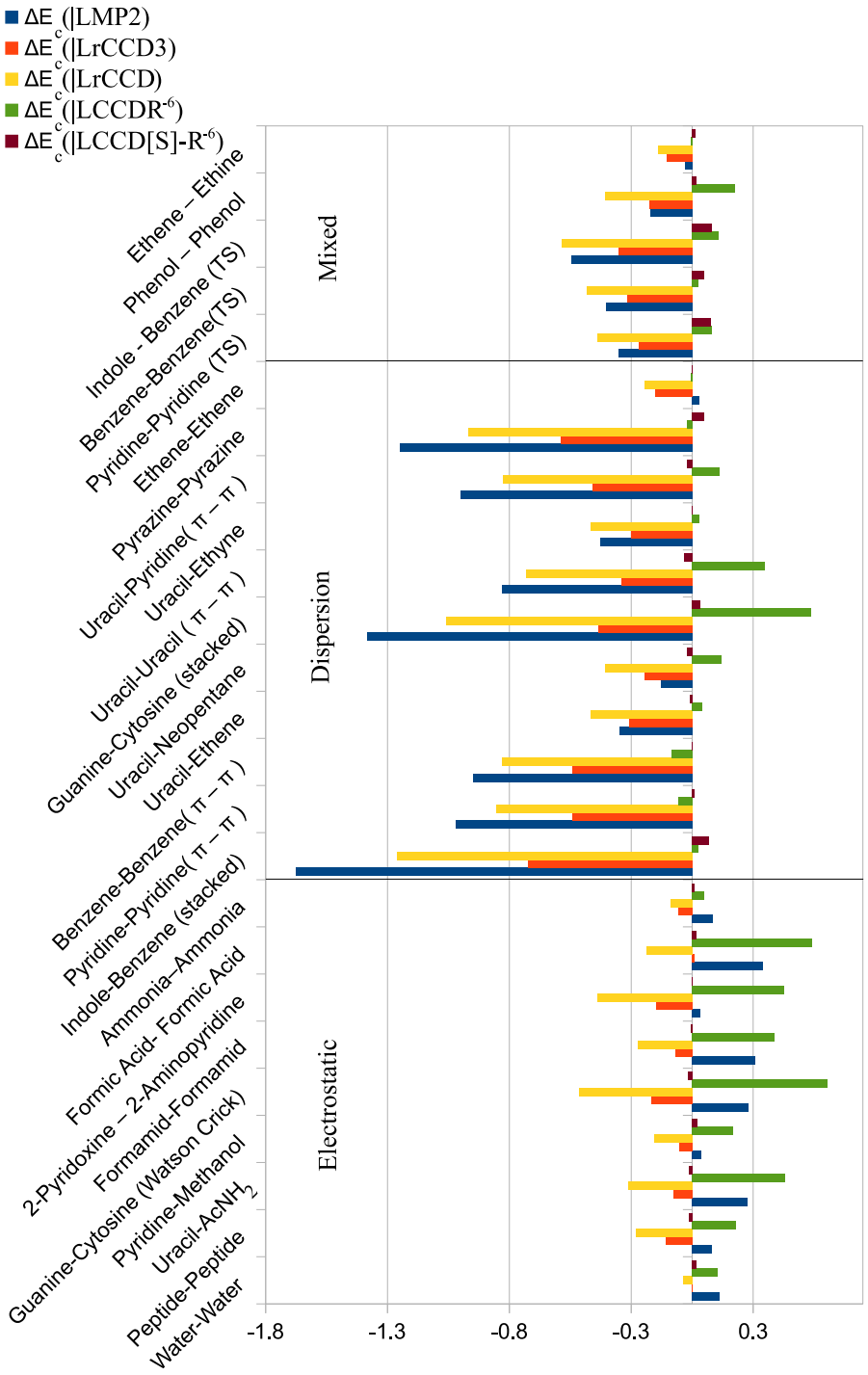

Figure 2.8: Deviations (in kcal/mol) of the correlation contributions to the interaction energies ΔE_c from the respective full LCCSD values, for LCCSD|LMP2 (black), LCCSD|LrCCD3 (blue), LCCSD|LrCCD (red), LCCSD|LCCD-R^{-6} (green), and LCCSD|LCCD[S]-R^{-6} (brown). [11]

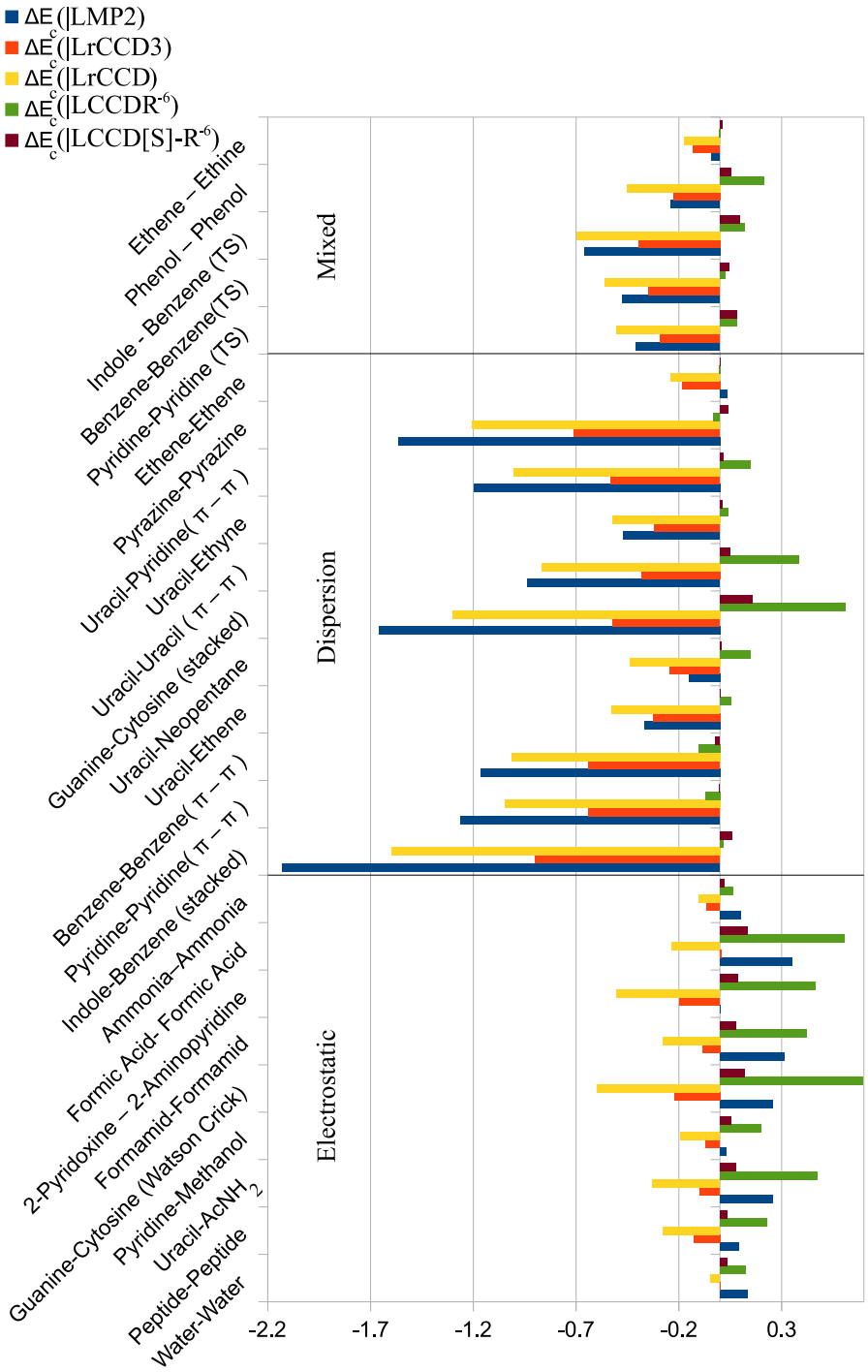

Figure 2.9: Deviations (in kcal/mol) of the correlation contributions to the interaction energies ΔE_c from the respective full LCCSD(T) values, for LCCSD(T)|LMP2 (black), LCCSD(T)|LrCCD3 (blue), LCCSD(T)|LrCCD (red), LCCSD(T)|LCCD-R^{-6} (green), and LCCSD(T)|LCCD[S]-R^{-6} (brown). [11]

Now comparing LCCSD|LCCD-R^{-6} to LCCSD|LMP2 and LCCSD|LrCCD3, it provides a better accuracy for van der Waals and mixed systems. For electrostatic dominated systems however, ΔE_c is regularly underestimated compared to the reference. In those cases, LCCSD|{LMP2, LrCCD3} perform better. Nevertheless LCCSD|LCCD-R^{-6} seems to provide better relative energies of different conformers, as seen in the example of the two guanine-cytosine complexes. The differences to the reference here are 1.57 kcal/mol for LCCSD|LMP2, 0.22 kcal/mol for LCCSD|LrCCD3 and 0.07 kcal/mol for LCCSD|LCCD-R^{-6}.

We move on by including the perturbative fourth order singles correction on top of LCCSD|LCCD-R^{-6}, which corresponds to the LCCSD|LCCD[S]-R^{-6} method. In doing so, the above described underestimation in electrostatic systems is significantly reduced. Interaction energies provided by LCCSD|LCCD[S]-R^{-6} conform outstandingly with the full LCCSD reference. The same is the case for LCCSD(T)|LCCD[S]-R^{-6} vs. the full LCCSD(T) reference. For the triples correction the deviations are in fact slightly larger, but this can be attributed to the fact that the doubles amplitudes entering the evaluation of the triples correction are those of LCCSD|LCCD-R^{-6}, which is lacking the singles feedback to the close pairs.

Table 2.5: RMS and maximum deviations in ΔE_c (total correlation contribution to interaction energy) and ΔE_T (triples component thereof) of the five different hybrid methods from the full (all pairs strong) LCCSD and LCCSD(T) references, respectively. All values in kcal/mol. [12]

	LCCSD				
	|LMP2	|LrCCD3	|LrCCD	|LCCD-R^{-6}	|LCCD[S]-R^{-6}
$\delta_{RMS}(\Delta E_c)$	0.637	0.293	0.550	0.240	0.033
$\delta_{MAX}(\Delta E_c)$	1.626	0.673	1.212	0.556	0.082
	LCCSD(T)				
	|LMP2	|LrCCD3	|LrCCD	|LCCD-R^{-6}	|LCCD[S]-R^{-6}
$\delta_{RMS}(\Delta E_c)$	0.816	0.384	0.712	0.299	0.067
$\delta_{MAX}(\Delta E_c)$	2.130	0.900	1.599	0.693	0.160
$\delta_{RMS}(\Delta E_T)$	0.181	0.092	0.162	0.059	0.059
$\delta_{MAX}(\Delta E_T)$	0.504	0.227	0.387	0.138	0.138

Taking a look at the RMS and maximum deviations of ΔE_c, compiled in tab.2.5, significant decreases are visible when going from LCCSD|LMP2 to LCCSD(T)|LrCCD3 and finally to LCCSD(T)|LCCD[S]-R^{-6}. RMS decreases from 0.64 to 0.29 and to 0.03 kcal/mol respectively and the maximum deviation from 1.63 to 0.67 and finally to 0.08 kcal/mol. When including the perturbative triples correction, RMS decreases from 0.82 to 0.38 and to 0.07; maximum deviation from 2.13 to 0.90 and to 0.16 kcal/mol. These numbers clearly show that LrCCD3 as already seen before, is significant improvement over LMP2, and LCCD[S]-R^{-6} an even further improvement.

As is discussed above, the fourth order singles correction to the close pair energy is essential. For hydrogen bonded systems, LCCSD|LCCD-R^{-6}'s performance is worse than that of LCCSD|LrCCD3. This indicates fortuitous error cancellation in LrCCD3 between the missing singles correction and the fourth order diagrams in LCCD-R^{-6}.

[12] Reprinted from ref.[24] with the permission of AIP Publishing.

Table 2.6: Comparison of pure pair theories LCCD|LrCCD3, and LCCD|LCCD-R^{-6} to full LCCD. Given are the correlation contribution to the interaction energy ΔE_c, and its intra- and inter-molecular component δE_{intra} and δE_{inter} [21;56;57] (all values in kcal/mol). [13]

	LCCD	LCCD|		LCCD	LCCD|	
		|LrCCD3	|LCCD-R^{-6}		LrCCD3	LCCD-R^{-6}
		Water-Water			Pyridine-Pyridine (π-π)	
δE_{inter}	-2.228	-2.322	-2.204	-7.772	-8.350	-7.800
δE_{intra}	1.519	1.529	1.515	2.263	2.373	2.269
ΔE_c	-0.709	-0.793	-0.688	-5.509	-5.976	-5.530
		Uracil-AcNH_2			Guanine-Cytosine (stacked)	
δE_{inter}	-8.762	-9.267	-8.741	-17.821	-18.802	-17.765
δE_{intra}	6.927	6.996	6.924	8.954	9.129	8.967
ΔE_c	-1.835	-2.271	-1.817	-8.867	-9.673	-8.798
	Guanine-Cytosine (Watson Crick)					
δE_{inter}	-13.319	-14.118	-13.285			
δE_{intra}	10.086	10.208	10.088			
ΔE_c	-3.232	-3.910	-3.197			

This can be confirmed by comparison to the LCCD method, which does not have singles at all. LCCD|X calculations were performed for a subset of the above systems against the full LCCD reference and the results are compiled in tab.2.6.

The investigated hybrid methods are LCCD|LrCCD3 and LCCD|LCCD-R^{-6}. In fact, LCCD|LCCD-R^{-6} conforms excellently with the LCCD reference. The maximum deviation is of 0.07 kcal/mol. LCCD|LrCCD3 however has clearly larger deviations; with a maximum deviation of 0.8 kcal/mol. The origin for those deviations is primarily from the inter-molecular contribution.

[13] Reprinted from ref.[24] with the permission of AIP Publishing.

2.3.3 Computational cost

This section is focused on the comparison of computational cost between the different methods for weak pair treatment. For that, calculations for one of the bigger complexes of the test set, namely the T-shaped indole-benzene, were performed. The details of these calculations, as well as the used hardware specifications, are shown in tab.2.7. In contrast to the calculations of section.2.3, intra-monomer pairs were divided intro strong, close and weak pairs on the basis of inter-orbital distance, as it is usually done in routine calculations. The results for the case, where all intra-monomer pairs were set to strong is also provided in the supplementary data (A). Inter-molecular pairs are all close, in both cases.
It can be seen that the average time for an iteration increases by 16% when going from LMP2 weak pair treatment to LrCCD3, and by another 17% going from LrCCD3 over to LCCD[S]-R^{-6}. The amount of time for the computation of the non-ring CCD doubles diagrams (D13-16), as well as the *a posteriori* fourth order singles contribution [S] is practically negligible. Hence LCCSD|LCCD[S]-R^{-6} virtually has the same cost as a LCCSD|LrCCD calculation. As can also be seen, that moving from LMP2 weak pair treatment to LCCD[S]-R^{-6}, the time for integrals transformation and the required disk space for the integrals basically stays the same and introduces no additional cost.

[14]Reprinted from ref.[24] with the permission of AIP Publishing.

Table 2.7: Computational details for the calculations on the T-shaped indole-benzene complex (total number of pairs: 703, number of inter-molecular pairs: 330). In the hybrid calculations the intra-molecular pairs may be strong, close, or weak (number of strong/close/weak pairs: s=181, c=413, w=109). Timings for the transformation of the 3-external and 4-external integrals, for evaluating the individual groups of diagrams, and the time for converging the (hybrid) coupled cluster calculations are given in seconds. The file sizes of the 3-external and 4-external integral distributions are given in GBytes. The interaction energies E_{int} and their inter- and intra-molecular components δE_{inter}, δE_{intra} are also provided (in kcal/mol). The calculations were performed on seven AMD opteron 6180 SE cores. [14]

					LCCSD|		
part	diagrams[a]	full	LMP2	LrCCD3	LrCCD	LCCD-R^{-6}	LCCD[S]-R^{-6}
LMP2			92	92	75	76	76
LrCCD3	D1,D3,D6			89	86	83	83
LrCCD	D2,D4,D5,D7-9				347	348	348
LCCD-R^{-6}	D13-16					4	4
[S]	D18-21						0.3
sum parts			92	181	508	511	511
whole iteration		3404	1382	1603	1908	1875	1875
nr. of iterations		10	9	9	9	10	10
integral transformation (3-ext)		5265	3830	3809	3968	3772	3772
integral transformation (4-ext)		9281	3501	3382	3419	3384	3384
3-ext file size		90.646	64.271	64.271	64.271	64.271	64.271
4-ext file size		263.293	61.860	61.860	61.860	61.860	61.860
δE_{inter}		-7.440	-8.223	-7.837	-8.160	-7.372	-7.403
δE_{intra}		3.236	3.504	3.314	3.393	3.204	3.204
E_{int}		-3.909	-4.431	-4.234	-4.478	-3.871	-3.902

[a] additional diagrams beyond LMP2 according to figs.2.1,2.2,2.3,2.5,2.6.

2.4 Conclusion

The local correlation techniques overcomes the high scaling of the computational cost in the conventional coupled cluster methods. The local approximations allow for a substantial reduction of the amplitudes, which serves as the essential parameters to define the wave function, and are the actual unknowns to be determined within the coupled cluster calculations. Unfortunately the computationally efficient local pair approximations have so far been not very successful for providing desirable accuracy for interaction energies in inter-molecular complexes.
In this work a new concept for the local pair approximation has been developed. It truncates the pair-list of localized occupied orbitals, exploiting

- the R^{-6} decay of the pair energies with respect to the inter-orbital distance R between the two LMOs of the pair, and
- the order of the respective energy contribution within the Møller Plesset partitioning.

Pairs are subdivided into strong, close, weak, distant and very distant pairs based on this inter-orbital distance. The full LCCSD treatment is only given to the strong pairs. The distant pairs are treated at the LMP2 level, while the very distant pairs are disregarded. This work focuses on the close and weak pairs which usually provide the major contribution to the interaction energy in the inter-molecular systems.
A hierarchy of hybrid methods that are denoted as LCCSD|X, where X stands for the respective level of close or weak pair treatment was developed and tested.
We started from the conventional LCCSD|LMP2*uc* treatment, where the "*uc*" stands for uncoupled, i.e. close pair amplitudes are not allowed to contribute to the LCCSD strong pair residuum. The progressively improved models included for X:

1. LMP2
   Coupling to strong from close pairs is essential.

2. LrCCD3
   Inclusion of all third order ring diagrams decaying as R^{-6}. Provides very good accuracy, while being only slightly more expensive than LMP2. This method provides a promising model for treating the weak pairs.

3. LrCCD
   Inclusion of all diagrams up to fourth order decaying as R^{-6} within the ring approximation. Provides similar accuracy as LCCSD|LMP2, but at a higher cost.

4. LCCD-R^{-6}
   Addition of all diagrams decaying as R^{-6} up to the fourth order, not only the ring ones, allows for further improvement of accuracy.

5. LCCD[S]-R^{-6}
   Adding the fourth order singles terms, the error of the approximate local CC treatment reduces to minute values. The perturbative contribution to the close pair energy is very computationally efficient. This technique appears to be an excellent treatment for the close pairs.

To summarize, as is shown in this chapter, it is highly advisable from the point of accuracy and efficiency to extend the conventional local correlation treatment to that developed in the present work, namely to the LCCSD(T)|LCCD[S]-R^{-6} method.

Chapter 3

Embedded local direct ring CCD for periodic systems

The content of this chapter is already published in the Journal of Chemical Theory and Computation, ref.[25]. The manuscript was revised for the context of this thesis. Fig.3.1 was done by Prof. Dr. Martin Schütz. The finite-cluster calculations, e.g. in tab.3.1, were done by Dr. Denis Usvyat. The implementation and testing of the presented embedded fragment-based method was realized by the author.

3.1 Introduction

In this chapter we switch from molecules to crystals. Periodic systems due to their infinite nature are very challenging for accurate quantum chemical models. The techniques applicable to periodic systems can be divided into two categories,

- pure periodic approaches[6;58–66]
- fragmentation approaches[67–79]

The essential advantage of the first ones is that the computation actually treats the actual system of interest. Yet this comes with a high complexity and cost. There are not many implementations of pure periodic methods that reach MP2 level,[59;61;63;64;80] and even less that go beyond.[62;81] In the second category - fragment approach - the total energy or its part is computed for a finite cluster, or a series of clusters, which mimic the crystal under study. Since only a fragment of the actual periodic crystal is treated, it allows for the exploitation of molecular quantum chemical methods and programs. And the advantage of the fragmentation methods is that the cluster or fragment size can be adjusted to cost of the calculation and treatment. Smaller cluster sizes can be applied to investigate the trends of high-order incremental energy contributions, defects and local phenomena in solids. Big clusters can be used for large energy contributions, corresponding to a lower level of accuracy.
Fragmentation approaches however are not entirely unproblematic either. They are based upon the hypothesis that certain properties of the crystal can be represented via a finite cluster. The accuracy of this assumption, however, strongly depends on the quantity of interest, the kind of interactions and the embedding used for the cluster. A difficult quantity to describe with a finite cluster approach is, e.g., the total lattice energy per unit cell, because of the long-range electrostatic contributions. At the same time, the error introduced by a finite cluster representation for the correlation energy, and even more so for the post-MP2 energy correction, is much lower because the corrections are small and the interactions are of a shorter range.[67;69;71;73;74] A very high accuracy still, even for such unproblematic quantities, can only be achieved with a proper embedding scheme. Embedding is presently quite an intensive field of research.[74–76;82–94]
In this chapter we present an interface for a local direct ring-coupled cluster doubles (d-LrCCD) treatment of a fragment, embedded in the periodic Hartree-Fock solution, to compute the correlation energy or the post-MP2 incremental energy correction corresponding to this fragment. The

CRYSTAL[95] and CRYSCOR[61] programs carry out the periodic part of the computation. This includes Hartree-Fock calculation, Wannier function generation and local MP2 calculation. The interface is specified by generating the necessary information for the fragment calculation and providing the files for it. Finally for the finite cluster part the local coupled cluster code[21;36] of MOLPRO[96] program package is utilized. The interface presented as is marks the first step in a project aiming at a fragment-based local CCSD(T) treatment, as e.g. seen in the first chapter, embedded in the periodic local MP2. Presently the approach uses the d-rCCD residual described in sec.2.2.2.[18] d-LrCCD based on the HF reference is not guaranteed to be more accurate than LMP2. In fact MP2 shows a better accuracy for intermediately polarizable systems like hydrogen bonded crystals. But for systems with a small band-gap MP2 is known to deteriorate while d-rCCD shows more stability. This makes d-rCCD still a useful model in its own right. Based on this, further improvements on the fragment's level of coupled cluster treatment can be done beyond d-rCCD. The present state can thus also be seen as proof of principle of such an approach.
In this chapter the proposed fragment approach is shown and tested. In section 3.2 the theory of the fragment approach is outlined; first the fragment is defined and then the technical implementation is described. Then this method is used to investigated physisorption of H_2 and argon on graphane in section 3.3. Section 3.4 summarizes and concludes this chapter.

3.2 Theory

This section first defines the fragment of this approach and then describes the technical implementation. In the following index notations are i, j, k, l for Wannier functions (WFs) and r, s, t, u for localized virtual orbitals, i.e. projected atomic orbitals (PAOs). In the periodic case these indices represent the orbitals in the reference cell. In case an orbital is placed outside the

reference cell the latter is denoted by an additional calligraphic index. Note that direct space orbitals can be chosen real due to time reversal symmetry.

3.2.1 Fragment definition

In order to cut out a cluster of the periodic case, spatial fragmentation of both occupied and virtual space needs to be allowed. The description of the systems in the periodic and the finite cluster treatments is based on spatially local orbitals, where this is possible. Wannier functions (WFs) span the occupied space and are obtained from a periodic Hartree-Fock calculation.[97–100] A fragment is defined as a subset of WFs and their WF-pairs. Projected atomic orbitals (PAOs) span the pair specific virtual spaces (pair domains). The fragmentation here is defined as the PAOs belonging to the relevant atoms in the spatial vicinity of the WF-pairs. The fragments specified by WF and WF pair subsets automatically inherit these pair domains. This setup links the cluster to the periodic calculation. Fragments embedded in this periodic mean field solution can in principle be treated by any higher-level correlation method.
As already pointed out, the WFs and PAOs are obtained from a periodic calculation. The WFs are constructed by localization of the occupied Bloch orbitals.[98] Projection of the reciprocal AO images from the periodic occupied space and Fourier-transformation of the resulting orbitals in the direct space gives the PAOs.[101]

The pair-list and the correlation energy

When restricting the pair-list for the fragment, a special care is needed for the proper definition of the local correlation energy. Correlation energy

per cell in periodic local treatment is defined as[58]:

$$
\begin{aligned}
{}^{\text{cell}}E^{\text{periodic}}_{\text{corr.}} &= \sum_{i\,j\mathcal{J}} \sum_{r\mathcal{R}\,s\mathcal{S}\in[i\,j\mathcal{J}]} \tilde{T}^{i\,j\mathcal{J}}_{r\mathcal{R}\,s\mathcal{S}}\,(i\,r\mathcal{R}|j\mathcal{J}\,s\mathcal{S}) \\
&= \frac{1}{2}\Bigg(\sum_{i\,j\mathcal{J}} \sum_{r\mathcal{R}\,s\mathcal{S}\in[i\,j\mathcal{J}]} \tilde{T}^{i\,j\mathcal{J}}_{r\mathcal{R}\,s\mathcal{R}}\,(i\,r\mathcal{R}|j\mathcal{J}\,s\mathcal{S}) \\
&\quad + \sum_{i\mathcal{I}\,j} \sum_{r\mathcal{R}\,s\mathcal{S}\in[i\mathcal{I}\,j]} \tilde{T}^{i\mathcal{I}\,j}_{r\mathcal{R}\,s\mathcal{S}}\,(i\mathcal{I}\,r\mathcal{R}|j\,s\mathcal{S})\Bigg),
\end{aligned}
\tag{3.1}
$$

where the $(i\,r\mathcal{R}|j\mathcal{J}\,s\mathcal{S})$ are 4-index electron repulsion integrals (ERIs) in the chemical notation,

$$
\begin{aligned}
(i\,r\mathcal{R}|j\mathcal{J}\,s\mathcal{S}) = \int d\mathbf{r}_1 \int d\mathbf{r}_2 \psi_i(\mathbf{r}_1)\psi_a(\mathbf{r}_1 - \mathbf{R}_{\mathcal{R}}) \\
\frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|}\psi_j(\mathbf{r}_2 - \mathbf{R}_{\mathcal{J}})\psi_b(\mathbf{r}_2 - \mathbf{R}_{\mathcal{S}}),
\end{aligned}
\tag{3.2}
$$

where the $\tilde{T}^{i\,j\mathcal{J}}_{r\mathcal{R}\,s\mathcal{S}}$ are the double amplitudes in contra-variant form,

$$
\tilde{T}^{i\,j\mathcal{J}}_{r\mathcal{R}\,s\mathcal{S}} = 2\tilde{T}^{i\,j\mathcal{J}}_{r\mathcal{R}\,s\mathcal{S}} - \tilde{T}^{i\,j\mathcal{J}}_{s\mathcal{S}\,r\mathcal{R}'}
\tag{3.3}
$$

and $[i\,j\mathcal{J}]$ denotes the pair domain of the WF pair $i\,j\mathcal{J}$. Note that the summation range for one of the indices in (3.1) is always restricted to the reference cell. Formally the summation range of the $\mathcal{J}$ runs to infinity. In practice the expression for the periodic local correlation energy (i.e. the range of $\mathcal{J}$) is truncated on the basis of the pair energies' R^{-6} decay with the inter-orbital distance.

In the residual equations the inter-pair couplings include also pairs which do not exist in the energy expression (eq.3.1). These are pairs where all indices live outside the reference cell. The internal Fock-coupling in the periodic LMP2 residual $R^{r\mathcal{R}\,s\mathcal{S}}_{i\,j\mathcal{J}}$ involves , e.g., the term, (assuming orthogonality of the virtuals)

$$
R^{r\mathcal{R}\,s\mathcal{S}}_{i\,j\mathcal{J}} \Leftarrow -\sum_{k\mathcal{K}} f_{i\,k\mathcal{K}} T^{k\mathcal{K}\,j\mathcal{J}}_{s\mathcal{S}\,r\mathcal{R}},
\tag{3.4}
$$

Where $f_{ik\mathcal{K}}$ is an element of the fock matrix $\mathbf{f}$. The summation over $\mathcal{K}$ also formally goes to infinity here. The long range contributions however are cut-off by the exponential decay of the Fock matrix. Eq.(3.4) demonstrates that some amplitudes with both occupied indices outside the reference cell still need to be incorporated. In the periodic case this is not an essential problem. These amplitudes can be trivially obtained by translating the amplitudes with the first index in the reference cell (by virtue of translational symmetry):

$$T^{i\mathcal{I}\,j\mathcal{J}}_{s\mathcal{S}\,r\mathcal{R}} = T^{i,j(\mathcal{J}\ominus\mathcal{I})}_{s(\mathcal{S}\ominus\mathcal{I})\,r(\mathcal{R}\ominus\mathcal{I})} \tag{3.5}$$

The symbolic operation over the cell indices $\ominus$ represents here the actual operation applied to the corresponding translational vectors.
The pair-list which will be used in the finite cluster computation can now be determined. For the correlation energy per cell the first index of the pair should run over the WFs of the reference cell only, as is given in eq.(3.1). We will denote these pairs as r.c.-pairs. The fragment pair-list should also include the pairs of the residual equations with both indices outside the reference cell (see eq.(3.4)). Such pairs will be denoted as emb.-pairs. The pair-list of the fragment $\{ij\}^{\text{fragment}}$ then is the union of these two sets of pairs.
If we evaluate the finite cluster's local correlation energy using the standard molecular expression

$$E^{\text{fragment}}_{\text{corr.}} = \sum_{i\geq j\in\{ij\}^{\text{fragment}}} \left(2-\delta_{ij}\right) \sum_{rs\in[ij]} \tilde{T}^{ij}_{rs}(ir|js), \tag{3.6}$$

with the pair-list $\{ij\}^{\text{fragment}}$, it would not correspond to the energy per cell, due to the presence of emb.-pairs (compare to eq.(3.1)). It rather has to be

calculated as

$$
\begin{aligned}
{}^{\text{cell}}E_{\text{corr.}}^{\text{fragment}} = & \sum_{\substack{i \geq j \in \{ij\}^{\text{fragment}} \\ i,j \in \text{r.c.}}} \left(2 - \delta_{ij}\right) \sum_{rs \in [ij]} \tilde{T}_{rs}^{ij}(is|js) \\
& + \sum_{\substack{i > j \in \{ij\}^{\text{fragment}} \\ i \in \text{r.c.} \\ j \notin \text{r.c.}}} \sum_{rs \in [ij]} \tilde{T}_{rs}^{ij}(ir|js) \\
& + \sum_{\substack{i > j \in \{ij\}^{\text{fragment}} \\ i \notin \text{r.c.} \\ j \in \text{r.c.}}} \sum_{rs \in [ij]} \tilde{T}_{rs}^{ij}(ir|js). \qquad (3.7)
\end{aligned}
$$

For ${}^{\text{cell}}E_{\text{corr.}}^{\text{fragment}}$ to be evaluated according to eq.(3.7) the program dealing with the finite cluster requires the information whether a certain WF belongs to the reference cell [denoted as "r.c."] or not. This information, together with the pair-list $\{ij\}^{\text{fragment}}$, and the pair domains $[ij]$, has to be passed from the periodic program via the interface to the finite cluster program.

The tighter the fragment cutting threshold, the longer the pair-list $\{ij\}^{\text{fragment}}$ and the closer the fragment results approach the periodic ones. However the finite cluster computation gets more expensive with longer pair-lists since the benefit of translational symmetry can no longer be used. For these reasons the inter-orbital distance criterion R^{fragment} specifying $\{ij\}^{\text{fragment}}$ has to be chosen smaller than in an ordinary periodic LMP2 calculation.

The r.c.-pairs are obtained according to the inter-orbital cut-off distance R^{fragment} mentioned above. The emb.-pairs are obtained by meeting two conditions:

1. only WFs appearing in th r.c.-pairlist are allowed
2. the emb.-pairs formed by these WFs also have to meet the cut-off tolerance $R^{\text{fragement}}$

The second is equivalent to the condition that any of the emb.-pairs should match some level r.c.-pair by a translation. For additional reduction of

computational cost, an option to restrict the emb.-pairlist to diagonal pairs only can be activated. As seen later in sec.3.3 such an approximation does not affect the interaction energy significantly. The pair hierarchy approach of local coupled cluster methods is used toward decreasing the fragment's pair-list truncation error.[36] As seen in chapter 2 only a restricted set of pairs, strong pairs, are treated at the CC level and the remaining weak pairs are treated at a lower level, such as LMP2. In the present context this translates to the periodic LMP2 treatment of the pairs outside of $\{ij\}^{\text{fragment}}$. This is done via the incremental energy correction,

$$^{\text{cell}}E_{\text{corr.}} = {}^{\text{cell}}E^{\text{fragment}}_{\text{d-LrCCD}} + {}^{\text{cell}}E^{\text{periodic}}_{\text{LMP2}} - {}^{\text{cell}}E^{\text{fragment}}_{\text{LMP2}}. \tag{3.8}$$

In the equation above, the third term is obtained by summing up the respective periodic pair energies of the fragment. The first term $^{\text{cell}}E^{\text{fragment}}_{\text{d-LrCCD}}$, however, is processed within the embedded finite-cluster framework. That means when undergoing a consistency test like the one in tab.3.1, i.e. reducing the treatment level of this term to LMP2 the first and third terms do not exactly cancel. This comes from the couplings present in the third term, but not in the first. In the LMP2 residual equations, see eq.(3.4), the couplings of the inter-pairs are rather short ranged due to the exponential decay (in non-conducting systems) of the off-diagonal Fock matrix elements. Consequently, regardless of the fragment specifications, the results of the trivial "embedded-fragment LMP2|periodic LMP2" hybrid scheme, analogously defined to eq.(3.8), should closely match that of the periodic LMP2.

Energy partitioning in calculations of physisorption

Modeling of physisorption of molecules on surfaces is an important application for this fragment technique. In periodic LMP2 applications to physisorption it is convenient to partition the local correlation interaction energy into the

intra-slab

$$\Delta E_{\text{intra-slab}} = \sum_{\substack{i\,j\mathcal{J} \\ i,j\in\text{slab}}} \Big(E_{i\,j\mathcal{J}}(\text{all}) - E_{i\,j\mathcal{J}}(\text{slab}) \Big), \tag{3.9}$$

intra-adsorbate

$$\Delta E_{\text{intra-ads.}} = \sum_{\substack{i\,j\mathcal{J} \\ i,j\in\text{adsorbate}}} \Big(E_{i\,j\mathcal{J}}(\text{all}) - E_{i\,j\mathcal{J}}(\text{ads.}) \Big), \tag{3.10}$$

and inter-slab-adsorbate

$$\Delta E_{\text{inter}} = \sum_{\substack{i\,j\mathcal{J} \\ i\in\text{slab},j\in\text{adsorbate} \\ j\in\text{slab},i\in\text{adsorbate}}} E_{i\,j\mathcal{J}}(\text{all}) \tag{3.11}$$

components.[47;61;73;102] Here, $E_{i\,j\mathcal{J}}$ stands for the individual pair energies

$$E_{i\,j\mathcal{J}}(\text{all, slab or ads.}) = \sum_{r\mathcal{R}\,s\mathcal{S}\in[i\,j\mathcal{J}]} \tilde{T}^{i\,j\mathcal{J}}_{r\mathcal{R}\,s\mathcal{S}}\,(i\,r\mathcal{R}|j\mathcal{J}\,s\mathcal{S}), \tag{3.12}$$

The labels "(all)", "(slab)", and "(adsorbate)" indicate the whole system, the isolated slab, and the adsorbate, respectively. There are two benefits in partitioning of the energy in such a manner.

- A better insight in the physics of the interaction and
- possible reduction of the cost of the computation.

The latter can be achieved trough a more drastic truncation of the intra-slab and intra-adsorbate pair-list in comparison to the inter-slab-adsorbate pair-list.[61;71] Within the finite cluster interface, this can be done by using different inter-orbital cut-off tolerances R^{fragment} for the specification of the inter and intra pair-lists. In order to model isolated adsorbate molecules, the definition of the fragment's adsorbate WFs includes those of the reference cell only. That requires an additional factor of 2 in the second and

Figure 3.1: Schematic representation of an adsorbate-slab system and the inter-slab-adsorbate pairs occurring in a typical local correlation calculation. The adsorbate and slab WFs are displayed by smaller blue and larger red circles, respectively, and are denoted by combined intra-cell $(1, 2)$ and cell-counting $(\bar{2}, \bar{1}, \ldots, 2)$ indices. The latter are given by the calligraphic font and decorated with an over-bar if the cell corresponds to a "negative" translation. Panel (a) presents the periodic case, while (b) and (c) fragments cut from the periodic systems. Panel (c) shows a fragment containing the adsorbate WF from the reference cell only. The inter-slab-adsorbate WF pairs, restricted to up to the second nearest neighbors, are represented by arrows: the square denotes the first WF index of a pair and the arrowhead the second. In Panel (b) different pairs (arrows) are distinguished: black solid arrows refer to pairs with *both* WFs inside the r.c., blue dashed arrows to pairs with just one WF inside the r.c., and green dotted arrows to pairs with no WF inside the r.c. The latter are emb.-pairs and do not contribute to the correlation energy directly via eq.(3.7). [15]

third terms of the energy expression (3.7) of the finite cluster for the inter-slab-adsorbate pairs.

Different factors for the pair energies are necessary in the cluster model. Fig.3.1 illustrates this statement:
Panel (a) shows a full periodic system. The WFs of the slab and adsorbate are represented by large red and small blue spheres, respectively. There are two WFs per cell; one adsorbate WF, named WF1, one slab WF, named WF2. The cell indices are given in a calligraphic font. The "negative" cells are marked with an over-bar. The arrows represent WF pairs that contribute to the correlation energy per cell. The arrow's square marks the first WF and the arrowhead the second WF in that pair. The inter-slab-adsorbate pair-list in Fig.3.1(a) includes pairs connecting up to the second-nearest neighboring WFs. The arrows in Fig.3.1(a) correspond to the first variant of the expression for the correlation energy expression per cell in eq.(3.1). There the first index is always in the reference cell. The pair energies are invariant with respect to pair index permutation:

$$E_{i\,j\mathcal{J}} = E_{j\mathcal{J}\,i}.$$

Eq.(3.1) can be symmetrized, as is shown in the second equality of this equation. The number of terms can be reduced by nearly a factor of 2 by combining the translational and permutational symmetries,

$$E_{i\,j\mathcal{J}} = E_{j\,i\ominus\mathcal{J}}. \tag{3.13}$$

Usually periodic LMP2 calculations utilize this, but for simplicity we have omitted it from eq.(3.1) and fig.3.1(a). For this illustrative example of Fig.3.1, in addition to the standard permutational invariance

$$E_{1\mathcal{0},2\mathcal{0}} = E_{2\mathcal{0},1\mathcal{0}},$$

[15]Reprinted with permission from ref.[25] and Prof. Schütz. ©2016 ACS.

eq.(3.13) yields also the relations

$$E_{1\mathit{0},2\mathit{1}} = E_{2\mathit{0},1\bar{\mathit{1}}},$$
$$E_{2\mathit{0},1\mathit{1}} = E_{1\mathit{0},2\bar{\mathit{1}}}.$$

Panel (b) of Fig.3.1 shows a fragment cut out of the periodic system of panel (a). The fragment includes the inter-slab-adsorbate WF pairs up to the second-nearest-neighbor. This fragment comprises the WFs from three adjacent unit cells. The inter-slab-adsorbate pairs of this fragment are again represented as (colored) arrows, with the square representing the first, and the arrowhead the second WF of the pair. Even though cells no longer exist in the fragment the cell indices of the WFs are kept, for easier comparison with the periodic case of panel (a). The total correlation energy of this fragment clearly does not correspond to the correlation energy per cell in the periodic case. However it is straightforward to link the periodic and molecular pairs and define the related "energy per cell" in the fragment calculation. Essentially we can distinguish the pairs according to their number of WFs inside the reference cell:

1. pairs with both WFs inside the r.c. (solid black arrows in 3.1(b))
2. pairs with only one WF inside the r.c. (dashed blue arrows in 3.1(b))
3. pairs with no WF inside the r.c. (dotted green arrows in 3.1(b))

Firstly the pairs with both WFs inside the r.c. appear with the same factor in the periodic- and the fragment- based energy expression (eqs. (3.1) and (3.7)). Next there are twice as many pairs with only one WF inside the r.c. in the fragment, than in the periodic structure. Hence, when aiming for energy per cell, the corresponding pair energy has to be scaled by a factor of $\frac{1}{2}$. Pairs with no WF inside the r.c. do not contribute directly to the correlation energy per cell and should be omitted in the energy expression (eq.3.7). Their influence solely originates from the couplings in the residual equations (see sec.3.2.1). The factors mentioned above were employed in

the derivation of eq.(3.7). Yet in eq.(3.7) there is an additional factor of 2 for all non-diagonal pairs that come from the index-permutation symmetry, i.e. the triangularity of the pair-list, which is always employed in fragment calculations.
Panel (c) considers the case where the fragment is chosen to contain the adsorbate WFs only from the reference cell. This represents a situation where a single adsorbate molecule rather than an adsorbate mono-layer is adsorbed on a surface. Here, all pairs involving adsorbate WFs outside the r.c. are absent, including all emb.-pairs. Contrary to the case shown in panel (b) the number of pairs appearing in such a fragment exactly matches that of the periodic energy-per-cell expression. This can easily be verified by comparing panel (c) and (a) and using translational symmetry in the latter,

$$E_{2\mathit{0},1\mathit{1}} = E_{2\bar{\mathit{1}},1\mathit{0}},$$
$$E_{2\mathit{0},1\bar{\mathit{1}}} = E_{2\mathit{1},1\mathit{0}}.$$

Consequently in this case, the energies of pairs with one WF outside the r.c., should not be scaled by a factor of $\frac{1}{2}$ and the standard molecular energy expression, eq.(3.6), can be used without any additional factors. In case of the intra-slab correlation energy, however, this simplification does not hold and eq.(3.7) should be used instead.

3.2.2 Direct-LrCCD residuum

The theory of d-LrCCD has been already discussed in sec.2.2.2. Here we reiterate essential parts of it relevant for the present context. The diagrams included in the residual are D1-5, which are given in fig.2.1. All these diagrams do not require ERI classes, which are not available in the CRYSCOR code base at the moment.

The spin-free d-LrCCD residual equations in the WF-PAO basis read:

$$\begin{aligned} R^{ij}_{rs} = 0 = (ir|js) &+ S_{rr'}T^{ij}_{r't}f_{ts} + f_{rt}T^{ij}_{ts'}S_{s's} \\ &- S_{rr'}T^{ik}_{r's'}f_{jk}S_{s's} - S_{rr'}f_{ik}T^{kj}_{r's'}S_{s's} \\ &+ S_{rr'}\tilde{T}^{ik}_{r't}(kt|sj) + (ri|kt)\tilde{T}^{kj}_{ts'}S_{s's} \\ &+ S_{rr'}\tilde{T}^{ik}_{r't}(kt|lu)\tilde{T}^{lj}_{us'}S_{s's}. \end{aligned} \tag{3.14}$$

The Fock matrix is represented by $\mathbf{f}$, the PAO overlap matrix by $\mathbf{S}$, the covariant and contra-variant amplitudes by T^{rs}_{ij} and $\tilde{T}^{rs}_{ij}$, respectively.
The WF pair indices occurring in eq.(3.14) are those of the full pair-list of the fragment, including both r.c.- and emb.-pairs. The PAO indices are restricted to the respective pair domains. This restricts all the summation ranges and tensor sizes appearing in the formulae. This applies to the ERIs, the two-internal and two-external Fock matrix and the PAO overlap matrix. All these tensors are constructed in the periodic format by the `CRYSCOR` program, and thus the occupied and virtual indices refer to real periodic WFs and their orthogonal complement, i.e. real periodic PAOs. In addition, the Fock matrix carries the information on the embedding periodic mean field potential, electrostatics and exchange.
The LMP2 residual consists of the first five terms in eq.(3.14). The quantities for the fragment needed in these terms are available from the periodic LMP2 calculation. In additon the converged LMP2 amplitudes are provided for the fragment. At this time, these amplitudes only serve as a starting guess for the d-LrCCD iterations, i.e. solving residual equations (3.14). Yet the presented interface embedding scheme and these periodic amplitudes will in the future allow to increase the level of embedding from mean-field to LMP2, such that the converged LMP2 pair amplitudes, beyond the fragment, are added in eq.(3.14), but not altered in the iterations (emb2.-pairs). This would represent the scheme similar to MP2 feedback to the strong pair CCSD residual in local Coupled Cluster methods.[36]
In order to evaluate the last three terms of eq.(3.14) density fitting approximation can be utilized for the ERIs (see sec.1.2.1). These last three terms

then are a contraction of three-index intermediates.[50]

$$\begin{aligned} S_{rr'}\tilde{T}^{ik}_{r't}(kt|sj) &= A^{ik}_{P}(P|sj), \\ (ri|kt)\tilde{T}^{kj}_{ts'}S_{s's} &= (ri|P)A^{js}_{P}, \\ S_{rr'}\tilde{T}^{ik}_{r't}(kt|lu)\tilde{T}^{lj}_{us'}S_{s's} &= B^{ir}_{P}A^{js}_{P}, \end{aligned} \tag{3.15}$$

with

$$\begin{aligned} A^{ir}_{P} &= S_{rr'}\tilde{T}^{ik}_{r't}d^{kt}_{P}, \\ B^{ir}_{Q} &= A^{ir}_{P}(P|Q). \end{aligned} \tag{3.16}$$

As pointed out earlier, the fragment-based computation's occupied and virtual index range is specified by the restricted lists of WFs, WF pairs and their respective pair domains. The contractions of eq.(3.15) and eq.(3.16) are handled in the finite-cluster framework, thus the auxiliary function indices $P, Q, \ldots$ need to be restricted. The periodic LMP2 method of `CRYSCOR` itself exploits restricted ranges of auxiliary functions.[103] To this end, a specific fit domain is constructed for each combination of WF i and PAO center $R\mathcal{R}$. The auxiliary orbitals within these fit domains are then used to fit the respective product densities $ir\mathcal{R}$ where the PAOs r are centered on the atom R. Such a multiple-fit-domain technique is presently not applicable for a fragment-based scheme because eq.(1.15) would then correspond to a non-robust fit of the ERIs.[104] That is why a single universal fit-domain (fit-cluster) has to be used for all DF-factorized terms. In the present approach the fit-cluster is defined to coincide with the PAO centers of the fragment, i.e. the PAO centers associated with the union of all pair-domains in the fragment. The integrals $(ir|P)$ and the metric $(P|Q)$ for this fit-cluster are evaluated during the LMP2 computation. It should be noted that the set of three-index integrals $(ir|P)$ which need to be evaluated by `CRYSCOR` is bigger than that required for the actual periodic LMP2 calculation when using such a universal fit-cluster (see sec.3.2.3).

3.2.3 Technical implementation

For both, the periodic LMP2 and the fragment d-LrCCD, the periodic HF determinant serves as the reference wave function. It is evaluated by the CRYSTAL[95] program and represented by WFs, which are constructed from the non-local occupied Bloch functions.[98] The Interface is split into two parts, one on the CRYSCOR side and one on the MOLPRO side.
The first part is incorporated into the periodic LMP2 module of CRYSCOR.[101] In the initial part a usual LMP2 calculation is performed. Here after the CRYSTAL HF calculation is done, CRYSCOR reads the converged Fock matrix in AO basis and the WF coefficients and transforms both to the reciprocal k-space. Then, in the k-space, the reciprocal images of the PAO coefficients are constructed and the Fock Matrix is transformed to WF and PAO basis, resulting in the internal and external Fock matrix. These reciprocal images are Fourier-back-transformed into direct space.[101] From CRYSTAL's AO overlap matrix the direct space PAO overlap matrix is evaluated analogously. These quantities are used in the periodic LMP2 calculation and, after being adjusted to the fragment specification, in the fragment-based d-LrCCD computation.
As stated before R^{fragment}, the inter-orbital distance criterion, specifies the subset of WF-pairs for which the fragment-based d-LrCCD computation is executed. Typically a smaller value is set for R^{fragment} compared to the distance criteria which specifies the periodic LMP2 pair hierarchy. The r.c.-pairs not included in $\{ij\}^{\text{fragment}}$ still contribute to the correlation energy at the periodic LMP2 level, see eq.(3.8).
A list of all WFs in the cluster pair-list $\{ij\}^{\text{fragment}}$, denoted $\{i\}^{\text{fragment}}$, is the basis for the fragment definition. The centers in each orbital domain of $\{i\}^{\text{fragment}}$ together form a union of atoms which build a cluster. That cluster represents the fragment in terms of atomic centers. The CRYSCOR interface part then writes a MOLPRO input. The geometry of this cluster,

the AO and fitting basis set specifications of the periodic calculation, the orbital domains of $\{i\}^{\text{fragment}}$'s WFs and the number of electrons (twice the number of WFs in $\{i\}^{\text{fragment}}$) is written to this input in the correct format. It should be noted here that the cluster specification in the MOLPRO input is only a formal collection of centers in order to specify the virtual and auxiliary space, in terms of PAOs and fitting functions respectively. The occupied space of the fragment is determined by the WF-list $\{i\}^{\text{fragment}}$ and thus still corresponds to the periodic HF. In fact, neither the specification in the MOLPRO input of the exact cluster geometry nor of the exponents or contraction coefficients of the basis set are important since no integral evaluation is done by MOLPRO. But type and order of the atoms is obligatory and has to coincide with that of the periodic computation, for correct processing of the transferred quantities on the MOLPRO side. With the basis set information the number of PAOs and auxiliary functions for each center of the cluster is defined. With all the above specifications the correct block sizes, pointers, offsets, etc. are set automatically, so the block structure of the passed integrals properly transfer from CRYSCOR to MOLPRO's d-LrCCD module.[21]

The internal and external Fock as well as the PAO overlap matrix, essential to the d-LrCCD calculation, are cut out from the periodic counterpart following the definition of the fragment and passed to MOLPRO's side of the interface.

The 2- and 3-index ERIs are computed during the periodic LMP2 calculation and are also transferred and used in the following d-LrCCD calculation. The 3-index integrals' $(i\,r\mathcal{R}|P\mathcal{P})$ PAO indices $r\mathcal{R}$ are restricted to the corresponding united pair domain $[i*]$, in LMP2 and d-LrCCD. The united pair domain $[i*]$ is the union of the pair domains $[i\,j\mathcal{J}]$ for all the WFs $j\mathcal{J}$ the WF i is pairing with. For the cluster for example,

$$[i*]^{\text{fragment}} \quad = \quad \cup[i\,j\mathcal{J}], \forall(i\,j\mathcal{J}) \in \{ij\}^{\text{fragment}}. \tag{3.17}$$

Since the cluster tolerance R^{fragment} is typically set smaller than the corresponding periodic LMP2 one (specifying pairs to be treated fully, without

employing the multipole approximation for the ERIs) $[i*]^{\text{fragment}}$ is already contained within the united pair domain $[i*]$ of the periodic calculation.
The situation complicates for the auxiliary functions. The 3-index ERIs $(i\,r\mathcal{R}|P\mathcal{P})$, for the periodic LMP2, are restricted to the united fit domain

$$[i*]_{\text{fit}} \quad = \quad \cup[j\mathcal{J}\,s\mathcal{S}]_{\text{fit}}, \forall(i\,j\mathcal{J}) \in \{ij\}. \tag{3.18}$$

That means the union of all fit-domains of the $jJ\,s\mathcal{S}$ product densities, where jJ pairs with i, are required. However on the d-LrCCD side, all product densities have a universal fitting function range, not a local one, hence run over the whole cluster. Therefore, it can become rather large and exceed the related $[i*]_{\text{fit}}$ for certain WFs i. In addition, index i of ERIs $(i\,r\mathcal{R}|P\mathcal{P})$ is restricted to the reference cell and only those integrals are evaluated. Through applying translational symmetry other integrals $(i\mathcal{I}\,r\mathcal{R}|P\mathcal{P})$ are generated

$$(i\mathcal{I}\,r\mathcal{R}|P\mathcal{P}) = (ir\,(\mathcal{R}\ominus\mathcal{I})\,|P\,(\mathcal{P}\ominus\mathcal{I}))\,. \tag{3.19}$$

As stated above index $P\mathcal{P}$ on the d-LrCCD side can be located on any center of the cluster, thus the range of auxiliary functions for 3-index ERIs to be computed is increased by $\mathcal{P}\ominus\mathcal{I}$. Therefor, the set of required 3-index integrals goes beyond the size required for the periodic LMP2 alone. After computation the 3-index ERIs are translated according to eq.(3.19). Once the whole set required for the fragment calculation is obtained, it is written to file for later use by `MOLPRO`'s d-LrCCD module. After completion of the periodic LMP2 calculation, the LMP2 pair energies, the assembled 4-index ERIs $(ir|js)$, as well as the LMP2 doubles amplitudes, are all cut to fit the fragment and stored in file for `MOLPRO` to read.
The second part of the interface, the fragment calculation with `MOLPRO`, starts with a dummy Hartree Fock calculation and a dummy orbital localization to set all the needed buffers for the subsequent d-LrCCD calculation. The default number of correlated orbitals in `MOLPRO` is the number of occupied valence orbitals corresponding to the fragment geometry, which

is written in the input. The actual number of WFs in the fragment, the length of the list $\{i\}^{\text{fragment}}$, can differ very much from that. That is why the correct number of correlated electrons is set explicitly in the input to twice the length of $\{i\}^{\text{fragment}}$, as stated above. Also the number of core electrons is set to zero. And, as mentioned earlier, the orbital domain of each WF in $\{i\}^{\text{fragment}}$ is defined explicitly in the input.
After the initial dummy part, the actual fragment d-LrCCD calculation starts. For that the `MOLPRO` part of the interface reads the needed quantities, provided by `CRYSCOR`,

- internal and external Fock matrices
- PAO overlap matrix
- 2- and 3-index ERIs for density fitting
- 4-index ERIs
- LMP2 amplitudes

and saves them in the proper files and records of its d-LrCCD module. For the 3-index ERIs a resort is needed to conform with `MOLPRO`'s PAO and auxiliary index block structure, proceeded by a standard molecular d-LrCCD calculation. From the provided 2- and 3-index ERIs the fitting coefficients, see eq.(1.14), are constructed. To evaluate the update on the amplitudes in the d-LrCCD iterations, the pair-specific transformation matrices $\mathbf{W}^{[ij]}$, which transform from PAO to the pseudo-canonical basis, are constructed on the `MOLPRO` side.[105] They are calculated using the provided PAO Fock and overlap matrices. After the d-LrCCD calculation converged, instead of the standard molecular expression for the correlation energy, eq.(3.7) is used. To achieve that the unit cell index of each WF in $\{i\}^{\text{fragment}}$ is transferred from `CRYSCOR` to `MOLPRO`. For systems where the periodic LMP2 calculation partitions the correlation energy into inter- and intra-slab/adsorbate contributions, as shown in sec.3.2.1, the d-LrCCD correlation energy is also partitioned accordingly. For this, the individual pair

type information for each pair from $\{ij\}^{\text{fragment}}$ is transferred from CRYSCOR to MOLPRO.

3.3 Calculations

3.3.1 Test systems and calculation variables

To test the method, we investigate physisorption on graphane (hydrogenated graphene[106;107]) of hydrogen molecules and argon atoms. Fig.3.2 presents these test systems. The geometry of graphane was taken from ref.[71], a B3LYP-optimized structure: 1.536 Å for the C-C bond, 1.101 Å for the C-H bond, and 107.4 degrees for the H-C-C angle. For the bond length of H_2 the CCSD optimized value of 0.746 Å was taken. The supplementary material (B) also includes the geometry information in form of a CRYSTAL input. The computations were carried out for the arrangement of the adsorbed H_2 mono-layer corresponding to the energetically most stable adequate structure according to the recent benchmark.[71] There the adsorption occurs in the perpendicular to the surface orientation, atop graphane's downward-pointing carbon atoms, covering every third of these sites. The argon layer was calculated analogously. In order to compare the periodic calculations to molecular results, several, progressively growing dimers were also considered. Those molecular systems which can serve as finite prototypes for the H_2-graphane systems, were:

- H_2–C_4H_{10}
- H_2–$C_{13}H_{22}$
- H_2–$C_{22}H_{34}$
- H_2–$C_{37}H_{52}$

Their geometries, i.e. C-C and C-H bond lengths, as well as angles were chosen the same as in graphane. The H_2 adsorbate was placed atop the molecular centers. For the border H atoms the C-H bond directions were chosen to coincide with the corresponding cut C-C bonds of graphane. The triple-zeta quality basis sets were used in every calculation. The Hartree-Fock reference was evaluated without diffuse orbitals on hydrogen and carbon atoms, because of convergence problems of the periodic SCF. The correlation energy was computed both without and with diffuse AOs, the latter implying addition of d- and f-AOs for carbon and p- and d-AOs for hydrogen. The diffuse high-angular momentum orbitals, known to be essential for dispersion[71;108] description, were added to the virtual space only by means of the dual basis set procedure of ref.[101]. In the following the basis set including diffuse orbitals on hydrogen and carbon is denoted AVTZ; without - VTZ. Diffuse d- and f- orbitals for argon were included in every calculation, since they did not cause any numerical problems at the periodic Hartree-Fock level. The basis sets' specifications are presented in the supplementary material (B). Wannier functions were constructed on the `CRYSTAL` side by the scheme of ref.[98]. The PAOs were constructed in `CRYSCOR` by projecting the reciprocal-space images of AOs from the occupied space, afterwards transforming the result back in direct space.[101] Both LMP2 and d-LrCCD computations employed Weigend and coworkers' auxiliary `MP2FIT` basis sets, optimized for the orbital aug-cc-pVTZ basis.[109] The local density fitting procedure presented in ref.[103], was used for the 4-index ERIs of the periodic LMP2, as well as for d-LrCCD (see sec.3.2.3). The detailed specifications of the technical variables for the calculations (k-meshes, truncation tolerances etc.) can be found in the supplementary data (B).

[16]Reprinted with permission from ref.[25]. ©2016 American Chemical Society.

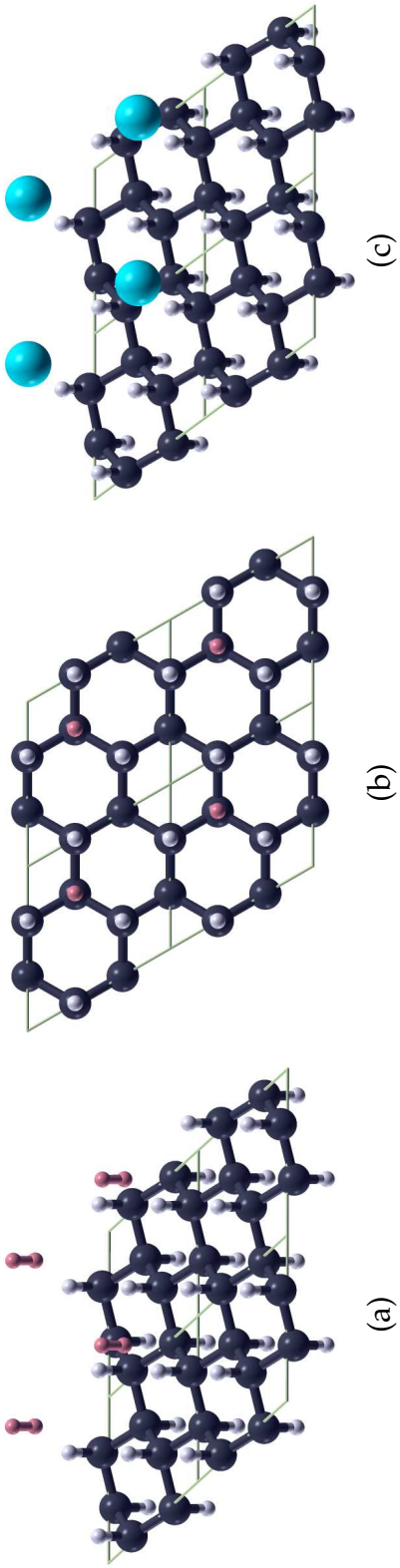

Figure 3.2: The H_2-graphane (side-top projection (a), top projection (b)) and Ar-graphane (c) systems. Dark gray spheres refer to the carbon atoms, light gray to the hydrogen atoms of graphane, pink to the hydrogen atoms of the adsorbate and blue to the argon atoms. Green bars are used to identify the supercell. The picture are made with XCRYSDEN.[110] [16]

3.3.2 Results and discussion

Firstly the presented fragment method is compared to the periodic reference at the LMP2 level for the test system of H_2-graphane, shown in fig.3.2. Here the fragment calculation remains at the LMP2 level for testing the consistency. The results of periodic, finite-cluster, fragment, and fragment/periodic hybrid are comprised in tab.3.1. Shown are the total correlation contribution $\Delta E_{\text{corr.}}$ to the interaction energy, an its inter-slab-adsorbate, intra-slab, and intra-adsorbate contributions for different sizes of cluster and fragment. In order to reach an accuracy of half a kJ/mol fairly large cluster sizes are required in the finite cluster approach. Even the cluster of H_2–$C_{37}H_{52}$, the largest one, does not fully reach the periodic value for the inter-component of $\Delta E_{\text{corr.}}$. Also the intra-H2 and intra-graphane components of $\Delta E_{\text{corr.}}$ deviate slightly from the periodic reference values. The presented fragment approach, i.e. embedding of the fragment in the periodic Hartree-Fock solution, nearly removes those differences, but only when the pair cut-off tolerance $R_{\text{intra}}^{\text{fragment}}$ is sufficiently large ($\geq$ 3 Å). The $\Delta E_{\text{corr.}}$'s inter-H2-graphane component, with a value of -5.82 kJ/mol, doesn't reach the periodic reference value of -5.97 kJ/mol either. In contrast the hybrid fragment/periodic scheme achieves this (see eq.(3.8)). The incremental periodic LMP2 correction accounts for the long-range r.c.-pairs not included in the fragment. The calculations demonstrate that the fragment/periodic hybrid LMP2 results are virtually independent of the fragment size, and match the periodic reference accurately. This confirms the correct implementation of the interface. The minor deviations of the order of a few hundredths of a kJ/mol can be attributed to the lack of certain inter-pair couplings in the fragment LMP2 computation.

As a next step we increase the fragment level to d-LrCCD. The reference here is the fragment/periodic hybrid method with largest cut-off, since an implementation of a fully periodic d-LrCCD is not available yet. The results of these computations are comprised in tab.3.2. As can be seen, the energy convergence of the d-LrCCD|LMP2 fragment/periodic hybrid

Table 3.1: LMP2 correlation contribution to the interaction energy and its intra- and inter-components for the H_2-graphane system with the separation of 3.67 Å between the center of H_2 and the midplane of graphane. The interaction energies were computed via: (i) periodic LMP2; (ii) finite-cluster LMP2 for several representative dimers; (iii) LMP2 for the periodic-HF-embedded fragments cut from the periodic structure using different cut-off distances R^{fragment}. The interaction energies in the latter case were computed for the fragments alone or via the hybrid scheme of eq.(3.8). All the calculations employed the VTZ basis. [17]

Cluster type or $R^{\text{fragment}}_{\text{inter}}/R^{\text{fragment}}_{\text{intra}}$ (Å)	ΔE_{inter} (kJ/mol)	$\Delta E_{\text{intra-slab}}$ (kJ/mol)	$\Delta E_{\text{intra-ads.}}$ (kJ/mol)	$\Delta E_{\text{corr.}}$ (kJ/mol)
Periodic LMP2				
	-5.966	0.515	0.094	-5.357
Finite-cluster LMP2				
H_2–C_4H_{10}	-4.123	0.411	0.042	-3.670
H_2–$C_{13}H_{22}$	-5.211	0.590	0.054	-4.567
H_2–$C_{22}H_{34}$	-5.595	0.630	0.051	-4.914
H_2–$C_{37}H_{52}$	-5.796	0.651	0.054	-5.091
Fragment LMP2				
4.0/2.0	-4.406	0.376	0.092	-3.937
6.9/2.0	-5.817	0.375	0.092	-5.351
6.9/3.0	-5.818	0.527	0.092	-5.200
6.9/4.0	-5.824	0.543	0.092	-5.190
Fragment LMP2 | Periodic LMP2				
4.0/2.0	-5.961	0.544	0.094	-5.322
6.9/2.0	-5.955	0.550	0.094	-5.311
6.9/3.0	-5.958	0.547	0.094	-5.317
6.9/4.0	-5.963	0.548	0.094	-5.322

scheme with fragment size is not as rapid as in the above presented trivial LMP2|LMP2 case. Naturally, there are additional long range couplings in thed-LrCCD method which are not captured by the periodic LMP2. Regardless, d-LrCCD|LMP2 converges much faster than the pure d-LrCCD fragment method, not to mention the finite-cluster approach. The binding

[17] Reprinted with permission from ref.[25]. ©2016 American Chemical Society.

energy for H_2-graphane is noticeably weaker than LMP2, by more than 1 kJ/mol, as anticipated. This is mainly due to the decrease in the inter-H_2-graphane attraction. The difference between the d-LrCCD finite-cluster intra-components of the interaction energy and the one evaluated by the embedded fragment, or the fragment/periodic hybrid methods, is here more substantial compared to the case of LMP2 (tab.3.1), which suggests that the proper embedding, as an essential part when aiming at high accuracy, even if the electrostatic contribution to the binding is small. The effect of the additional approximation, restriction of emb.-pairs to diagonal ones, is rather small and leads to a deviation in $\Delta E_{\text{corr.}}$ of 0.1 kJ/mol at most. In addition for large cut-off distances $R_{\text{intra}}^{\text{fragment}}$, there is a significant reduction in the number of pairs, and therefore cost, due to this approximation.
The embedded d-LrCCD scheme's performance was tested on argon atom and hydrogen molecule adsorbed on graphane, as shown in fig.3.2. The interaction energy for these systems is displayed in fig.3.3 as a function of graphane-adsorbate distance. The plots show the curves for both periodic LMP2 and embedded d-LrCCD. In the case of argon adsorption, the potential curve is much deeper compared to H_2, as anticipated. Due to dispersion being the main attractive force in both cases, the higher polarizability of argon than H_2 results in a stronger Ar-graphane binding. The error of MP2 when comparing different types of intermolecular systems is not systematic, e.g. see sec.2.3.2 fig.2.9. For example, MP2 is known to show excellent accuracy for complexes with hydrogen bonds. On the other hand, MP2 significantly overestimates interaction when investigating strongly polarizable van der Waals bound systems, and underestimating it for weakly polarizable ones, because of the uncoupled Hartree-Fock treatment of dispersion.[111] Returning to our case, MP2 underestimates the binding of hydrogen molecules adsorbed on graphane, when considering CCSD(T) reference, by about 10%.[71] There is no accurate benchmarks for argon on graphane in the literature, but due to the higher polarizability of argon

[18]Reprinted with permission from ref.[25]. ©2016 American Chemical Society.

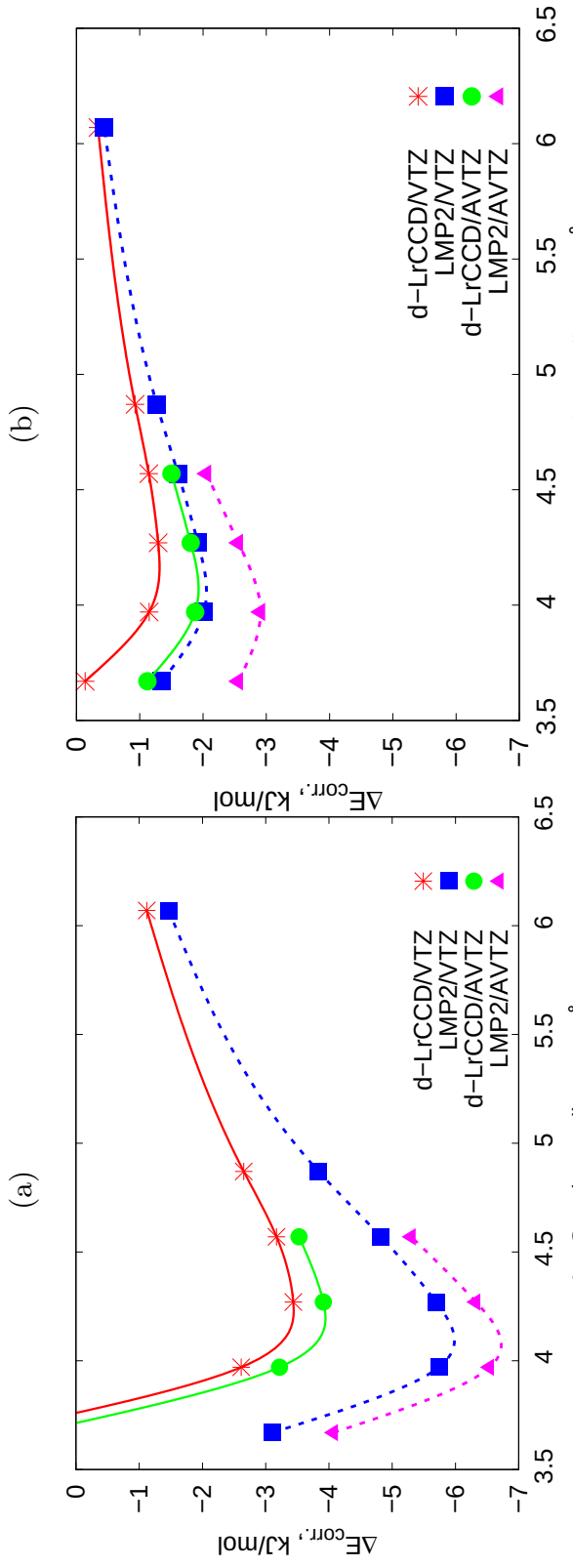

Figure 3.3: Periodic LMP2 and embedded d-LrCCD counterpoise-corrected potential energies curves for adsorption of Ar atom (a) or H_2 molecule (b) on graphane atop a carbon atom (see Fig.3.2). The VDZ and AVDZ basis sets, specified in section 3.3.1 and supplementary material B were used. The symbols indicate the explicitly computed energies, while the connecting curves were obtained by the cubic spline interpolations between these points. [18]

over H_2, it is likely that MP2 overestimates the Ar-graphane interaction energy.

The treatment of dispersion by d-LrCCD at the coupled-perturbed Hartree-Fock level[112] leads to a loss of interaction energy compared to LMP2, in both studied cases. The correction for argon is substantially larger than for H_2 and should bring the former also to the pronouncedly underbinding regime. This is in agreement with the molceular benchmarks of d-LrCCD in sec.2.3.1, which states the general tendency to underestimate intermolecular interaction energies of this method. For a study of relative stabilities between different compounds this however still prove useful, at least in comparison to LMP2. By addition of diffuse orbitals the deepening of the potential curve points out their essentiality for describing dispersion, which is in agreement with previous studies.

[19]Reprinted with permission from ref.[25].

Table 3.2: d-LrCCD correlation contribution to the interaction energy and its intra- and inter-components for the H_2-graphane system with the separation of 3.67 Å between the center of H_2 and the midplane of graphane. The interaction energies were computed via (i) finite-cluster d-LrCCD for several representative dimers; (iii) d-LrCCD for the periodic-HF-embedded fragments cut from the periodic structure using different cut-off distances R^{fragment}. The interaction energies in the latter case were computed for the fragments alone or via the hybrid scheme of eq.(3.8). In the fragment calculations, the approximation that restricts the emb.-pairs to the diagonal ones only was also tested (marked as "Diag" vs "All"). In addition to the energies, the number of atoms, WFs, WF pairs and 3-index ERIs (for the fragment d-LrCCD calculations) are also given. The calculations employed the VTZ basis. [19]

Cluster type or $R^{\text{fragment}}_{\text{inter}}/R^{\text{fragment}}_{\text{intra}}$ (Å)	$N_{\text{WF}}/N_{\text{atoms}}$	Emb.-pairs	$N_{\text{WF-pairs}}$	$N_{\text{3idx-ERIs}}$ (millions)	ΔE_{inter} (kJ/mol)	$\Delta E_{\text{intra-slab}}$ (kJ/mol)	$\Delta E_{\text{intra-ads.}}$ (kJ/mol)	$\Delta E_{\text{corr.}}$ (kJ/mol)
Finite-cluster d-LrCCD								
H_2–C_4H_{10}	14/16		105		-3.217	0.401	0.124	-2.692
H_2–$C_{13}H_{22}$	38/37		741		-3.977	0.573	0.157	-3.247
H_2–$C_{22}H_{34}$	62/58		1953		-4.237	0.623	0.162	-3.452
H_2–$C_{37}H_{52}$	101/91		5151		-4.379	0.651	0.170	-3.558
Fragment d-LrCCD								
4.0/2.0	62 / 59	Diag	249	70	-3.534	0.197	0.177	-3.160
		All	390	103	-3.512	0.194	0.176	-3.141
6.9/2.0	127 / 114	Diag	413	207	-4.556	0.212	0.210	-4.134
		All	554	269	-4.531	0.210	0.209	-4.112
6.9/3.0	145 / 130	Diag	767	415	-4.546	0.314	0.209	-4.023
		All	1547	769	-4.489	0.308	0.208	-3.973
6.9/4.0	165 / 144	Diag	1111	653	-4.544	0.317	0.209	-4.017
		All	3004	1579	-4.439	0.315	0.206	-3.918
Fragment d-LrCCD|Periodic LMP2								
4.0/2.0	62 / 59	Diag	249	70	-5.089	0.365	0.179	-4.545
		All	390	103	-5.067	0.363	0.178	-4.527
6.0/2.0	98 / 89	Diag	345	138	-4.736	0.385	0.209	-4.143
		All	486	186	-4.712	0.383	0.208	-4.121
6.9/2.0	127 / 114	Diag	413	207	-4.694	0.387	0.212	-4.095
		All	554	269	-4.670	0.385	0.212	-4.073
8.0/2.0	167 / 147	Diag	501	324	-4.680	0.391	0.214	-4.075
		All	642	404	-4.655	0.388	0.213	-4.053
9.0/2.0	212 / 184	Diag	594	475	-4.672	0.390	0.215	-4.067
6.9/1.0	119 / 109	Diag	279	129	-4.701	0.451	0.212	-4.039
		All	288	133	-4.702	0.451	0.212	-4.037
6.9/2.0	127 / 114	Diag	413	207	-4.694	0.387	0.212	-4.095
		All	554	269	-4.670	0.385	0.212	-4.073
6.9/3.0	145 / 130	Diag	767	415	-4.686	0.334	0.212	-4.140
		All	1547	769	-4.628	0.328	0.210	-4.090
6.9/4.0	165 / 144	Diag	1111	653	-4.684	0.323	0.212	-4.149
		All	3004	1579	-4.579	0.320	0.209	-4.050

3.4 Conclusion

This chapter introduced an interface between the two quantum chemical programs CRYSCOR and MOLPRO enabling correlated computations on a fragment of a periodic system embedded in the periodic mean field solution, by utilizing the local coupled cluster module of the MOLPRO package. At the moment the fragment treatment is restricted to direct local ring CCD (d-LrCCD). The long-range correlation part outside the fragment is calculated at the periodic LMP2 level leading to the hybrid d-LrCCD|LMP2 approach.
The embedded fragment's virtual and occupied space is built from Wannier functions and PAOs obtained in the periodic part of the method. Therefore the periodic framework, done by CRYSCOR, produces all transformed matrix elements of the Hamiltonian needed for the correlation treatment of the fragment. This means Fock matrix and ERIs in Wannier/PAO basis are passed from CRYSCOR to MOLPRO for the subsequent fragment computation. This represents a seamless embedding of the fragment in the real periodic mean field. This is superior compared to any finite-cluster or a point-multipole type embedding scheme, which are commonly used for solids.
This work is a first step towards a universal interface for a local coupled cluster treatment of fragments embedded in the periodic solution, with the final goal of treatment up to LCCSD(T). The current stage is embedding at the mean field level, but an increase to the periodic LMP2 level, and at length an approximation to the periodic coupled cluster model such as LCCD[S]-R^{-6} (see sec.2.2.2) when they become available. The current fragment level d-LrCCD sums the ring-type diagrams to infinity, leading to a correct treatment of small-gap systems. In this respect, for such systems, the d-LrCCD method is a much better low-order reference than LMP2 in hierarchical incremental protocols.[113] Generally d-LrCCD tends

to underestimate dispersion when using a HF reference. However, since this underbinding is quite systematic, this method is a better choice for comparative studies of adsorption with various adsorbate types of large and small polarizability than MP2.

Chapter 4

Summary

In the course of this work a new and enhanced weak pair treatment for local coupled cluster (LCC) methods and a fragment-based embedded approach for periodic systems, both based on approximate coupled cluster models, are presented.
To avoid the high cost of a full CCSD calculation, CCSD calculations with a local pair approximation are usually employed. There the strong pairs are treated at the full LCCSD level, and the weak pairs, where the inter-orbital distance of the pair is larger than a certain threshold, are treated at a lower level. Normally that lower level method is LMP2. Despite the fact that LMP2 does not provide sufficient accuracy many cases, e.g. delocalized π-systems.
In this work, we propose an alternative scheme exploiting the R^{-6} decay of a pair energy with respect to the inter-orbital distance R. The presented methods for close/weak pair treatment is based on coupled cluster formalism with the inclusion of third and fourth order diagrams, that have a decay rate of R^{-6}. These hybrid methods are denoted LCCSD|X, where X stands for the level of close/weak pair treatment.
A first substantial improvement over the LMP2 treatment was provided by the LCCSD|LrCCD3 model. LCCSD|LrCCD3 comprises all diagrams of the ring approximation decaying as R^{-6} and being up to the third order in

the fluctuation potential. LrCCD3 weak pair treatment results in a superior description of van der Waals dominated systems and is on a par with LMP2 in electrostatic dominated cases.
In order to improve on LrCCD3, it was extended to the LCCD[S]-R^{-6} method, abandoning the ring approximation and including all diagrams decaying as R^{-6} and being up to the fourth order. The fourth order singles correction [S] is evaluated through an *a posteriori* perturbative estimate. LCCSD|LCCD[S]-R^{-6} provides another quite significant improvement now over LrCCD3. The interaction energies for inter-molecular systems, computed using this method, are in very good agreement with the full LCCSD reference, while being still very cost effective.
Further in this thesis, an interface between the two quantum chemical packages `MOLPRO` and `CRYSCOR` is presented. This interface enables correlated calculations on a fragment of a periodic system, which is embedded in the periodic mean field solution. In order to do that it bundles the coupled cluster module of the `MOLPRO` program with `CRYSCOR`'s LMP2. The fragment is treated at the d-LrCCD level. A hybrid approach framework, named d-LrCCD|LMP2, adds the long-range correlation part outside the fragment computed at the periodic LMP2 level to the fragments d-LrCCD. The needed quantities, that reflect the periodic embedding, are evaluated within a periodic LMP2 calculation and passed to `MOLPRO` for the following fragment treatment. This seamless embedding of the fragment in the real periodic mean field, is superior to usually used embedding types like finite-cluster and point-multipole. The current fragment treatment of d-LrCCD has some advantages over LMP2: it sums the ring-type diagrams to infinity for a correct description of small-gap systems. The presented scheme is a first step towards a universal interface for a LCC treatment of fragments embedded in the periodic solution.

Appendix A

Supplementary data for chapter 2

All of the supplementary material is reprinted from refs.[21; 24]with the permission of AIP Publishing.

Table A.1: The intermolecular interaction energies E_{int}, the correlation contributions ΔE_{c} and the intra- δE_{intra} and intermolecular δE_{inter} components thereof, calculated at the full LCCSD level (i.e. all pairs are strong) as well as using the hybrid LCCSD|X schemes, with X=LMP2, X=LrCCD3, X=LrCCD, X=LCCD-R^{-6}, X=LCCD[S]-R^{-6}, for a set of 25 dimers. All energies are in kcal/mol. The RMS and maximum deviations with respect to the LCCSD calculation are also given.

		LCCSD|				
	full	LMP2	LrCCD3	LrCCD	LCCD-R^{-6}	LCCD[S]-R^{-6}
			Electrostatic dominated			
			Water dimer			
δE_{inter}	-2.318	-2.206	-2.324	-2.365	-2.205	-2.293
δE_{intra}	1.594	1.593	1.598	1.604	1.586	1.586
ΔE_{c}	-0.724	-0.612	-0.726	-0.761	-0.620	-0.707
E_{int}	-4.349	-4.238	-4.351	-4.386	-4.244	-4.332
			Peptide-Peptide			
δE_{inter}	-5.663	-5.620	-5.793	-5.945	-5.458	-5.652
δE_{intra}	3.070	3.106	3.092	3.121	3.046	3.046
ΔE_{c}	-2.593	-2.514	-2.701	-2.824	-2.412	-2.606
E_{int}	-7.162	-7.081	-7.269	-7.392	-6.980	-7.174

Appendix A. Supplementary data for chapter 2

Table A.1: The intermolecular interaction energies E_{int}, the correlation contributions ΔE_{c} and the intra- δE_{intra} and intermolecular δE_{inter} components thereof, calculated at the full LCCSD level (i.e. all pairs are strong) as well as using the hybrid LCCSD|X schemes, with X=LMP2, X=LrCCD3, X=LrCCD, X=LCCD-R^{-6}, X=LCCD[S]-R^{-6}, for a set of 25 dimers. All energies are in kcal/mol. The RMS and maximum deviations with respect to the LCCSD calculation are also given.

			Uracil-AcNH$_2$			
δE_{inter}	-9.148	-8.968	-9.253	-9.476	-8.728	-9.124
δE_{intra}	7.254	7.299	7.281	7.319	7.215	7.215
ΔE_{c}	-1.894	-1.668	-1.972	-2.156	-1.513	-1.908
E_{int}	-17.113	-16.888	-17.192	-17.376	-16.733	-17.128
			Pyridine-Methanol			
δE_{inter}	-4.624	-4.634	-4.698	-4.823	-4.437	-4.585
δE_{intra}	2.622	2.669	2.641	2.666	2.605	2.605
ΔE_{c}	-2.002	-1.966	-2.057	-2.157	-1.832	-1.980
E_{int}	-6.226	-6.189	-6.281	-6.381	-6.056	-6.203
			Guanine -Cytosine (Watson Crick)			
δE_{inter}	-13.860	-13.731	-14.081	-14.445	-13.248	-13.823
δE_{intra}	10.570	10.673	10.624	10.690	10.514	10.514
ΔE_{c}	-3.290	-3.058	-3.457	-3.755	-2.735	-3.309
E_{int}	-27.729	-27.497	-27.897	-28.194	-27.174	-27.749
			Formamid dimer			
δE_{inter}	-7.962	-7.738	-8.055	-8.242	-7.591	-7.936
δE_{intra}	6.001	6.037	6.026	6.057	5.969	5.969
ΔE_{c}	-1.961	-1.702	-2.030	-2.185	-1.622	-1.967
E_{int}	-14.032	-13.773	-14.101	-14.256	-13.694	-14.038
			2-Pyridoxine - 2-Aminopyridine			
δE_{inter}	-10.226	-10.314	-10.427	-10.731	-9.812	-10.192
δE_{intra}	6.478	6.601	6.530	6.592	6.442	6.442
ΔE_{c}	-3.748	-3.713	-3.897	-4.139	-3.370	-3.750
E_{int}	-14.150	-14.118	-14.302	-14.543	-13.775	-14.155
			Formic acid dimer			

Appendix A. Supplementary data for chapter 2

Table A.1: The intermolecular interaction energies E_{int}, the correlation contributions ΔE_{c} and the intra- δE_{intra} and intermolecular δE_{inter} components thereof, calculated at the full LCCSD level (i.e. all pairs are strong) as well as using the hybrid LCCSD|X schemes, with X=LMP2, X=LrCCD3, X=LrCCD, X=LCCD-R^{-6}, X=LCCD[S]-R^{-6}, for a set of 25 dimers. All energies are in kcal/mol. The RMS and maximum deviations with respect to the LCCSD calculation are also given.

δE_{inter}	-10.152	-9.899	-10.156	-10.389	-9.611	-10.084
δE_{intra}	8.958	8.996	8.972	9.008	8.909	8.909
ΔE_{c}	-1.194	-0.902	-1.184	-1.380	-0.702	-1.175
E_{int}	-16.339	-16.049	-16.330	-16.527	-15.849	-16.322
			Ammonia dimer			
δE_{inter}	-2.039	-1.958	-2.107	-2.147	-1.986	-2.028
δE_{intra}	0.903	0.907	0.915	0.922	0.900	0.900
ΔE_{c}	-1.136	-1.051	-1.192	-1.225	-1.086	-1.128
E_{int}	-2.523	-2.440	-2.581	-2.614	-2.474	-2.516
			dispersion dominated			
			Indole-Benzene (stacked)			
δE_{inter}	-13.234	-15.660	-14.116	-14.884	-13.236	-13.193
δE_{intra}	4.555	5.356	4.765	4.994	4.582	4.582
ΔE_{c}	-8.679	-10.305	-9.352	-9.891	-8.655	-8.611
E_{int}	-1.632	-3.258	-2.305	-2.845	-1.608	-1.565
			Pyridine-Pyridine (π-π)			
δE_{inter}	-7.707	-9.148	-8.346	-8.795	-7.791	-7.724
δE_{intra}	2.388	2.857	2.536	2.671	2.415	2.415
ΔE_{c}	-5.319	-6.291	-5.810	-6.124	-5.376	-5.310
E_{int}	-1.976	-2.948	-2.467	-2.781	-2.034	-1.968
			Benzene-Benzene(π-π)			
δE_{inter}	-7.035	-8.368	-7.672	-8.086	-7.144	-7.063
δE_{intra}	1.965	2.397	2.110	2.236	1.990	1.990
ΔE_{c}	-5.070	-5.971	-5.562	-5.850	-5.154	-5.074
E_{int}	-1.103	-2.003	-1.594	-1.882	-1.186	-1.106
			Uracil-Ethene			

Appendix A. Supplementary data for chapter 2

Table A.1: The intermolecular interaction energies E_{int}, the correlation contributions ΔE_{c} and the intra- δE_{intra} and intermolecular δE_{inter} components thereof, calculated at the full LCCSD level (i.e. all pairs are strong) as well as using the hybrid LCCSD|X schemes, with X=LMP2, X=LrCCD3, X=LrCCD, X=LCCD-R^{-6}, X=LCCD[S]-R^{-6}, for a set of 25 dimers. All energies are in kcal/mol. The RMS and maximum deviations with respect to the LCCSD calculation are also given.

δE_{inter}	-5.364	-5.821	-5.682	-5.898	-5.316	-5.369
δE_{intra}	1.897	2.054	1.958	2.014	1.892	1.892
ΔE_{c}	-3.467	-3.766	-3.724	-3.885	-3.424	-3.478
E_{int}	-2.056	-2.357	-2.316	-2.476	-2.015	-2.069
Uracil-Neopentane						
δE_{inter}	-6.083	-6.313	-6.317	-6.525	-5.943	-6.086
δE_{intra}	1.895	1.994	1.932	1.979	1.875	1.875
ΔE_{c}	-4.188	-4.318	-4.384	-4.546	-4.068	-4.211
E_{int}	-2.105	-2.237	-2.304	-2.466	-1.987	-2.130
Guanine-Cytosine (stacked)						
δE_{inter}	-18.237	-20.212	-18.727	-19.555	-17.693	-18.145
δE_{intra}	9.193	9.834	9.295	9.503	9.135	9.135
ΔE_{c}	-9.045	-10.378	-9.431	-10.053	-8.558	-9.010
E_{int}	-14.455	-15.790	-14.845	-15.465	-13.971	-14.423
Uracil-Uracil (π-π)						
δE_{inter}	-11.953	-13.067	-12.297	-12.809	-11.599	-11.932
δE_{intra}	4.867	5.202	4.919	5.042	4.811	4.811
ΔE_{c}	-7.086	-7.865	-7.378	-7.768	-6.788	-7.121
E_{int}	-6.684	-7.464	-6.977	-7.367	-6.387	-6.721
Uracil-Ethyne						
δE_{inter}	-5.072	-5.628	-5.379	-5.602	-5.036	-5.064
δE_{intra}	2.145	2.326	2.201	2.260	2.139	2.139
ΔE_{c}	-2.926	-3.302	-3.178	-3.343	-2.898	-2.925
E_{int}	-2.540	-2.915	-2.791	-2.955	-2.511	-2.538
Uracil-Pyridine (π-π)						
δE_{inter}	-10.090	-11.483	-10.602	-11.107	-9.967	-10.100

Table A.1: The intermolecular interaction energies E_{int}, the correlation contributions ΔE_{c} and the intra- δE_{intra} and intermolecular δE_{inter} components thereof, calculated at the full LCCSD level (i.e. all pairs are strong) as well as using the hybrid LCCSD|X schemes, with X=LMP2, X=LrCCD3, X=LrCCD, X=LCCD-R^{-6}, X=LCCD[S]-R^{-6}, for a set of 25 dimers. All energies are in kcal/mol. The RMS and maximum deviations with respect to the LCCSD calculation are also given.

δE_{intra}	3.712	4.153	3.815	3.951	3.700	3.700
ΔE_{c}	-6.379	-7.329	-6.787	-7.157	-6.267	-6.400
E_{int}	-4.290	-5.243	-4.701	-5.070	-4.181	-4.313
			Pyrazine dimer			
δE_{inter}	-0.015	-11.178	-10.089	-10.635	-9.436	-9.364
δE_{intra}	0.005	3.823	3.395	3.560	3.257	3.257
ΔE_{c}	-0.010	-7.355	-6.694	-7.074	-6.179	-6.108
E_{int}	-1.975	-3.171	-2.510	-2.891	-1.995	-1.924
			Ethene dimer			
δE_{inter}	-2.367	-2.353	-2.555	-2.613	-2.376	-2.370
δE_{intra}	0.741	0.758	0.777	0.791	0.745	0.745
ΔE_{c}	-1.626	-1.595	-1.777	-1.822	-1.631	-1.626
E_{int}	-0.789	-0.757	-0.939	-0.984	-0.793	-0.788
			mixed			
			Pyridine-Pyridine (TS)			
δE_{inter}	-5.538	-6.040	-5.841	-6.073	-5.475	-5.478
δE_{intra}	2.406	2.607	2.490	2.553	2.423	2.423
ΔE_{c}	-3.132	-3.433	-3.351	-3.520	-3.052	-3.055
E_{int}	-2.254	-2.557	-2.475	-2.644	-2.176	-2.179
			Benzene-Benzene(TS)			
δE_{inter}	-5.172	-5.724	-5.527	-5.757	-5.161	-5.140
δE_{intra}	2.101	2.300	2.188	2.252	2.116	2.116
ΔE_{c}	-3.071	-3.424	-3.338	-3.505	-3.045	-3.023
E_{int}	-1.618	-1.969	-1.883	-2.050	-1.590	-1.568
			Indole - Benzene (TS)			
δE_{inter}	-7.440	-8.223	-7.849	-8.173	-7.348	-7.374

Appendix A. Supplementary data for chapter 2

Table A.1: The intermolecular interaction energies E_{int}, the correlation contributions ΔE_{c} and the intra- δE_{intra} and intermolecular δE_{inter} components thereof, calculated at the full LCCSD level (i.e. all pairs are strong) as well as using the hybrid LCCSD|X schemes, with X=LMP2, X=LrCCD3, X=LrCCD, X=LCCD-R^{-6}, X=LCCD[S]-R^{-6}, for a set of 25 dimers. All energies are in kcal/mol. The RMS and maximum deviations with respect to the LCCSD calculation are also given.

δE_{intra}	3.236	3.522	3.345	3.431	3.252	3.252
ΔE_{c}	-4.204	-4.701	-4.505	-4.741	-4.096	-4.122
E_{int}	-3.909	-4.408	-4.211	-4.448	-3.803	-3.828
			Phenol dimer			
δE_{inter}	-6.669	-7.002	-6.901	-7.141	-6.481	-6.639
δE_{intra}	3.121	3.281	3.177	3.237	3.109	3.109
ΔE_{c}	-3.548	-3.721	-3.724	-3.904	-3.372	-3.530
E_{int}	-5.346	-5.517	-5.519	-5.700	-5.168	-5.326
			Ethene - Ethine			
δE_{inter}	-1.527	-1.588	-1.658	-1.707	-1.536	-1.517
δE_{intra}	0.839	0.872	0.866	0.879	0.845	0.845
ΔE_{c}	-0.687	-0.716	-0.792	-0.827	-0.691	-0.672
E_{int}	-1.115	-1.144	-1.219	-1.255	-1.119	-1.100
			RMS			
δE_{inter}		0.940	0.378	0.729	0.266	0.036
δE_{intra}		0.310	0.086	0.180	0.029	0.029
ΔE_{c}		0.637	0.293	0.550	0.240	0.033
E_{int}		0.637	0.294	0.550	0.240	0.033
			Max deviation			
δE_{inter}		2.426	0.882	1.650	0.612	0.092
δE_{intra}		0.800	0.209	0.438	0.057	0.057
ΔE_{c}		1.626	0.673	1.212	0.556	0.082
E_{int}		1.626	0.673	1.213	0.555	0.080

Table A.2: The intermolecular interaction energies E_{int} from the full LCCSD(T) and hybrid LCCSD(T)|X calculations, with X=LMP2, X=LrCCD3, X=LrCCD, X=LCCD-R^{-6}, X=LCCD[S]-R^{-6}, and the values of the triples correction ΔE_{T}, for a set of 25 dimers. All energies are in kcal/mol. The RMS and maximum deviations with respect to the LCCSD calculation are also given.

		LCCSD(T)|				
	full	LMP2	LrCCD3	LrCCD	LCCD-R^{-6}	LCCD[S]-R^{-6}
			electrostatic dominated			
			Water dimer			
E_{int}	-4.522	-4.390	-4.524	-4.566	-4.398	-4.486
ΔE_{T}	-0.173	-0.152	-0.173	-0.180	-0.154	-0.154
			Peptide-Peptide			
E_{int}	-7.691	-7.599	-7.818	-7.968	-7.463	-7.657
ΔE_{T}	-0.529	-0.518	-0.549	-0.576	-0.483	-0.483
			Uracil-AcNH$_2$			
E_{int}	-17.731	-17.473	-17.831	-18.061	-17.257	-17.652
ΔE_{T}	-0.618	-0.585	-0.639	-0.685	-0.524	-0.524
			Pyridine-Methanol			
E_{int}	-6.652	-6.621	-6.723	-6.845	-6.451	-6.598
ΔE_{T}	-0.426	-0.432	-0.442	-0.464	-0.395	-0.395
			Guanine-Cytosine (Watson Crick)			
E_{int}	-28.550	-28.292	-28.773	-29.147	-27.857	-28.432
ΔE_{T}	-0.821	-0.795	-0.876	-0.953	-0.683	-0.683
			Formamid dimer			
E_{int}	-14.596	-14.284	-14.682	-14.875	-14.172	-14.516
ΔE_{T}	-0.564	-0.510	-0.581	-0.618	-0.478	-0.478
			2-Pyridoxine - 2-Amonipyridine			
E_{int}	-15.072	-15.075	-15.271	-15.577	-14.989	-14.989
ΔE_{T}	-0.923	-0.957	-0.969	-1.034	-1.214	-0.834
			Formic acid			
E_{int}	-16.909	-16.557	-16.901	-17.144	-16.303	-16.776

Table A.2: The intermolecular interaction energies E_{int} from the full LCCSD(T) and hybrid LCCSD(T)|X calculations, with X=LMP2, X=LrCCD3, X=LrCCD, X=LCCD-R^{-6}, X=LCCD[S]-R^{-6}, and the values of the triples correction ΔE_{T}, for a set of 25 dimers. All energies are in kcal/mol. The RMS and maximum deviations with respect to the LCCSD calculation are also given.

ΔE_{T}	-0.569	-0.508	-0.571	-0.617	-0.454	-0.454
Ammonia dimer						
E_{int}	-2.721	-2.624	-2.789	-2.828	-2.662	-2.704
ΔE_{T}	-0.199	-0.184	-0.208	-0.214	-0.188	-0.188
dispersion dominated						
Indole-Benzene (stacked)						
E_{int}	-3.447	-5.578	-4.347	-5.047	-3.432	-3.389
ΔE_{T}	-1.815	-2.319	-2.042	-2.202	-1.824	-1.824
Pyridine-Pyridine (π-π)						
E_{int}	-3.065	-4.327	-3.706	-4.114	-3.137	-3.071
ΔE_{T}	-1.089	-1.379	-1.239	-1.333	-1.103	-1.103
Benzene-Benzene(π-π)						
E_{int}	-2.158	-3.323	-2.797	-3.171	-2.261	-2.180
ΔE_{T}	-1.055	-1.320	-1.203	-1.289	-1.075	-1.075
Uracil-Ethene						
E_{int}	-2.739	-3.108	-3.068	-3.268	-2.685	-2.738
ΔE_{T}	-0.683	-0.751	-0.752	-0.792	-0.670	-0.670
Uracil-Neopentane						
E_{int}	-2.876	-3.029	-3.122	-3.318	-2.728	-2.870
ΔE_{T}	-0.771	-0.792	-0.818	-0.852	-0.741	-0.741
Guanine-Cytosine (stacked)						
E_{int}	-16.401	-18.064	-16.925	-17.706	-15.792	-16.244
ΔE_{T}	-1.946	-2.274	-2.081	-2.240	-1.821	-1.821
Uracil-Uracil (π-π)						
E_{int}	-8.252	-9.188	-8.636	-9.121	-7.869	-8.202
ΔE_{T}	-1.568	-1.724	-1.659	-1.754	-1.482	-1.482

Table A.2: The intermolecular interaction energies E_{int} from the full LCCSD(T) and hybrid LCCSD(T)|X calculations, with X=LMP2, X=LrCCD3, X=LrCCD, X=LCCD-R^{-6}, X=LCCD[S]-R^{-6}, and the values of the triples correction ΔE_{T}, for a set of 25 dimers. All energies are in kcal/mol. The RMS and maximum deviations with respect to the LCCSD calculation are also given.

Uracil-Ethyne						
E_{int}	-3.167	-3.635	-3.484	-3.690	-3.127	-3.154
ΔE_{T}	-0.627	-0.720	-0.693	-0.735	-0.616	-0.616
Uracil-Pyridine (π-π)						
E_{int}	-5.609	-6.810	-6.143	-6.614	-5.463	-5.596
ΔE_{T}	-1.319	-1.567	-1.442	-1.544	-1.282	-1.282
Pyrazine dimer						
E_{int}	-3.240	-4.804	-3.947	-4.443	-3.269	-3.197
ΔE_{T}	-1.266	-1.633	-1.437	-1.552	-1.273	-1.273
Ethene dimer						
E_{int}	-1.065	-1.031	-1.248	-1.303	-1.070	-1.065
ΔE_{T}	-0.276	-0.274	-0.309	-0.319	-0.277	-0.277
mixed						
Pyridine-Pyridine (TS)						
E_{int}	-2.857	-3.267	-3.149	-3.362	-2.775	-2.778
ΔE_{T}	-0.603	-0.710	-0.674	-0.718	-0.599	-0.599
Benzene-Benzene(TS)						
E_{int}	-2.220	-2.691	-2.566	-2.777	-2.194	-2.173
ΔE_{T}	-0.602	-0.722	-0.683	-0.727	-0.604	-0.604
Indole - Benzene (TS)						
E_{int}	-4.762	-5.425	-5.161	-5.462	-4.644	-4.669
ΔE_{T}	-0.853	-1.017	-0.950	-1.014	-0.841	-0.841
Phenol dimer						
E_{int}	-6.046	-6.281	-6.270	-6.496	-5.829	-5.987
ΔE_{T}	-0.699	-0.765	-0.751	-0.797	-0.661	-0.661
Ethene - Ethine						

Appendix A. Supplementary data for chapter 2

Table A.2: The intermolecular interaction energies E_{int} from the full LCCSD(T) and hybrid LCCSD(T)|X calculations, with X=LMP2, X=LrCCD3, X=LrCCD, X=LCCD-R^{-6}, X=LCCD[S]-R^{-6}, and the values of the triples correction ΔE_{T}, for a set of 25 dimers. All energies are in kcal/mol. The RMS and maximum deviations with respect to the LCCSD calculation are also given.

E_{int}	-1.065	-1.031	-1.248	-0.600	-1.070	-1.262
ΔE_{T}	-0.276	-0.274	-0.309	-0.215	-0.277	-0.162
			RMS			
E_{int}		0.817	0.385	0.712	0.299	0.067
ΔE_{T}		0.181	0.092	0.162	0.059	0.059
			Maximum deviation			
E_{int}		2.131	0.900	1.600	0.693	0.157
ΔE_{T}		0.504	0.227	0.387	0.138	0.138

Table A.3: Intermolecular interaction energies and its components are compared on the level of LCCSD|X with X=LMP2 *uc* (for uncoupled), X=d-LrCCD *uc*, X=LMP2, X=*direct*-LrCCD, X=LrCCD3, X=LrCCD4, to a full LCCSD calculation (all pairs strong) for the set of electrostatic dominated dimers. Correlation contributions ΔE_c to the respective intermolecular interaction energies, and their intra- and intermolecular components δE_{intra}, and δE_{inter}, respectively, all in [kcal/mol]. The RMS and maximum devivations relative to the CCSD calculation for all dimers is given in tab.A.5. The prefix "L" for local was omitted in the method names, for brevity.

	CCSD	MP2	CCSD| MP2*uc*	CCSD| d-rCCD*uc*	CCSD| MP2	CCSD| d-rCCD	CCSD| rCCD3	CCSD| rCCD4
				electrostatic dominated				
				water dimer				
δE_{inter}	-2.318	-2.205	-2.205	-1.863	-2.206	-1.795	-2.324	-2.365
δE_{intra}	1.594	1.277	1.362	1.362	1.593	1.512	1.596	1.604
ΔE_c	-0.724	-0.929	-0.844	-0.501	-0.612	-0.283	-0.727	-0.761
E_{int}	-4.349	-4.554	-4.469	-4.126	-4.238	-3.909	-4.352	-4.386
				peptide-peptide				
δE_{inter}	-5.663	-5.621	-5.621	-4.427	-5.620	-4.200	-5.793	-5.945
δE_{intra}	3.070	2.364	2.274	2.274	3.106	2.792	3.092	3.121
ΔE_c	-2.593	-3.257	-3.348	-2.153	-2.514	-1.408	-2.701	-2.824
E_{int}	-7.162	-7.825	-7.915	-6.721	-7.081	-5.976	-7.269	-7.392
				uracil-AcNH_2				
δE_{inter}	-9.148	-8.970	-8.970	-7.458	-8.968	-7.160	-9.253	-9.476
δE_{intra}	7.254	6.501	6.194	6.194	7.299	6.892	7.281	7.319
ΔE_c	-1.894	-2.469	-2.775	-1.264	-1.668	-0.268	-1.972	-2.156
E_{int}	-17.113	-17.689	-17.995	-16.483	-16.888	-15.487	-17.192	-17.376
				pyridine-methanol				
δE_{inter}	-4.624	-4.640	-4.640	-3.778	-4.634	-3.600	-4.698	-4.823
δE_{intra}	2.622	-2.758	2.017	2.018	2.669	2.416	2.641	2.666
ΔE_c	-2.002	-7.398	-2.623	-1.760	-1.966	-1.184	-2.057	-2.157
E_{int}	-6.226	-6.983	-6.847	-5.985	-6.189	-5.408	-6.281	-6.381
				guanine -cytosine (Watson Crick)				
δE_{inter}	-13.860	-13.749	-13.749	-11.345	-13.731	-10.854	-14.081	-14.445
δE_{intra}	10.570	9.667	8.880	8.881	10.673	10.002	10.624	10.690
ΔE_c	-3.290	-4.083	-4.869	-2.465	-3.058	-0.852	-3.457	-3.755
E_{int}	-27.729	-28.520	-29.307	-26.904	-27.497	-25.291	-27.897	-28.194

Table A.4: Intermolecular interaction energies and its components are compared on the level of LCCSD|X with X=LMP2 *uc* (for uncoupled), X=d-rCCD *uc*, X=LMP2, X=*direct*-LrCCD, X=LrCCD3, X=LrCCD4, to a full LCCSD calculation (all pairs strong) for the set of dispersion dominated dimers. Correlation contributions ΔE_c to the respective intermolecular interaction energies, and their intra- and intermolecular components δE_{intra}, and δE_{inter}, respectively, all in [kcal/mol]. The RMS and maximum devivations relative to the CCSD calculation for all dimers is given in tab.A.5. The prefix "L" for local was omitted in the method names, for brevity.

	CCSD	MP2	CCSD| MP2*uc*	CCSD| d-rCCD*uc*	CCSD| MP2	CCSD| d-rCCD	CCSD| rCCD3	CCSD| rCCD4
				dispersion dominated				
				pyridin-pyridine (π-π)				
δE_{inter}	-7.707	-9.150	-9.150	-6.486	-9.148	-6.060	-8.346	-8.795
δE_{intra}	2.388	0.567	0.592	0.593	2.857	1.953	2.536	2.671
ΔE_c	-5.319	-8.583	-8.558	-5.893	-6.291	-4.108	-5.810	-6.124
E_{int}	-1.976	-5.241	-5.215	-2.551	-2.948	-0.765	-2.467	-2.781
				pyridin-pyridine (TS)				
δE_{inter}	-5.538	-6.043	-6.043	-4.717	-6.040	-4.436	-5.841	-6.073
δE_{intra}	2.406	1.390	1.436	1.436	2.607	2.139	2.490	2.553
ΔE_c	-3.132	-4.653	-4.607	-3.281	-3.433	-2.297	-3.351	-3.520
E_{int}	-2.254	-3.777	-3.731	-2.405	-2.557	-1.421	-2.475	-2.644
				benzene-benzene(π-π)				
δE_{inter}	-7.035	-8.369	-8.369	-5.891	-8.368	-5.504	-7.672	-8.086
δE_{intra}	1.965	0.293	0.281	0.282	2.397	1.559	2.110	2.236
ΔE_c	-5.070	-8.076	-8.089	-5.609	-5.971	-3.946	-5.562	-5.850
E_{int}	-1.103	-4.110	-4.121	-1.643	-2.003	0.022	-1.594	-1.882
				benzene-benzene(TS)				
δE_{inter}	-5.172	-5.729	-5.729	-4.350	-5.724	-4.086	-5.527	-5.757
δE_{intra}	2.101	1.111	1.113	1.114	2.300	1.835	2.188	2.252
ΔE_c	-3.071	-4.618	-4.616	-3.236	-3.424	-2.251	-3.338	-3.505
E_{int}	-1.618	-3.163	-3.159	-1.781	-1.969	-0.796	-1.883	-2.050
				uracil-ethene				
δE_{inter}	-5.364	-5.817	-5.817	-4.283	-5.821	-3.994	-5.682	-5.898
δE_{intra}	1.897	0.949	0.849	0.849	2.054	1.576	1.958	2.014
ΔE_c	-3.467	-4.868	-4.968	-3.434	-3.766	-2.418	-3.724	-3.885
E_{int}	-2.056	-3.460	-3.560	-2.026	-2.357	-1.009	-2.316	-2.476
				uracil-neopentane				
δE_{inter}	-6.083	-6.315	-6.315	-4.718	-6.313	-4.418	-6.317	-6.525
δE_{intra}	1.895	0.881	0.856	0.856	1.994	1.541	1.932	1.979
ΔE_c	-4.188	-5.434	-5.459	-3.862	-4.318	-2.877	-4.384	-4.546
E_{int}	-2.105	-3.353	-3.377	-1.781	-2.237	-0.796	-2.304	-2.466
				guanine-cytosine (stacked)				
δE_{inter}	-18.237	-20.212	-20.212	-15.028	-20.212	-14.140	-18.727	-19.555
δE_{intra}	9.193	6.752	5.932	5.931	9.834	8.193	9.295	9.503
ΔE_c	-9.045	-13.460	-14.280	-9.098	-10.378	-5.947	-9.431	-10.053
E_{int}	-14.455	-18.875	-19.699	-14.514	-15.790	-11.360	-14.845	-15.465

Table A.5: Intermolecular interaction energies and its components are compared on the level of LCCSD|X with X=LMP2 *uc* (for uncoupled), X=d-LrCCD *uc*, X=LMP2, X=*direct*-LrCCD, X=LrCCD3, X=LrCCD4, to a full LCCSD calculation (all pairs strong) for the set of dimers not dominated by a particular interaction. Correlation contributions ΔE_c to the respective intermolecular interaction energies, and their intra- and intermolecular components δE_{intra}, and δE_{inter}, respectively, all in [kcal/mol]. The RMS and maximum devivations relative to the CCSD calculation for all dimers is given. The prefix "L" for local was omitted in the method names, for brevity.

	CCSD	MP2	CCSD| MP2*uc*	CCSD| d-rCCD*uc*	CCSD| MP2	CCSD| d-rCCD	CCSD| rCCD3	CCSD| rCCD4
				mixed				
uracil-uracil (π-π)								
δE_{inter}	-11.953	-13.078	-13.078	-9.569	-13.067	-8.972	-12.297	-12.809
δE_{intra}	4.867	2.844	2.695	2.695	5.202	4.141	4.919	5.042
ΔE_c	-7.086	-10.235	-10.383	-6.874	-7.865	-4.831	-7.378	-7.768
E_{int}	-6.684	-9.833	-9.982	-6.475	-7.464	-4.431	-6.977	-7.367
uracil-ethyne								
δE_{inter}	-5.072	-5.623	-5.623	-4.119	-5.628	-3.831	-5.379	-5.602
δE_{intra}	2.145	1.331	1.131	1.133	2.326	1.836	2.201	2.260
ΔE_c	-2.926	-4.292	-4.492	-2.986	-3.302	-1.995	-3.178	-3.343
E_{int}	-2.540	-3.905	-4.104	-2.601	-2.915	-1.608	-2.791	-2.955
uracil-pyridine (π-π)								
δE_{inter}	-10.090	-11.483	-11.483	-8.271	-11.483	-7.746	-10.602	-11.107
δE_{intra}	3.712	1.799	1.650	1.651	4.153	3.121	3.815	3.951
ΔE_c	-6.379	-9.684	-9.834	-6.621	-7.329	-4.625	-6.787	-7.157
E_{int}	-4.290	-7.597	-7.747	-4.535	-5.243	-2.539	-4.701	-5.070
RMS deviation								
δE_{inter}		0.900	0.900	1.613	0.899	2.021	0.364	0.719
δE_{intra}		1.921	1.545	1.545	0.293	0.467	0.079	0.171
ΔE_c		2.591	2.380	0.389	0.613	1.558	0.285	0.549
E_{int}		2.194	2.381	0.389	0.613	1.557	0.286	0.550
maximum deviation								
δE_{inter}		1.975	1.975	3.209	1.975	4.097	0.639	1.318
δE_{intra}		5.381	3.261	3.262	0.642	1.000	0.149	0.310
ΔE_c		5.397	5.235	0.825	1.333	3.097	0.492	1.008
E_{int}		4.420	5.244	0.825	1.335	3.095	0.491	1.010

Appendix A. Supplementary data for chapter 2

Table A.6: Deviations (in kcal/mol) of the correlation contributions to the interaction energies ΔE_c from the respective full LCCSD values, for the LCCSD|LMP2, LCCSD|LrCCD3 , LCCSD|LrCCD, LCCSD|LCCD-R^{-6}, and LCCSD|LCCD[S]-R^{-6} methods. The RMS and maximum deviations are also given.

Δ E$_c$ wrt LCCSD					
	LCCSD|				
	LMP2	LrCCD3	LrCCD	LCCD-R^{-6}	LCCD[S]-R^{-6}
Electrostatic dominated					
Water-Water	0.112	-0.002	-0.037	0.105	0.017
Peptide-Peptide	0.080	-0.108	-0.230	0.181	-0.013
Uracil-AcNH_2	0.226	-0.078	-0.262	0.381	-0.014
Pyridine-Methanol	0.036	-0.055	-0.155	0.170	0.022
Guanine-Cytosine (Watson Crick)	0.232	-0.167	-0.465	0.556	-0.019
Formamid-Formamid	0.259	-0.069	-0.224	0.338	-0.006
2-Pyridoxine - 2-Aminopyridine	0.035	-0.149	-0.391	0.378	-0.002
Formic Acid - Formic Acid	0.292	0.010	-0.186	0.492	0.019
Ammonia - Ammonia	0.085	-0.056	-0.089	0.050	0.008
Dispersion dominated					
Indole-Benzene (stacked)	-1.626	-0.673	-1.212	0.024	0.068
Pyridine-Pyridine($\pi-\pi$)	-0.971	-0.490	-0.804	-0.057	0.010
Benzene-Benzene($\pi-\pi$)	-0.900	-0.492	-0.780	-0.084	-0.004
Uracil-Ethene	-0.300	-0.258	-0.418	0.043	-0.011
Uracil-Neopentane	-0.130	-0.196	-0.358	0.120	-0.023
Guanine-Cytosine (stacked)	-1.333	-0.387	-1.008	0.487	0.035
Uracil-Uracil ($\pi-\pi$)	-0.779	-0.292	-0.682	0.298	-0.035
Uracil-Ethyne	-0.376	-0.252	-0.416	0.029	0.001
Uracil-Pyridine($\pi-\pi$)	-0.951	-0.409	-0.778	0.112	-0.021
Pyrazine-Pyrazine	-1.199	-0.538	-0.918	-0.023	0.049
Ethene-Ethene	0.031	-0.151	-0.196	-0.005	0.000
Mixed					
Pyridine-Pyridine (TS)	-0.302	-0.219	-0.388	0.079	0.077
Benzene-Benzene(TS)	-0.354	-0.268	-0.434	0.026	0.047
Indole - Benzene (TS)	-0.497	-0.301	-0.537	0.108	0.082
Phenol - Phenol	-0.173	-0.175	-0.356	0.176	0.018
Ethene - Ethine	-0.029	-0.104	-0.140	-0.004	0.015
δ_{RMS}(ΔE$_c$)	0.637	0.293	0.554	0.240	0.033
δ_{MAX}(ΔE$_c$)	1.626	0.673	1.212	0.556	0.082

Appendix A. Supplementary data for chapter 2

Table A.7: Deviations (in kcal/mol) of the correlation contributions to the interaction energies ΔE_c from the respective full LCCSD(T) values, for the LCCSD(T)|LMP2, LCCSD(T)|LrCCD3, LCCSD(T)|LrCCD, LCCSD(T)|LCCD-R^{-6}, and LCCSD(T)|LCCD[S]-R^{-6} methods. The RMS and maximum deviations are also given.

Δ E_c wrt LCCSD(T)					
	LCCSD|				
	LMP2	LrCCD3	LrCCD	LCCD-R^{-6}	LCCD[S]-R^{-6}
Electrostatic dominated					
Water-Water	0.133	-0.002	-0.044	0.124	0.036
Peptide-Peptide	0.091	-0.128	-0.278	0.227	0.033
Uracil-$AcNH_2$	0.259	-0.099	-0.329	0.475	0.080
Pyridine-Methanol	0.030	-0.071	-0.193	0.201	0.053
Guanine-Cytosine (Watson Crick)	0.258	-0.222	-0.597	0.693	0.118
Formamid-Formamid	0.312	-0.086	-0.279	0.424	0.080
2-Pyridoxine - 2-Aminopyridine	0.000	-0.196	-0.502	0.466	0.086
Formic Acid- Formic Acid	0.353	0.009	-0.234	0.607	0.134
Ammonia - Ammonia	0.099	-0.066	-0.105	0.061	0.019
Dispersion dominated					
Indole-Benzene (stacked)	2.130	-0.900	-1.599	0.016	0.059
Pyridine-Pyridine($\pi-\pi$)	-1.261	-0.640	-1.048	-0.071	-0.005
Benzene-Benzene($\pi-\pi$)	-1.165	-0.640	-1.014	-0.104	-0.023
Uracil-Ethene	-0.368	-0.327	-0.527	0.056	0.002
Uracil-Neopentane	-0.151	-0.243	-0.439	0.150	0.007
Guanine-Cytosine (stacked)	-1.661	-0.521	-1.302	0.612	0.160
Uracil-Uracil ($\pi-\pi$)	-0.935	-0.383	-0.867	0.384	0.051
Uracil-Ethyne	-0.469	-0.318	-0.524	0.039	0.012
Uracil-Pyridine($\pi-\pi$)	-1.199	-0.532	-1.003	0.148	0.016
Pyrazine-Pyrazine	-1.566	-0.709	-1.205	-0.030	0.041
Ethene-Ethene	0.033	-0.184	-0.239	-0.006	-0.001
Mixed					
Pyridine-Pyridine (TS)	-0.409	-0.290	-0.504	0.084	0.081
Benzene-Benzene(TS)	-0.474	-0.348	-0.559	0.024	0.045
Indole - Benzene (TS)	-0.661	-0.397	-0.698	0.120	0.095
Phenol - Phenol	-0.238	-0.227	-0.453	0.214	0.056
Ethene - Ethine	-0.042	-0.130	-0.174	-0.006	0.013
$\delta_{RMS}(\Delta E_c)$	0.816	0.384	0.716	0.299	0.067
$\delta_{MAX}(\Delta E_c)$	2.130	0.900	1.599	0.693	0.160

Table A.8: Computational details for the calculations on the T-shaped indole-benzene complex (total number of pairs: 703, number of intermolecular pairs: 330). In the hybrid calculations the intramolecular pairs may be strong or close (number of strong/close pairs: s=373, c=330). Timings for the transformation of the 3-external and 4-external integrals, for evaluating the individual groups of diagrams, and the time for converging the (hybrid) coupled cluster calculations are given in seconds. The file sizes of the 3-external and 4-external integral distributions are given in GBytes. The interaction energies E_{int} and their inter- and intramolecular components δE_{inter}, δE_{intra} are also provided (in kcal/mol). The calculations were performed on seven AMD opteron 6180 SE cores.

					LCCSD|		
part	diagrams[a]	full	LMP2	LrCCD3	LrCCD	LCCD-R^{-6}	LCCD[S]-R^{-6}
LMP2			78	75	59	63	63
LrCCD3	D1,D3,D6			87	90	86	86
LrCCD	D2,D4,D5,D7-9				422	377	377
LCCD-R^{-6}	D13-16					8	8
[S]	D18-21						0.4
sum parts			78	162	571	534	534
whole iteration		3404	1642	1822	2190	2215	2215
nr. of iterations		10	10	10	10	10	10
intergral transformation (3-ext)		5265	3949	3939	3981	4010	4010
intergral transformation (4-ext)		9281	3583	3550	3705	3759	3759
3-ext file size		90.646	66.618	66.618	66.618	66.618	66.618
4-ext file size		263.293	71.495	71.495	71.495	71.495	71.495
δE_{inter}		-7.440	-8.223	-7.849	-8.173	-7.348	-7.374
δE_{intra}		3.236	3.522	3.345	3.431	3.252	3.252
E_{int}		-3.909	-4.408	-4.211	-4.448	-3.803	-3.828

a) additional diagrams beyond LMP2.

The diagrams D1,D3,D6,D13-D16, D18 and D19 decay as R^{-6}.

The diagrams D3-D5, D7 and D8 decay either as R^{-6} or exponentially depending on the pair type of the corresponding amplitudes.

The diagram D9 decays as R^{-12} and the diagrams D20 and D21 decay exponentially.

Appendix B

Supplementary data for chapter 3

Figs. B.1, B.2 give the geometries of the Ar-graphane and H_2-graphane systems. The separation between the center of the asorbate and the midplane of graphane is 3.67Å. The geometries are given in the form of `CRYSTAL` inputs.[95;114] The used basis sets are also given in form of `CRYSTAL` inputs. Hydrogen's basis is given in fig.B.3, carbon in fig.B.4 and argon in fig. B.5. Argon's core was represented by the Stuttgart relativistic pseudopotential.[115] The black part of the basis set represents VTZ in the text. The blue parts represent the diffuse AOs, not used in the periodic HF, but adde later only for some of the correlated calculations, using the dual basis set procedure of ref.[101] This basis set is called AVTZ in the text. In the periodic HF calculations the following values were chosen for the integral prescreening tolerances TOLINTEG: 7 7 7 20 75.[95;114] In the HF calculations and for construction of WFs[98] and PAOs,[101] the 24x24 k-mesh was used. The PAO-domains contained two atoms fro bonding WFs (graphane and H_2) and one atom for the lone-pair WFs (argon) in the LMP2 and d-LrCCD computations. This represents a standard value of 0.985 for the Boughton-Pulay domain criterion.[37] In the periodic LMP2 computations local density fitting was used. The fit-domain contained 12 atoms.[103] The density fitting's auxiliary basis set was aug-cc-pVTZ-mp2fit of Weigend and coworkers.[109] In the LMP2 calculations, the intra-slab and intra-adsorbate pairs were re-

stricted by the interorbital distance of 6 Å, the inter-adsorbate-slab pairs by 12 Å. For the latter the two-electron integrals with the interorbital distance larger than 8 Å were computed by means of the multipole approximation. The fragment definition for the d-LrCCD|LMP2 potential energy curves implied a single adsorbate and the intra-slab cutoff distance $R_{\text{intra}}^{\text{fragment}}$ of 3 Å. The number of the inter-slab pairs for the fragment was obtained at the point 3.67 Å of the slab-adsorbate separation, using a the cutoff $R_{\text{inter}}^{\text{fragment}}$ of 6.9Å, and was then kept fixed for the other points of the potential curve. For the emb.-pairs, the diagonal-pair approximation was used.

```
SLAB
72
2.5382195
2
1 0.666666666667 0.333333333333  1.329934859345
6 0.666666666667 0.333333333333  0.2293833912644
SUPERCELL
-2 -1
 1  2
FRACTION
ATOMINSE
1
218  .3333333333  .66666667 3.67
END
```

Figure B.1: Input geometry of argon on graphane (for the adsorbate-slab separation of 3.67Å) according to the `Crystal` input syntax.[95;114]

```
SLAB
72
2.5382195
2
1 0.666666666667 0.333333333333  1.329934859345
6 0.666666666667 0.333333333333  0.2293833912644
SUPERCELL
-2 -1
 1  2
FRACTION
ATOMINSE
2
1  .3333333333  .66666667 3.29921687
1  .3333333333  .66666667 4.04078314
END
```

Figure B.2: Input geometry of H_2 on graphane (for the adsorbate-slab separation of 3.67Å) according to the `Crystal` input syntax.[95;114]

```
1 6
0 0 3 1. 1.
 34.061341 0.0060251978
 5.1235746 0.045021094
 1.1646626 0.20189726
0 0 1 0. 1.
 0.32723041 1.
0 0 1 0. 1.
 0.16000000 1.
0 2 1 0. 1.
      1.4070000             1.0000000
0 2 1 0. 1.
      0.3880000             1.0000000
0 3 1 0. 1.
      1.0570000             1.0000000


0 2 1 0. 1.
      0.1020000             1.0000000
0 3 1 0. 1.
      0.2470000             1.0000000
```

Figure B.3: Specification of the basis set for hydrogen atoms according to the `Crystal` input syntax.[114] The basis set without the blue part is denoted as VTZ. The blue part consists of diffuse orbitals, and the compete basis set is denoted as AVTZ. Addition of the blue part in the calculations was processed via the dual basis set scheme (the keyword GUESDUAL[114]). It also requires modification of the number-of-shells specification from 6 to 8 (the first line).[114]

```
6  12
0 0 5 2.0 1.00
       8506.03840         0.533736640E-03
       1275.73290         0.412502320E-02
       290.311870         0.211713370E-01
       82.0562000         0.824178600E-01
       26.4796410         0.240128580
0 0 1 2.0 1.00
       9.24145850          1.00000000
0 0 1 0.0 1.00
       3.36435300          1.00000000
0 0 1 0.0 1.00
      0.871741640          1.00000000
0 0 1 0.0 1.00
      0.500000000          1.00000000
0 0 1 0.0 1.00
      0.20          1.00000000
0 2 4 2.0 1.00
       34.7094960         0.533009740E-02
       7.95908830         0.358658140E-01
       2.37869720         0.142002990
      0.815400650         0.342031050
0 2 1 0.0 1.00
      0.50          1.00000000
0 2 1 0.0 1.00
      0.20          1.00000000
0 3 1  0.  1.1
  1.09700000        1.0000000000
0 3 1  0.  1.1
  0.31800000        1.0000000000
0 4 1  0.  1.1
  0.76100000        1.0000000000


0 3 1  0.  1.0
  0.10000000        1.0000000000
0 4 1  0.  1.0
  0.26800000        1.0000000000}
```

Figure B.4: Specification of the basis set for carbon atoms according to the `Crystal` input syntax.[114] The basis set without the blue part is denoted as VTZ. The blue part consists of diffuse orbitals, and the compete basis set is denoted as AVTZ. Addition of the blue part in the calculations was processed via the dual basis set scheme (the keyword GUESDUAL[114]). It also requires modification of the number-of-shells specification from 12 to 14 (the first line).[114]

```
218 13
INPUT
  8.   0   2   2   2   1 0
   10.2617210000000        68.6677880100000               0
   3.95272500000000        24.0427662900000               0
   5.39271400000000        27.7307633100000               0
   2.69996700000000        4.04545904000000               0
   8.08623500000000       -8.13747696000000               0
   4.01663200000000       -1.66452808000000               0
   5.20845900000000       -3.40009845000000               0
0  0  3  2.  1.
 174.669655 0.002587
 12.695768 0.062313
 2.917834 -1.042158
0  0  1  0.  1.
 0.670840 1.000000
0  0  1  0.  1.
 0.299112 1.000000
0  0  1  0.  1.
 0.131402 1.000000
0 2 3 6. 1.
 19.887221 0.023346
3.776172 -0.224621
 1.211516 1.137596
0 2 1 0. 1.
 0.538499 1.000000
0 2 1 0. 1.
 0.229575 1.000000
0 2 1 0. 1.
 0.095103 1.000000
0 3 1 0. 1.
 1.4509  1.000000
0 3 1 0. 1.
 0.4385 1.000000
0 3 1 0. 1.
 0.1325 1.000000
0 4 1 0. 1.
 0.9305 1.000000
0 4 1 0. 1.
 0.2800 1.000000
```

Figure B.5: Specification of the basis set for argon atoms according to the `Crystal` input syntax.[114] The same basis for argon was used in both VTZ and AVTZ calculations, as it already contains diffuse d- and f-orbitals.

Appendix C

Detailed description and manual for the interface of chapter 3

In this appendix chapter, firstly a description for the MOLPRO and CRYSCOR interface parts is given, focusing on the code details. Secondly a brief manual for a calculation with the presented interface for embedded local direct ring CCD is given, describing the keywords introduced.

C.1 Code description

C.1.1 Cryscor part

The interface on CRYSCOR side consists of a module. In case the interface is activated via keyword (see sec.C.2) , the first difference to a normal LMP2 calculation is is an additional bucket-sort of the 3-index integrals. During the LMP2 calculation the three index integrals are assembled as normally for LMP2 to use and sorted in direct space with a bucket-sort in the density fitting module. The interface part adds an additional bucket sort afterwards, where the previously assembled three index integrals are resorted and stored in a separate file for later cutting to the fragment framework. The resort is needed for making the auxiliary functions the

fastest, since the P-blocks of MOLPRO differ from CRYSCOR's. The LMP2 calculation then continues normally.
After the usual LMP2 calculation is done, the rest of the needed quantities is attended to. This is realized in a main routine calling one subroutine after the other each handling one of the below listed quantities, after establishing the cluster framework. For the establishment of the fragment framework, the CRYSCOR internal pairs are scanned for pairs matching the distance criteria R^{fragment}, explained in chap.3; with an optional scan for the emb-pairs. From that list a subset of atoms is created. This subset contains all atoms included in the pair domains of the fragments pair-list. Since CRYSCOR's WFs are labeled only inside the unitcell, a conversion matrix labeling the WFs for MOLPRO is built. Analogous this is done for the atoms, and WF-pairs. The conversion is made that the fragment's order matches CRYSCOR's.
For this cluster framework the external Fock matrix is cut out using the subroutine "constr_dom_mat" according to fragment pair-list. The overlap matrix is handled analogously. The internal Fock matrix is created by filling in the corresponding values through the conversion matrix. The two index integrals are handled by running over the cluster atoms and getting the corresponding values via the "get_JPQ_slice" subroutine. For three index ERIs, the subroutine runs over the previously stored integrals reading, for a specific WF and atom, the block of auxiliary functions for a specific auxiliary atom. Then checking if the block is present in the fragment, going through all translational symmetry possibilities, and if so saving it at the corresponding array spaces for MOLPRO. Since the 4-index ERIs and the amplitudes are already stored in matrices for every pair, the subroutine goes through fragment pair-list and saves every pair part in the corresponding MOLPRO way.
Also a MOLPRO input is created, specifying the fragment's geometry, fake basis, number of electrons, needed keywords for the specific method and the domain information (see sec.3.2.3).

The file handling between both programs is done via four Fortran files, "fort.177" until "fort.180". Since the I/O-routines of the two programs are not compatible, this is done through Fortran basic I/O commands in specifically written subroutines on both sides, so that MOLPRO's variable definitions are matched. The files are split into two categories:

- files fort.177 and fort.178. These two contain the offsets needed to read the needed quantity out of
- files fort.179 and fort.180. These two contain the actual values of needed quantity, e.g. the Fock matrix.

The first file of each category (files fort.177, fort.179) handle integers; the second (fort.178, fort.180) double precision.
In order for MOLPRO code to read the correct quantity, e.g. external Fock matrix, the correct offset is needed. The code can read the quantity using the corresponding number of the list below. For example to read the external Fock matrix, the number 2 is used, to get the correct offset out of fort.178, in order to read the external Fock matrix out of fort.180. In summary, the needed quantity can be read using just the number of the lists below as a variable.

- **Integers**,
  1. Number of strong pairs
  2. Number of close pairs
  3. Total number of pairs
  4. Molpro pairlist
  5. Pairtype
  6. Blocklengths for auxiliary functions
- **Double Precision**,
  1. Pairweights

2. External Fock matrix
3. Overlap matrix
4. Doubles amplitudes
5. Two index integrals
6. Internal Fock matrix
7. Three index integrals
8. Four index integrals
9. Reference energy
10. Cryscor MP2 energy
11. Fragment MP2 energy
12. Cryscor MP2 energy, for intra slab
13. Fragment MP2 energy, for intra slab
14. Cryscor MP2 energy, for intra adsorbate
15. Fragment MP2 energy, for intra adsorbate
16. Cryscor MP2 energy, for inter slab-adsorbate
17. Fragment MP2 energy, for inter slab-adsorbate
18. Leonard Jones energy

C.1.2 Molpro part

On the `MOLPRO` side first a dummy Hartree Fock calculation with one iteration is executed, afterwards the Fock matrices are overwritten. After the localisation, the pair-list is overwritten with the provided one. After the following dummy LMP2 computation, the amplitudes are replaced and used as a starting guess for the following d-LrCCD calculation.
After the generation of the three index ERIs the provided ones are read and resorted to introduce the correct auxiliary block length and stored in the corresponding order in `MOLPRO` three index ERI file.

The pair specific transformation matrices $\mathbf{W}^{[ij]}$, which transform from PAO to the pseudo-canonical basis are constructed from the provided external Fock and PAO overlap matrices. The overlap matrix is replaced before their construction. Before starting the d-LrCCD iterations the four index integrals are replaced by the periodic ones.
The two index integrals in `MOLPRO` are usually created anew every time needed, and are for the interface read instead.

C.2 Manual

In this section a brief manual is given. It is presented in the steps of a standard calculation:

1. The initial `CRYSTAL` calculation is to be done normally.
2. The usually following localisation is also done normally. (no `symmWF`)
3. The `CRYSCOR` input needs to include the following keywords:
   - `NOSYM12`, for switching of the permutational symmetry.
   - `CLUSCOR`, for activating the interface module.

   and optionally,
   - `MOG_DIST`, for specifying the fragment cut-off tolerances in bohr. The first number is the intra tolerance; the second the inter. If there are no "mol-atoms" the first number applies. The default values are 1Åand 3Å. (1Åis the default in a normal case)
     For example:
     ```
     MOG_DIST
     5.67 13.1
     ```
   - `NO2LAYR`, for restricting the emb.-pairs to diagonal ones.

- `NFITCEL`, for setting the value of included cells in the densitfy fitting atoms set for the interface. Because of translational symmetry this some times can get to a very large value, depending on the cut-off tolerances and the resulting fragment. The default value is 75. An example:
  `NFITCEL`
  `123`
- `LPairlst`, for specifying the array length of the fragment pairlist, because it cannot be known before hand. The default value is "10000". Example:
  `LPairlst`
  1000

4. The `CRYSCOR` calculation then produces four fortran files for the passed quantities, "fort.177 - fort.180", and the `MOLPRO` input. In the `MOLPRO` input the spaces in the geometry and domain specification need to be removed. And the `MP2FIT` basis used in `CRYSCOR` needs to be specified. The memory allocation might need adjustment, depending on the fragment. It is set to 1536 words on default.
5. The four Fortran files "fort.177-180" need to be copied to the scratch of the desired computer for `MOLPRO` to read. And the fragment calculation can now be started.

The following two figures are example inputs for `CRYSCOR` and the produced `MOLPRO` input. Note the additonal keywords:

- `cluscor=1`
- `clscr_noref=1`

used for activating the interface module in the `MOLPRO` input and restricting the dummy HF to one iteration.

```
READC14
KNET
24
MEMORY
14000
NOSYM12
CLUSCOR
NO2LAYR
MOG_DIST
5.6691 13.
NFITCEL
123
MOLATOMS
1
13
MOLPAIR
6. 6.
ENVPAIR
6. 6.
MOENPAIR
8. 12.
DOMDEF2
19
2 2 2 1 2 2 2 2 2 1 1 1 2 2 2 2 2 2 2
LENJONES
8.
LONTOL
0.0001
OPERTOL
0.001 10
DFITTING
DIRECT
G-AVTZ
NMINCENT
12
ENDDF
PRINPLOT
2
END
END
```

Figure C.1: Example input for CRYSCOR with keywords for activating the interface.

```
memory,1536,m
file,2,cac.wfu,new
gdirect
symmetry,nosym;orient,noorient,angstrom
GEOMETRY={
H1,, -0.732725, -1.269116,  1.329942
H2,,  3.663624, -1.269116,  1.329942
...
C72,,  5.129073,  8.883815, -0.229385
AR1,,  2.930899,  2.538233,  3.670019
}
basis={
 s,H ,   1.000000     ,   2.000000     ,   3.000000
 p,H ,   1.000000     ,   2.000000
 d,H ,   1.000000
 s,C ,   1.000000     ,   2.000000     ,   3.000000     ,   4.000000     ,
   5.000000     ,   6.000000
 p,C ,   1.000000     ,   2.000000     ,   3.000000
 d,C ,   1.000000     ,   2.000000
 f,C ,   1.000000
 s,AR,   1.000000     ,   2.000000     ,   3.000000     ,   4.000000
 p,AR,   1.000000     ,   2.000000     ,   3.000000     ,   4.000000
 d,AR,   1.000000     ,   2.000000     ,   3.000000
 f,AR,   1.000000     ,   2.000000
}
set,nelec=296
{hf,clscr_noref=1}
locali,boys
{df-lrpa
local,cluscor=1,nonorm=2;
local,rpa_exch=1;
local,rpa_exres=1;
local,rclose= 3.0000
local,rweak=0
local,rdist=0
local,rvdist=0
core,0;
domain,   1.1  ,C1 ,C65
...
domain, 148.1  ,C11 ,C59
dfit,basis=avtz/mp2fit}
```

Figure C.2: Example input for MOLPRO produced by CRYSCOR

Bibliography

[1] V. Fock, *Zeitschrift fuer Physik A Hadrons and Nuclei* **61**, 126 (1930).

[2] C. Pisani and R. Dovesi, *Int. J. Quantum Chem.* **17**, 501 (1980).

[3] R. N. Euwema, D. L. Wilhite, and G. T. Surratt, *Phys. Rev. B* **7**, 818 (1973).

[4] P. Pulay and S. Saebo, *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)* **69**, 357 (1986).

[5] C. Pisani, M. Busso, G. Capecchi, S. Casassa, R. Dovesi, L. Maschio, C. Zicovich-Wilson, and M. Schütz, *J. Chem. Phys.* **122**, 094113 (2005).

[6] L. Maschio, D. Usvyat, F. R. Manby, S. Casassa, C. Pisani, and M. Schütz, *Phys. Rev. B* **76**, 075101 (2007).

[7] E. J. Baerends, D. E. Ellis, and P. Ros, *Chem. Phys.* **2**, 41 (1973).

[8] J. L. Whitten, *J. Chem. Phys.* **58**, 4496 (1973).

[9] D. Bohm and D. Pines, *Phys. Rev.* **82**, 625 (1951).

[10] X. Ren, P. Rinke, C. Joas, and M. Scheffler, *J. Mater. Sci.* **47**, 7447 (2012).

[11] K. S. Singwi, M. P. Tosi, R. H. Land, and A. Sjölander, *Phys. Rev.* **176**, 589 (1968).

[12] F. Furche, *Phys. Rev. B* **64**, 195120 (2001).

[13] J. Harl and G. Kresse, *Phys. Rev. B* **77**, 045136 (2008).

[14] J. Harl and G. Kresse, *Phys. Rev. Lett.* **103**, 056401 (2009).

[15] L. Schimka, J. Harl, A. Stroppa, A. Grüneis, M. Marsman, F. Mittendorfer, and G. Kresse, *Nat. Mater.* **9**, 741 (2010).

[16] M. Rohlfing and T. Bredow, *Phys. Rev. Lett.* **101**, 266106 (2008).

[17] J. Paier, B. G. Janesko, T. M. Henderson, G. E. Scuseria, A. Grüneis, and G. Kresse, *J. Chem. Phys.* **132**, 094103 (2010).

[18] G. E. Scuseria, T. M. Henderson, and D. C. Sorensen, *J. Chem. Phys.* **129**, 231101 (2008).

[19] D. L. Freeman, *Phys. Rev. B* **15**, 5512 (1977).

[20] J. Toulouse, W. Zhu, A. Savin, G. Jansen, and J. G. Angyan, *J. Chem. Phys.* **135**, 084119 (2011).

[21] O. Masur, D. Usvyat, and M. Schütz, *J. Chem. Phys.* **139**, 164116 (2013).

[22] J. Řezáč, K. E. Riley, and P. Hobza, *J. Chem. Theory Comput.* **7**, 2427 (2011).

[23] P. Jurečka, J. Šponer, J. Černý, and P. Hobza, *Phys. Chem. Chem. Phys.* **8**, 1985 (2006).

[24] M. Schütz, O. Masur, and D. Usvyat, *J. Chem. Phys.* **140**, 244107 (2014).

[25] O. Masur, M. Schütz, L. Maschio, and D. Usvyat, *J. Chem. Theory Comput.* **12**, 5145, doi: 10.1021/acs.jctc.6b00651, (2016).

[26] S. Grimme and M. Parac, *ChemPhysChem* **3**, 292 (2003).

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

[28] S. Saebo and P. Pulay, *Annu. Rev. Phys. Chem.* **44**, 213 (1993).

[29] P. Pulay, *Chem. Phys. Lett.* **100**, 151 (1983).

[30] J. Pipek and P. G. Mezey, *J. Chem. Phys.* **90**, 4916 (1989).

[31] S. F. Boys, *Quantum Theory of Atoms, Molecules, and the Solid State, A Tribute to John C. Slater. Edited by Per-Olov Löwdin. New York: Academic Press* , 253 (1966).

[32] C. M. Zicovich-Wilson, R. Dovesii, and V. R. Saunders, *J. Chem. Phys.* **115**, 9708 (2001).

[33] N. Marzari and D. Vanderbilt, *Phys. Rev. B* **56**, 12847 (1997).

[34] M. Schütz, *J. Chem. Phys.* **113**, 9986 (2000).

[35] M. Schütz and H.-J. Werner, *J. Chem. Phys.* **114**, 661 (2001).

[36] H.-J. Werner and M. Schütz, *J. Chem. Phys.* **135**, 144116 (2011).

[37] J. W. Boughton and P. Pulay, *J. Comput. Chem.* **14**, 736 (1993).

[38] S. A. Kucharski and R. J. Bartlett, *Adv. Quant. Chem.* **18**, 218 (1986).

[39] M. Schütz and F. R. Manby, *Phys. Chem. Chem. Phys.* **5**, 3349 (2003).

[40] M. Schütz and H.-J. Werner, *Chem. Phys. Lett.* **318**, 370 (2000).

[41] M. Schütz, J. Yung, G. K. L. Chan, F. R. Manby, and H.-J. Werner, *J. Chem. Phys.* **138**, 054109 (2013).

[42] P. E. Maslen, A. Dutoi, M. S. Lee, Y. H. Shao, and M. Head-Gordron, *Mol. Phys* **103**, 425 (2005).

[43] W. Li, P. Piecuch, J. R. Gour, and S. Li, *J. Chem. Phys.* **131**, 114109 (2009).

[44] Z. Rolik and M. Kallay, *J. Chem. Phys.* **135**, 104111 (2011).

[45] Z. Rolik, L. Szegedy, I. Ladjanszki, B. Ladoczki, and M. Kallay, *J. Chem. Phys.* **139**, 094105 (2013).

[46] M. Schütz, *J. Chem. Phys.* **116**, 8772 (2002).

[47] D. Usvyat, K. Sadeghian, L. Maschio, and M. Schütz, *Phys. Rev. B* **86**, 045412 (2012).

[48] R. Sedlak, T. Janowski, M. Pitonik, J. Rezic, P. Pulay, and P. Hobza, *J. Chem. Theory Comput.* **9**, 3364 (2013).

[49] J. Paier, B. G. Janesko, T. M. Henderson, G. E. Scuseria, A. Grüneis, and G. Kresse, *J. Chem. Phys.* **132**, 094103 (2010).

[50] D. Kats and F. R. Manby, *J. Chem. Phys.* **138**, 144101 (2013).

[51] J. J. Shepherd, T. M. Henderson, and G. E. Scuseria, *J. Chem. Phys.* **140**, 124102 (2014).

[52] M. Schütz, G. Hetzer, and H.-J. Werner, *J. Chem. Phys.* **111**, 5691 (1999).

[53] P. Jurečka, J. Šponer, J. Černy, and P. Hobza, *Phys. Chem. Chem. Phys.* **8**, 1985 (2006).

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

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

[56] M. Schütz, G. Rauhut, and H.-J. Werner, *J. Phys. Chem. A,* **102**, 5997 (1998).

[57] D. Usvyat, K. Sadeghian, L. Maschio, and M. Schütz, *Phys. Rev. B* **86**, 045412 (2012).

[58] C. Pisani, L. Maschio, S. Casassa, M. Halo, M. Schütz, and D. Usvyat, *J. Comput. Chem.* **29**, 2113 (2008).

[59] A. Grüneis, M. Marsman, and G. Kresse, *J. Chem. Phys.* **133**, 074107 (2010).

[60] A. Grüneis, G. H. Booth, M. Marsman, J. Spencer, A. Alavi, and G. Kresse, *J. Chem. Theory Comput.* **7**, 2780 (2011).

[61] C. Pisani, M. Schütz, S. Casassa, D. Usvyat, L. Maschio, M. Lorenz, and A. Erba, *Phys. Chem. Chem. Phys.* **14**, 7615 (2012).

[62] G. Booth, A. Grüneis, G. Kresse, and A. Alavi, *Nature* **493**, 365 (2012).

[63] D. Usvyat, *J. Chem. Phys.* **139**, 194101 (2013).

[64] A. Grüneis, *Phys. Rev. Lett.* **115**, 066402 (2015).

[65] C. Pisani, M. Busso, G. Capecchi, S. Casassa, R. Dovesi, L. Maschio, C. Zicovich-Wilson, and M. Schütz, *J. Chem. Phys.* **122**, 094113 (2005).

[66] D. Usvyat, L. Maschio, F. R. Manby, S. Casassa, M. Schütz, and C. Pisani, *Phys. Rev. B* **76**, 075102 (2007).

[67] H. Stoll, *Phys. Rev. B* **46**, 6700 (1992).

[68] H. Stoll, *Chem. Phys. Lett.* **191**, 548 (1992).

[69] B. Paulus, *Phys. Rep.* **428**, 1 (2006).

[70] A. Hermann and P. Schwerdtfeger, *Phys. Rev. Lett.* **101**, 183005 (2008).

[71] D. Usvyat, *J. Chem. Phys.* **143**, 104704 (2015).

[72] J. Yang, W. Hu, D. Usvyat, D. Matthews, M. Schütz, and G. K.-L. Chan, *Science* **345**, 6197 (2014).

[73] R. Martinez-Casado, D. Usvyat, L. Maschio, G. Mallia, S. Casassa, J. Ellis, M. Schütz, and N. M. Harrison, *Phys. Rev. B* **89**, 205138 (2014).

[74] C. Müller and D. Usvyat, *J. Chem. Theory Comput.* **9**, 5590 (2013).

[75] P. J. Bygrave, N. L. Allan, and F. R. Manby, *J. Chem. Phys.* **137**, 164102 (2012).

[76] C. Müller, D. Usvyat, and H. Stoll, *Phys. Rev. B* **83**, 245136 (2011).

[77] F. R. Manby, D. Alfè, and M. J. Gillan, *Phys. Chem. Chem. Phys.* **8**, 5178 (2006).

[78] C. Tuma and J. Sauer., *Phys. Chem. Chem. Phys.* **8**, 3955 (2006).

[79] E. Voloshina, N. Gaston, and B. Paulus, *J. Chem. Phys.* **126**, 134115 (2007).

[80] L. Maschio, *J. Chem. Theory Comput.* **7**, 2818 (2011).

[81] A. Grüneis, *J. Chem. Phys.* **143**, 102817 (2015).

[82] G. J. O. Beran and K. Nanda, *J. Phys. Chem. Lett.* **1**, 3480 (2010).

[83] S. Wen, K. Nanda, Y. Huanga, and G. J. O. Beran, *Phys. Chem. Chem. Phys.* **14**, 7578 (2012).

[84] C. R. Taylor, P. J. Bygrave, J. N. Hart, N. L. Allan, and F. R. Manby, *Phys. Chem. Chem. Phys.* **14**, 7739 (2012).

[85] B. Herschend, M. Baudin, and K. Hermansson, *J. Chem. Phys.* **120**, 4939 (2004).

[86] U. Birkenheuer, P. Fulde, and H. Stoll, *Theor. Chem. Acc.* **116**, 398 (2006).

[87] C. Müller and K. Hermansson, *Surf. Sci.* **603**, 3329 (2009).

[88] C. Huang, M. Pavone, and E. A. Carter, *J. Chem. Phys.* **134**, 154110 (2011).

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

[90] G. Knizia and G. K. L. Chan, *Phys. Rev. Lett.* **109**, 186404 (2012).

[91] O. Sode, M. Keçili, K. Yagi, and S. Hirata, *J. Chem. Phys.* **138**, 074501 (2013).

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

[93] T. Lan, A. A. Kananenka, and D. Zgid, *J. Chem. Phys.* **143**, 241102 (2015).

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

[95] R. Dovesi, R. Orlando, A. Erba, C. M. Zicovich-Wilson, B. Civalleri, S. Casassa, L. Maschio, M. Ferrabone, M. D. L. Pierre, P. DÁrco, Y. Noël, M. Causa, M. Rerat, and B. Kirtman, *Int. J. Quantum Chem.* **114**, 1287 (2014).

[96] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, and M. Schütz, *WIREs: Comput. Mol. Sci.* **2**, 242 (2012).

[97] N. Marzari and D. Vanderbilt, *Phys. Rev. B* **56**, 12847 (1997).

[98] C. M. Zicovich-Wilson, R. Dovesi, and V. R. Saunders, *J. Chem. Phys.* **115**, 9708 (2001).

[99] V. Smirnov, R. Evarestov, and D. Usvyat, *Int. J. Quantum Chem.* **88**, 642 (2002).

[100] S. Casassa, C. M. Zicovich-Wilson, and C. Pisani, *Theor. Chem. Acc.* **116**, 726 (2006).

[101] D. Usvyat, L. Maschio, C. Pisani, and M. Schütz, *Z. Phys. Chem.* **224**, 441 (2010).

[102] R. Martinez-Casado, D. Usvyat, L. Maschio, G. Mallia, S. Casassa, J. Ellis, M. Schütz, and N. M. Harrison, *Phys. Chem. Chem. Phys.* **16**, 21106 (2014).

[103] M. Schütz, D. Usvyat, M. Lorenz, C. Pisani, L. Maschio, S. Casassa, and M. Halo, *Accurate Condensed-Phase Quantum Chemistry, Series: Computation in Chemistry*, volume 27, chapter Density fitting for correlated calculations in periodic systems, CRC Press, 2010.

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

[105] C. Hampel and H.-J. Werner, *J. Chem. Phys.* **104**, 6286 (1996).

[106] D. C. Elias, R. R. Nair, T. M. G. Mohiuddin, S. V. Morozov, P. Blake, M. P. Halsall, A. C. Ferrari, D. W. Boukhvalov, M. I. Katsnelson, A. K. Geim, and K. S. Novoselov, *Science* **323**, 610 (2009).

[107] H. Sahin, O. Leenaerts, S. K. Singh, and F. M. Peeters, *Arxiv preprint arXiv:1502.05804* (2015).

[108] L. Maschio, D. Usvyat, M. Schütz, and B. Civalleri, *J. Chem. Phys.* **132**, 134706 (2010).

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

[110] A. Kokalj, *Comp. Mater. Sci.* **28**, 155 (2003).

[111] A. Hesselmann, *J. Chem. Phys.* **128**, 144112 (2008).

[112] A. Szabo and N. S. Ostlund, *J. Chem. Phys.* **67**, 4351 (1977).

[113] G. Sansone, L. Maschio, D. Usvyat, M. Schütz, and A. Karttunen, *J. Phys. Chem. Lett.* **7**, 131 (2016).

[114] R. Dovesi, V. R. Saunders, R. Roetti, R. Orlando, C. M. Zicovich-Wilson, F. Pascale, B. Civalleri, K. Doll, N. M. Harrison, I. J. Bush, P. D. Arco, and M. Llunell, *CRYSTAL09 User's Manual*, University of Torino, 2009.

[115] A. Nicklass, M. Dolg, H. Stoll, and H. Preuss, *J. Chem. Phys.* **102**, 8942 (1995).