

Implementation of the Incremental Scheme for Highly Efficient Correlation Methods

Inaugural-Dissertation zur Erlangung des Doktorgrades der
Mathematisch-Naturwissenschaftlichen Fakultät
der Universität zu Köln

vorgelegt von
Joachim Friedrich
aus Leideneck

Köln, 08.05.2007

Berichterstatter: Prof. Dr. M. Dolg
Prof. Dr. U. Deiters
Tag der mündlichen Prüfung: 25.06.2007

Contents

1	Introduction	1
2	General Theory	5
2.1	The Many-Particle Problem in Quantum Mechanics	5
2.2	Hartree-Fock Theory	6
2.3	Multi-Configuration Hartree-Fock Theory	8
2.4	The Concept of Size-Extensivity/Size-Consistency	8
2.5	Configuration Interaction Theory	9
2.5.1	Multi-Reference Configuration Interaction Theory	10
2.6	Coupled Cluster Theory	10
2.7	Localization	11
2.8	Periodic Systems	15
2.8.1	Lattice Translation Vectors	15
2.9	Finite-Cluster Approach	16
2.10	Graph Theory	17
2.10.1	METIS Graph-Partitioning	17
3	The Incremental Scheme	19
3.1	The Incremental Expansion for a Molecule	19
3.1.1	Exactness of the Incremental Expansion	21
3.1.2	Approximations to the Incremental Scheme	23
3.1.3	Energy Screening	26
3.1.4	Formal Scaling	27
3.2	Derivation of the Correlation Energy for a Solid	28
3.3	Incremental Correlation Energy for Small Domains	29
3.3.1	Expansion of the Fragmental Energies in a Further Expansion	30
3.3.2	Separate Treatment of the Translational Symmetry	31
3.4	Incremental Expansion for non-Disjoint Cells	33
3.4.1	The Prefactors for Disjoint Cells	34
3.4.2	The Prefactors for non-Disjoint Cells	35

3.5	Incremental Expansion for a Multi-reference Case	36
3.5.1	Truncation of the Virtual Space	37
3.6	Treatment of Symmetry	37
3.6.1	Symmetric One-Site Domains	38
3.7	Error Analysis in the Incremental Expansion	39
4	Automatization of the Incremental Scheme	43
4.1	Obtaining Groups of Occupied Orbitals	43
4.1.1	Excitation Spaces for One-Site Domains	44
4.1.2	Construction of the n -Site Domains	45
4.1.3	Obtaining Correlation Energies	46
4.2	Truncation by Distance	46
5	Implementation	47
5.1	Interfaces	47
5.2	Foster-Boys Localization	47
5.3	Parallelization	48
5.3.1	The Server	48
5.3.2	The Client	49
5.3.3	The Wrapper	49
5.4	Visualization of the Domains	52
5.5	Symmetry for Molecules	53
5.5.1	Generating the Operators of the Point Group	54
5.5.2	Symmetry Adapted One-Site Domains	55
5.5.3	Symmetry Classified n -Site Domains	56
5.6	Symmetry for Periodic Systems	57
5.7	Algorithm to Obtain the Virtual Space	58
6	Applications	59
6.1	Molecules	59
6.1.1	Hydrocarbon Compounds	59
6.1.2	Transition Metal/Actinide Compounds	59
6.2	Intermolecular Interactions	68
6.2.1	Water Clusters	68
6.2.2	π - π /CH- π -Interactions	72
6.2.3	The Auophilic Attraction	74
6.2.4	DNA Base Pair	76
6.2.5	Reaction Pathways of the 4- <i>exo</i> /5- <i>endo</i> Cyclization	76
6.3	Symmetric Systems	81

6.3.1	A Water Cluster	81
6.3.2	Circular <i>Cis, Trans</i> -Cyclotriazine Cluster	86
6.4	Cluster Compounds	88
6.4.1	Boron Hydrides	88
6.4.2	Mercury Clusters	89
6.5	Potential Energy Surfaces	91
6.5.1	Octane	91
6.5.2	Hexayne	93
6.6	Polymers	96
6.7	Excited States	97
7	Symmetric Localization	101
8	Summary and Outlook	105
8.1	Summary	105
8.2	Outlook	106
	Literature	107
A	List of Abbreviations	117
B	The Incremental Code	119
B.1	Required Libraries	119
B.2	Quantum Chemistry Packages	120
B.3	Molpro Input Example	120
B.4	MOLCAS 6.4 Dependencies	120
B.5	Dalton Dependencies	120
C	Molecular Data	125
C.1	Exemplary Lists of Increments	125
C.2	Lists of Increments for CI-based Methods	127
C.3	4- <i>exo</i> /5- <i>endo</i> Cyclization	131

List of Figures

2.1	Parallelepiped spanned by the fundamental lattice vectors $\vec{a}_1, \vec{a}_2, \vec{a}_3$.	15
2.2	Example of a graph with 8 vertices.	17
2.3	The various phases of the multilevel graph bisection.	18
3.1	A system divided into four subsystems.	19
3.2	The four one-site increments in a system of 4 spatial parts.	20
3.3	All possible two-site increments within a space of four one-site increments.	20
3.4	A test system with 3 spatial parts 1, 2, 3.	22
3.5	Pictorial view of the involved sets, if the set of domains \mathbb{X} is partitioned into two sets \mathbb{A} and \mathbb{B} .	24
3.6	A 3×3 supercell built from a unit cell with atoms/centers of charge at the corners.	35
3.7	Error distribution of the incremental series for 10, 15 and 20 domains.	40
3.8	Error distribution of the incremental series with 20 domains using artificial order-dependent energies of ($\varepsilon_{\mathbb{X}} = X \cdot 10^{-10} \cdot 10^{\mathcal{O}}$).	42
5.1	A sketch of the required input data files for the incremental calculation	49
5.2	Data flow within the server/client structure.	50
5.3	Data flow for an incremental calculation within the MOLPRO environment.	51
5.4	Visualization of the concept of a wrapper class as data container for the network transfer.	51
5.5	Centers of charge for naphthalene colored by their membership to the one-site domains.	52
5.6	Centers of charge for naphthalene colored by their membership to the one-site domains.	53
6.1	RI-BP86/SVP optimized structures of some hydrocarbon molecules.	60
6.2	Optimized structures of TiCp_2Cl_2 , MoCl_6 , MoF_6 , $\text{Nb}_2\text{Cl}_{10}$, $[\text{Pu}(\text{H}_2\text{O})_8]^{3+}$	62
6.3	Centers of charge of MoF_6	64

6.4	Centers of charge of MoCl_6	64
6.5	RI-BP86/SVP optimized structure of Au_4Cl_4 11 (D_{2d}).	66
6.6	RI-BP86/SVP optimized structure of a set of eight water molecules.	68
6.7	$(\text{H}_2\text{O})_{11}$ cluster taken from Bulusu et al.	69
6.8	Sandwich structure of the benzene dimer at a benzene-benzene distance of 3.75 Å	73
6.9	Indole methane complex optimized by Ringer et al.	73
6.10	MP2 optimized structure of $\text{Au}_2(\text{PH-C}_2\text{H}_2\text{-S})_2$	75
6.11	RI-BP86/SVP optimized guanine-cytosine base pair.	76
6.12	Intermediates of the 4- <i>exo</i> /5- <i>endo</i> cyclization.	78
6.13	Convergence of the incremental correlation energies for intermediates of the 4- <i>exo</i> and 5- <i>endo</i> cyclization.	79
6.14	RI-BP86/SVP optimized structure of $(\text{H}_2\text{O})_6$ (C_3).	81
6.15	RI-BP86/SVP optimized structure of $(\text{N}_3\text{H}_3)_5$ (C_5).	86
6.16	RI-BP86/SVP optimized structures of B_5H_{11} 17 (C_1), Hg_{13} 18 (I_h), and Hg_{20} 19 (T_d).	88
6.17	RI-BP86/SVP optimized structure of n-octane with a fixed C4-C5 distance of 2.05 Å.	91
6.18	Potential energy scan along the C4-C5 distance of n-octane	92
6.19	Errors of the incremental correlation energy for n-octane.	92
6.20	RI-BP86/SVP optimized structure of dodeca-hexayne with a fixed C5-C6 distance of 1.55 Å.	93
6.21	Potential energy scan along the C5-C6 distance of dodeca-hexayne.	94
6.22	Errors of the incremental expansion of the CCSD correlation energy of hexayne.	95
6.23	Structure of the used hydrocarbon chain and the symmetric cell of $(\text{C}_2\text{H}_2)_n$	96
6.24	H_{28} test system	99
B.1	File dependencies for CI-based calculations within the MOLCAS 6.4 and DIESEL framework.	122
B.2	File dependencies for the incremental CCSD calculations within the DALTON framework.	123

List of Tables

6.1	Comparison of the incremental energies with the full CCSD calculations for the hydrocarbons in figure 6.1	61
6.2	Comparison of the incremental energies with the full CCSD calculations for a series of complexes	65
6.3	Comparison of the incremental energies with the full CCSD calculations for Au_4Cl_4	66
6.4	Incremental correlation Energy of $[\text{Pu}(\text{H}_2\text{O})_8]^{3+}$	66
6.5	Comparison of the convergence of the incremental energy of TiCp_2Cl_2 with respect to different dsp and t_{con} values	67
6.6	Comparison of the incremental energies for the $(\text{H}_2\text{O})_8$ -cluster	69
6.7	Convergence behavior of the incremental scheme with respect to the density parameter t_{dens}	70
6.8	Performance of the approximation of the incremental scheme with respect to a dynamic distance threshold	71
6.9	Performance of the approximation scheme eqn. 3.26	72
6.10	Performance of the dynamic distance threshold $R_{min}(\mathcal{O}_i)$	72
6.11	Comparison of the incremental energies with the full CCSD calculations for the benzene dimer and the methane indole complex	74
6.12	Comparison of the incremental energies with the full CCSD calculations for $\text{Au}_2(\text{PH-C}_2\text{H}_2\text{-S})_2$	75
6.13	Comparison of the incremental energies with the full CCSD calculations for the guanine-cytosine dimer	77
6.14	Performance of different quantum chemical methods for the 4- <i>exo</i> /5- <i>endo</i> cyclization.	80
6.15	Comparison of the error introduced by the incremental expansion and saving with respect to symmetry	82
6.16	Convergence of the incremental CCSD/6-31G** energies for a C_3 symmetric $(\text{H}_2\text{O})_6$ aggregate.	83
6.17	Convergence of the incremental CCSD/aug-cc-pVDZ energies for a C_3 symmetric $(\text{H}_2\text{O})_6$ aggregate.	84

6.18	Convergence of the incremental CCSD/cc-pVTZ energies for a C_3 symmetric $(H_2O)_6$ aggregate.	85
6.19	Convergence of the incremental scheme for a C_5 symmetric $(N_3H_3)_5$ aggregate.	87
6.20	Comparison of the incremental correlation energies with the full CCSD results for the B_5H_{11} -cluster.	89
6.21	Comparison of the incremental correlation energies with the full CCSD results for Hg clusters.	90
6.22	6-31G** incremental CCSD correlation energies per cell for <i>all-trans</i> polyacetylene.	96
6.23	6-31G** finite-cluster CCSD correlation energies per cell for <i>all-trans</i> polyacetylene.	97
6.24	Convergence behavior of incremental correlation energies for H_{28} , within the CI framework.	100
C.1	List of the incremental energies of MoF_6	125
C.2	List of the incremental energies of $MoCl_6$	126
C.3	List of the MR-CISD-incremental energies of H_{28}	127
C.4	List of the MR-ACPF-incremental energies of H_{28}	128
C.5	List of the MR-AQCC-incremental energies of H_{28}	129
C.6	List of the MR-CEPA(0)-incremental energies of H_{28}	130
C.7	Convergence behavior for the incremental RCCSD/6-31G** correlation energy of the intermediates of the cyclization of 14	131
C.8	Convergence behavior for the incremental RCCSD/6-31G** correlation energy of the intermediates of the cyclization of 14	132
C.9	Convergence behavior for the incremental RCCSD/6-31G** correlation energy of the intermediates of the cyclization of 14	133
C.10	Incremental correlation energies for different conformers on the PES of the 4- <i>exo</i> /5- <i>endo</i> cyclization of 14	134

Chapter 1

Introduction

The aim of quantum chemistry is to obtain a computational model for the qualitative and quantitative description of experimental observations on the basis of quantum mechanics. In order to do this one has to approximate the quantum mechanical many-body problem to a certain accuracy. A well established approximation is the Hartree-Fock (HF) approach, which is usually not sufficient to obtain chemical accuracy. The reason for this is the approximate averaged treatment of the electron repulsion in the HF approach. The electrons of different spin move independently of each other in the HF model. In reality all electrons avoid each other individually, which means that their movement is correlated. A way to include the correlation of the electrons is provided by density functional theory (DFT). DFT based methods are used today for a wide range of chemical systems and provide reliable results for many molecules. The major drawback of DFT is that it is not systematically improvable, since the Hohenberg-Kohn functional is unknown.

A common way to improve the HF wavefunction is to set up a many-body expansion on the basis of the HF orbitals. However, the application of these post HF methods is limited, since these approaches depend heavily on the size of the one-particle basis. The HF energy is invariant with respect to an arbitrary unitary transformation within the occupied space or within the virtual space. Therefore it is an interesting question to ask about a set of unitarily transformed orbitals where a many-body expansion based on configuration interaction (CI) or coupled cluster (CC) theory converges faster to the desired accuracy than for the canonical orbitals [1, 2]. An alternative to the canonical orbitals provide localized orbitals, which may be obtained efficiently by a Foster-Boys [3] or Pipek-Mezey [4] localization. Today we can find a variety of local correlation methods for the major quantum chemical methods. Møller-Plesset perturbation theory to second order (MP2) based correlation schemes were introduced by Pulay [5–8], Werner and Schütz [9, 10], Maslen and Head-Gordon [11–14], Ayala and Scuseria [15], Federov and Kitaura [16]. Local CI/multi-reference CI (MR-CI) ap-

proaches were developed by Walter et al. [17–19]. CC based local correlation methods were developed by Werner and Hampel [20], Werner and Schütz [21–24], Flocke and Bartlett [25], Subotnik and Head-Gordon [26, 27], Auer and Nooijen [28] and Christiansen et al. [29].

The extension of local correlation methods to solids is not yet solved completely. An early approach is the so-called local ansatz of Fulde and Stollhoff [30, 31]. MP2 became recently available in the CRYSTAL quantum chemistry package [32, 33] and a density fitting local MP2 was written by Usvyat and Schütz [34].

For higher level correlation methods it is possible to set up an incremental expansion of the correlation energy as introduced by H. Stoll [35–37]. The drawback of the incremental scheme is that up to now a lot of handwork is required to obtain the correlation energy. It was applied in a series of case studies for polymers [38–41], for solids [42–52], for molecules [53], for band structures [54–56] and for open-shell clusters [57]. A theoretical foundation of the incremental scheme was given by Fulde and Stoll [58, 59]. Conceptually similar to the incremental scheme is the divide and conquer approach [60] as well as other fragment based methods like the cluster in molecules (CIM) approach [61, 62], the molecular fractionation with conjugated caps (MFCC) [63], the generalized molecular fractionation with conjugate caps/molecular mechanics (GMFCC/MM) [64] or the systematic molecular fractionation [65].

The current work was done to reduce the scaling of multi-reference configuration interaction methods (MR-CI) and coupled cluster methods (CC) using the incremental scheme of Stoll and Nesbet [1, 35–37]. The main goals of this work are:

* Implementation

- generate a fully automatized procedure to obtain the incremental coupled cluster singles and doubles (CCSD) correlation energy, the open-shell restricted coupled cluster singles and doubles (RCCSD) correlation energy, the multi-reference configuration interaction singles and doubles (MR-CISD) correlation energy, the multi-reference averaged coupled pair functional (MR-ACPF) correlation energy, the multi-reference averaged quadratic coupled cluster (MR-AQCC) correlation energy and the multi-reference coupled electron pair approximation of zeroth order (MR-CEPA(0)) correlation energy
- generate a fully automated procedure to treat periodic systems within the framework of the incremental scheme

* Theory

- explore the potential accuracy of the approach using different correlation methods

- analyze the convergence behavior of the incremental scheme for molecules
- check the performance of the incremental scheme in combination with an approximate treatment of molecular symmetry

* Application

- molecules
- polymers
- excited states

Chapter 2

General Theory

2.1 The Many-Particle Problem in Quantum Mechanics

The quantum mechanical basis for many chemical problems beside relativistic effects is the time-independent Schrödinger equation:

$$\hat{H}\Psi = E\Psi \quad (2.1)$$

Here \hat{H} is the Hamilton operator, Ψ is the wave function and E is the energy eigenvalue. The non-relativistic electronic Hamiltonian for a molecule in Born-Oppenheimer approximation in atomic units is given as:

$$\hat{H} = -\frac{1}{2} \sum_i \nabla_i^2 - \sum_{\alpha} \sum_i \frac{Z_{\alpha}}{|\mathbf{R}_{\alpha} - \mathbf{r}_i|} + \sum_{i < j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{\alpha < \beta} \frac{Z_{\alpha} Z_{\beta}}{|\mathbf{R}_{\alpha} - \mathbf{R}_{\beta}|} \quad (2.2)$$

where \mathbf{r}_i are the coordinates of the electrons, \mathbf{R}_{α} are the coordinates of the nuclei and Z_{α} is the charge of the nucleus α . The first term describes the kinetic energy of the electrons, the second term corresponds to the interaction between the electrons and the nuclei, the third term is the electron-electron repulsion and the last term corresponds to the interaction between the nuclei. Unfortunately it is not possible to find an analytic solution for eqn. 2.1, for more than one electron in the potential of some nuclei. Therefore one relies on approximation schemes to obtain the energy with the desired accuracy.

The term causing the major complications in the electronic Hamiltonian eqn. 2.2 is the electron-electron repulsion. Unfortunately it is far too important to be neglected completely [66].

2.2 Hartree-Fock Theory

The Hartree-Fock approach is equivalent to the usage of a single Slater determinant Φ as approximation to the wavefunction Ψ in eqn. 2.1 [67] and the application of the variation principle in order to arrive at the optimal orbital set.

$$\Psi(x_1, x_2, \dots, x_{n-1}, x_n) = \Phi_0$$

$$\Phi_0 = \frac{1}{\sqrt{n!}} \begin{vmatrix} \chi_1(x_1) & \chi_2(x_1) & \cdots & \chi_n(x_1) \\ \chi_1(x_2) & \chi_2(x_2) & \cdots & \chi_n(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(x_{n-1}) & \chi_2(x_{n-1}) & \cdots & \chi_n(x_{n-1}) \\ \chi_1(x_n) & \chi_2(x_n) & \cdots & \chi_n(x_n) \end{vmatrix} \quad (2.3)$$

The spin-orbital $\chi(x)$ is a product of the spin function $\sigma(\omega)$ and the space function $\phi(\mathbf{r})$:

$$\chi_i(x) = \phi_i(\mathbf{r}) \cdot \sigma_i(\omega)$$

According to the variational principle we find the best approximation to the ground state energy for a single determinant wavefunction by:

$$\min[E(\{\chi_a\})] = \min \frac{\langle \Phi_0 | \hat{H} | \Phi_0 \rangle}{\langle \Phi_0 | \Phi_0 \rangle} \quad (2.4)$$

using the orthonormality of the spin orbitals

$$\langle \chi_a | \chi_b \rangle = \delta_{ab}$$

we have:

$$\langle \Phi_0 | \hat{H} | \Phi_0 \rangle = \sum_a \langle \chi_a | \hat{h} | \chi_a \rangle + \frac{1}{2} \sum_{ab} \langle \chi_a \chi_b | | \chi_a \chi_b \rangle \quad (2.5)$$

with

$$\hat{h}(i) = -\frac{1}{2} \nabla_i^2 - \sum_{\alpha} \frac{Z_{\alpha}}{|\mathbf{R}_{\alpha} - \mathbf{r}_i|}$$

$$\langle \chi_a \chi_b | | \chi_a \chi_b \rangle = \int dx_1 dx_2 \chi_a^*(1) \chi_b^*(2) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} (1 - \mathcal{P}_{12}) \chi_a(1) \chi_b(2)$$

$$\mathcal{P}_{12} \chi_a(1) \chi_b(2) = \chi_a(2) \chi_b(1)$$

The spin orbitals χ_a can be determined by the Fock equation:

$$\hat{f} | \chi_a \rangle = \varepsilon_a | \chi_a \rangle \quad (2.6)$$

with the Fock operator of the first particle $\hat{f}(1)$ being:

$$\hat{f}(1) = \hat{h}(1) + \sum_b \int dx_2 \chi_b^*(2) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} (1 - \mathcal{P}_{12}) \chi_b(2)$$

Since the Fock operator depends on the orbitals, we have to solve eqn. 2.6 in an iterative manner. The procedure to solve the HF equations is called self-consistent field method (SCF), which is often but not entirely correct synonymously used for the HF method. For a closed shell system we can use a restricted set of spin orbitals:

$$\begin{aligned} \chi_{2i}(x) &= \phi_{2i}(\mathbf{r}) \cdot \alpha(\omega) \\ \chi_{2i+1}(x) &= \phi_{2i}(\mathbf{r}) \cdot \beta(\omega) \end{aligned}$$

inserting the ansatz into the Fock equations and integrating out the spin functions, we get the Fock operator for a closed shell system:

$$\hat{f}(1) = \hat{h}(1) + \sum_b^{n/2} \int dr_2 \phi_b^*(2) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} (2 - \mathcal{P}_{12}) \phi_b(2) \quad (2.7)$$

The closed shell HF equations read:

$$\hat{f}|\phi_a\rangle = \varepsilon_a |\phi_a\rangle \quad (2.8)$$

In a system of n particles we use the $\frac{n}{2}$ lowest eigenvalues of eqn. 2.8 to build the Slater determinant Φ_0 . This is equivalent to divide the set of the orbitals $\{\phi_a\}$ into two disjoint subsets \mathbb{O} and \mathbb{V} . The set of the occupied orbitals \mathbb{O} is defined as:

$$\mathbb{O} = \bigcup_{\lambda=1}^{n/2} \phi_\lambda \text{ with } \varepsilon_{\lambda_1} < \varepsilon_{\lambda_2} \text{ for } \lambda_1 < \lambda_2$$

The virtual space is defined as:

$$\mathbb{V} = \{\phi_a\} \setminus \mathbb{O}$$

The introduction of a basis of N atomic orbitals for the spatial part of the spin orbitals according to eqn. 2.9:

$$\phi_a(\mathbf{r}) = \sum_{\lambda}^N c_{a\lambda} \phi_{\lambda}^{AO}(\mathbf{r}) \quad (2.9)$$

enables us to transform the Hartree-Fock equations into the Roothaan-Hall equations:

$$\mathbf{FC} = \mathbf{SC}\boldsymbol{\varepsilon} \quad (2.10)$$

Here \mathbf{S} is the overlap matrix of the atomic orbitals, \mathbf{F} is the Fock-matrix, \mathbf{C} is the coefficient matrix of the molecular orbitals and $\boldsymbol{\varepsilon}$ is the matrix of the eigenvalues (in the canonical case a diagonal matrix). Note that a unitary transformation U to the orbitals within the occupied space or within the virtual space has no effect on the total HF energy.

2.3 Multi-Configuration Hartree-Fock Theory

It is not always possible to build a fairly accurate wavefunction with a single determinant. Therefore in the multi-configuration self-consistent field method (MCSCF) a set of Slater determinants is used in the energy functional [67, 68].

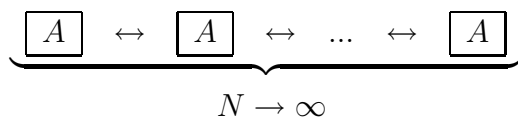
$$\min[E(\{\chi_a\}, \{c_i\})] = \min\left\langle \sum_i c_i \Phi_i | \hat{H} | \sum_i c_i \Phi_i \right\rangle \quad (2.11)$$

Where the set $\{\Phi_i\}$ is usually a small set of determinants with significant importance. A special case of the MCSCF is the complete active space SCF method (CASSCF). In this case all possible determinants which can be constructed from a certain number of (active) electrons and (active) orbitals are included in the functional eqn. 2.11.

2.4 The Concept of Size-Extensivity/Size-Consistency

In quantum chemistry one has to compare energies of different systems. Therefore it is very important that a given quantum chemical method yields an accuracy, independent of the size of the system under investigation. This is covered by the concept of size-extensivity [69–71] and the concept of size-consistency [72–74].

Consider a system of equal and possibly interacting subsystems A :



A correlation method is called size-extensive, if the correlation energy scales linearly with the size of the system:

$$\lim_{N \rightarrow \infty} \frac{E_{\text{corr}}(N \times A)}{N} = \text{const} > 0$$

This property guarantees the accuracy of a correlation method, independent of the total size of the system.

The term size-consistency was introduced for non-interacting subsystems A . If we consider a non interacting system of two parts A and B , we should be able to calculate the energy of the compound system by adding the two fragmental energies:

$$E_{AB} = E_A + E_B \quad (2.12)$$

If we can write the compound wavefunction $\hat{\psi}_{AB}|\text{vac}\rangle$ as a product of the separate wavefunctions $\hat{\psi}_A \hat{\psi}_B |\text{vac}\rangle$, we find that eqn. 2.12 holds [68]. Note that the antisymmetry is built into the wave operator $\hat{\psi}_X$ according to the anticommutation relations of the second quantized operators.

2.5 Configuration Interaction Theory

The HF and MCSCF wavefunctions can usually determine the energy of a molecule within an accuracy of 1%. Since this is not enough to treat chemical reactions, we have to go beyond these methods. The correlation energy is defined as:

$$E_{\text{corr}} = E_{\text{exact}} - E_{\text{HF}} \quad (2.13)$$

where E_{exact} is the non-relativistic energy in the limit of a complete one-particle basis and in the Born-Oppenheimer approximation. The optimal correlation energy for a given one-particle basis for all states can be obtained with the full configuration interaction method (FCI) [67, 68, 75, 76]. In the limit of a complete one-particle basis the FCI approach yields the exact correlation energy. In the FCI approach a linear combination of all possible determinants of the orbitals in \mathbb{O} and \mathbb{V} with $|\mathbb{O}|$ orbitals in every determinant is used to construct the wavefunction:

$$|\Psi_0^{\text{FCI}}\rangle = c_0|\Phi_0\rangle + \sum_i \sum_a c_i^a |\Phi_i^a\rangle + \sum_{ij} \sum_{ab} c_{ij}^{ab} |\Phi_{ij}^{ab}\rangle + \dots \quad (2.14)$$

$$i, j, \dots \in \mathbb{O}, \quad a, b, \dots \in \mathbb{V}$$

The coefficients for the determinants can be found by solving the eigenvalue equation:

$$\mathbf{HC} = \mathbf{CE} \quad (2.15)$$

where the \mathbf{H} is the Hamilton matrix, \mathbf{C} is the matrix with the expansion coefficients for all states and \mathbf{E} is the diagonal matrix with the energies of the states. $H_{st} = \langle \Phi_s | \hat{H} | \Phi_t \rangle$ is a matrix element between two arbitrary Slater determinants of eqn. 2.14. Due to the fast increase of the number of determinants in $|\Psi_0^{\text{FCI}}\rangle$ it is already impossible to use the FCI method for rather small molecules. The usual way to truncate the FCI wavefunction is to use only the most important classes of determinants. For instance we write the CI singles and doubles (CISD) wavefunction as:

$$|\Psi_0^{\text{CISD}}\rangle = c_0|\Phi_0\rangle + \sum_i \sum_a c_i^a |\Phi_i^a\rangle + \sum_{ij} \sum_{ab} c_{ij}^{ab} |\Phi_{ij}^{ab}\rangle \quad (2.16)$$

$$i, j, \dots \in \mathbb{O}, \quad a, b, \dots \in \mathbb{V}$$

There are three major problems associated with truncated CI wave functions: the lack of size-extensivity/consistency [68], the bad performance in the multi-reference case and the poor scaling behavior with respect to the one-particle basis (CISD $\propto \mathcal{N}^6$). The size-extensivity problem cannot be solved exactly within the framework of truncated CI theory, whereas it is very easy to construct a CI-wavefunction for a multi-reference case (vide infra).

2.5.1 Multi-Reference Configuration Interaction Theory

In order to get a proper multi-reference wavefunction within the framework of CI theory one uses a set of important determinants $\{\Phi_{act}\}$ and does single and double substitutions for every determinant of $\{\Phi_{act}\}$ (MR-CISD) [77]. Since the excitation-manifolds of the active determinants are not necessarily disjoint, one can construct the set of excited determinants by unification of excitation-manifolds for the active determinants. Due to the linear parameterization it is easily possible to obtain the required number of equations to determine the coefficients in the MR-CISD. The MR-CISD method yields very accurate energies for small molecules, but due to the lack of size-extensivity the results get poor for larger molecules.

2.6 Coupled Cluster Theory

Since the problem of size-consistency comes from the incapability to write a truncated CI wavefunction in direct product form, it is convenient to build the wavefunction in a way that the direct product form is always possible. This is done in the Coupled Cluster CC ansatz [68, 76]:

$$|\Psi_0^{CC}\rangle = \left[\prod_{\mu} (1 + t_{\mu} \hat{t}_{\mu}) \right] |\Psi_0^{HF}\rangle \quad (2.17)$$

where \hat{t}_{μ} is a general substitution operator. The \hat{t}_{μ} commute, if they are restricted to substitutions from the occupied HF-orbitals to the unoccupied HF-orbitals. Since the \hat{t}_{μ} are nilpotent [68], we can use the Taylor expansion of an exponential of an operator, for commuting \hat{t}_{μ} , to rewrite eqn. 2.17:

$$|\Psi_0^{CC}\rangle = \left[\prod_{\mu} e^{t_{\mu} \hat{t}_{\mu}} \right] |\Psi_0^{HF}\rangle = e^{\sum_{\mu} t_{\mu} \hat{t}_{\mu}} |\Psi_0^{HF}\rangle = e^{\hat{T}} |\Psi_0^{HF}\rangle \quad (2.18)$$

Analogous to the CI wavefunction we can truncate the cluster operator \hat{T} according to substitution classes:

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \dots + \hat{T}_N \quad (2.19)$$

with \hat{T}_n as:

$$\hat{T}_n = \left(\frac{1}{n!} \right)^2 \sum_{ij \dots ab \dots} t_{ij \dots ab \dots}^{ab \dots} \hat{a}_a^{\dagger} \hat{a}_b^{\dagger} \dots \hat{a}_j \hat{a}_i$$

where \hat{a}_a^{\dagger} are second quantized creation operators and \hat{a}_i are second quantized annihilation operators with $a, b, \dots \in \mathbb{V}$ and $i, j, \dots \in \mathbb{O}$ [78]. For the \hat{T}_n operator we have n summation indices for the occupied space and n summation indices for the virtual

space. Inserting the CC wavefunction of eqn. 2.17 into the Schrödinger equation and projecting with the HF state from the left we get:

$$\langle \Psi_0^{\text{HF}} | \hat{H} | \Psi_0^{\text{CC}} \rangle = E_{\text{CC}} \langle \Psi_0^{\text{HF}} | \Psi_0^{\text{CC}} \rangle = E_{\text{CC}} \quad (2.20)$$

Hereby we assumed intermediate normalization of the CC-wavefunction.

$$\langle \Psi_0^{\text{HF}} | \Psi_0^{\text{CC}} \rangle = 1$$

We get the equations for the CC amplitudes if we project with the excited determinants Φ^μ from the left:

$$\langle \Phi^\mu | \hat{H} | \Psi_0^{\text{CC}} \rangle = E_{\text{CC}} \langle \Phi^\mu | \Psi_0^{\text{CC}} \rangle \quad (2.21)$$

It is usually more convenient [78] to use the similarity transformed CC equations, where the Schrödinger equation is first multiplied by $e^{-\hat{T}}$ before the projection is done:

$$\begin{aligned} \langle \Psi_0^{\text{HF}} | e^{-\hat{T}} \hat{H} e^{\hat{T}} | \Psi_0^{\text{HF}} \rangle &= E_{\text{CC}} \\ \langle \Phi^\mu | e^{-\hat{T}} \hat{H} e^{\hat{T}} | \Psi_0^{\text{HF}} \rangle &= 0 \end{aligned} \quad (2.22)$$

For CCSD it can be shown that the energy in eqn. 2.20 is equal to the energy in eqn. 2.22 [68].

2.7 Localization

Local orbitals can be obtained by a unitary transformation of the canonical SCF orbitals. In order to keep the SCF energy invariant to these rotations we allow only rotations within the occupied space or the virtual space, respectively. Therefore we use in the closed shell case a unitary transformation of the form (in the full one-particle space):

$$U = \begin{pmatrix} u_{11} & \cdots & u_{1,n} & & & \\ \vdots & \ddots & \vdots & & & 0 \\ u_{n,1} & \cdots & u_{n,n} & & & \\ & & & u_{n+1,n+1} & \cdots & u_{n+1,n+m} \\ & & & \vdots & \ddots & \vdots \\ & & & u_{n+m,n+1} & \cdots & u_{n+m,n+m} \end{pmatrix} \quad (2.23)$$

$n :=$ number of occupied orbitals

$m :=$ number of virtual orbitals

$$N = m + n$$

According to the general considerations on products of unitary matrices in eqn. 2.24 we can construct a unitary matrix by forming a product of unitary matrices. This leads to a product Ansatz for our unitary transformation of the orbitals, where the complete transformation matrix is constructed as a product of 2×2 rotations.

$$\begin{aligned}
 UU^\dagger &= (U_1 \cdot U_2 \cdot \dots \cdot U_q) (U_1 \cdot U_2 \cdot \dots \cdot U_q)^\dagger \\
 &= (U_1 \cdot U_2 \cdot \dots \cdot U_{q-1}) \cdot U_q \cdot U_q^\dagger \cdot (U_1 \cdot U_2 \cdot \dots \cdot U_{q-1})^\dagger \\
 &= (U_1 \cdot U_2 \cdot \dots \cdot U_{q-1}) \cdot \mathbb{1} \cdot (U_1 \cdot U_2 \cdot \dots \cdot U_{q-1})^\dagger \\
 &= \vdots \\
 &= \mathbb{1}
 \end{aligned} \tag{2.24}$$

The corresponding unitary (orthogonal) $N \times N$ matrices U_k^{ij} are given as:

$$U_k^{ij} := \begin{cases} u_{ii} = u_{jj} = \cos(\gamma) \\ u_{ij} = -u_{ji}, \quad u_{ij} = \sin(\gamma) \\ u_{aa} = 1, \quad a \neq \{i, j\} \\ u_{ab} = 0, \quad a, b \neq \{i, j\} \end{cases} \tag{2.25}$$

e.g.

$$U_k^{1l} = \left(\begin{array}{ccc|ccc} \cos(\gamma) & & \sin(\gamma) & & & \\ & 1 & & & & \\ & & \ddots & & & \\ -\sin(\gamma) & & & 1 & & 0 \\ & & & & \cos(\gamma) & \\ & & & & & 1 \\ & & & & & \ddots \\ \hline & & & & & 0 \\ & & & & & \mathbb{1} \end{array} \right)$$

Since the HF energy is invariant to any unitary transformation of the type in eqn. 2.23 it is not possible to construct a unique localization criterion. Since the most localization

procedures use a functional like eqn. 2.26:

$$D(\phi) = \sum_i \langle \phi_i \phi_i | \hat{O} | \phi_i \phi_i \rangle, \quad (2.26)$$

we start from eqn. 2.26 to obtain a more general form. Now we transform the orbitals a and b of eqn. 2.26 by a 2×2 rotation according to eqn. 2.25 and obtain our functional with respect to the angle γ and the orbital pair a and b :

$$\begin{aligned} D(U_k^{ab}) &= \langle \phi_a^2 | \hat{O} | \phi_a^2 \rangle + \langle \phi_b^2 | \hat{O} | \phi_b^2 \rangle - 2 \langle \phi_a^2 | \hat{O} | \phi_a^2 \rangle \sin^2(\gamma) \\ &+ 2 \langle \phi_b^2 | \hat{O} | \phi_a^2 \rangle \sin^2(\gamma) + 2 \langle \phi_a^2 | \hat{O} | \phi_b^2 \rangle \sin^2(\gamma) - 2 \langle \phi_b^2 | \hat{O} | \phi_b^2 \rangle \sin^2(\gamma) \\ &+ 2 \langle \phi_a^2 | \hat{O} | \phi_a^2 \rangle \sin^4(\gamma) - 2 \langle \phi_b^2 | \hat{O} | \phi_a^2 \rangle \sin^4(\gamma) - 2 \langle \phi_a^2 | \hat{O} | \phi_b^2 \rangle \sin^4(\gamma) \\ &+ 2 \langle \phi_b^2 | \hat{O} | \phi_b^2 \rangle \sin^4(\gamma) + \langle \phi_a \phi_b | \hat{O} | \phi_a^2 \rangle \sin(2\gamma) \\ &+ \langle \phi_a^2 | \hat{O} | \phi_a \phi_b \rangle \sin(2\gamma) - \langle \phi_b^2 | \hat{O} | \phi_a \phi_b \rangle \sin(2\gamma) \\ &- \langle \phi_a \phi_b | \hat{O} | \phi_b^2 \rangle \sin(2\gamma) - 2 \langle \phi_a \phi_b | \hat{O} | \phi_a^2 \rangle \sin^2(\gamma) \sin(2\gamma) \\ &- 2 \langle \phi_a^2 | \hat{O} | \phi_a \phi_b \rangle \sin^2(\gamma) \sin(2\gamma) + 2 \langle \phi_b^2 | \hat{O} | \phi_a \phi_b \rangle \sin^2(\gamma) \sin(2\gamma) \\ &+ 2 \langle \phi_a \phi_b | \hat{O} | \phi_b^2 \rangle \sin^2(\gamma) \sin(2\gamma) + 2 \langle \phi_a \phi_b | \hat{O} | \phi_a \phi_b \rangle \sin^2(2\gamma) \\ &+ \sum_{i \neq a, b} \langle \phi_i \phi_i | \hat{O} | \phi_i \phi_i \rangle \end{aligned} \quad (2.27)$$

In the next step we use the addition theorem of trigonometric functions until we have only linear terms of $\cos(4\gamma)$ and $\sin(4\gamma)$. Now we factor out $\cos(4\gamma)$ and $\sin(4\gamma)$ and define:

$$A_{ab} = \langle \phi_a \phi_b | \hat{O} | \phi_a \phi_b \rangle - \frac{1}{4} \langle \phi_a^2 - \phi_b^2 | \hat{O} | \phi_a^2 - \phi_b^2 \rangle \quad (2.28)$$

$$B_{ab} = \langle \phi_a \phi_b | \hat{O} | \phi_a^2 - \phi_b^2 \rangle \quad (2.29)$$

Finally we obtain:

$$D(U_k^{ab}) = A_{ab} - A_{ab} \cos(4\gamma) + B_{ab} \sin(4\gamma) + \sum_i \langle \phi_i \phi_i | \hat{O} | \phi_i \phi_i \rangle \quad (2.30)$$

Eqn. 2.30 can be simplified to an expression with a single trigonometric function where the maxima and minima are given analytically. In order to do this we define α by:

$$\begin{aligned} \frac{B_{ab}}{\tan(4\alpha)} &= -A_{ab} \\ \sin(4\alpha) &= \frac{B_{ab}}{\sqrt{A_{ab}^2 + B_{ab}^2}} \\ \cos(4\alpha) &= -\frac{A_{ab}}{\sqrt{A_{ab}^2 + B_{ab}^2}} \end{aligned}$$

Next we substitute the prefactor of the cosine A_{ab} in eqn. 2.30 and multiply the term with the sine function by $\frac{\sin(4\alpha)}{\sin(4\alpha)}$:

$$D(U_k^{ab}) = A_{ab} + \sum_i \langle \phi_i \phi_i | \hat{O} | \phi_i \phi_i \rangle + \frac{B_{ab}}{\sin(4\alpha)} [\cos(4\gamma) \cos(4\alpha) + \sin(4\gamma) \sin(4\alpha)] \quad (2.31)$$

$$\cos(\alpha - \beta) = \cos(\alpha) \cos(\beta) + \sin(\alpha) \sin(\beta) \quad (2.32)$$

Using the identity 2.32 and the definition of $\sin(4\alpha)$ we simplify our functional to:

$$D(U_k^{ab}) = A_{ab} + \sqrt{A_{ab}^2 + B_{ab}^2} \cos(4\gamma - 4\alpha) + \sum_i \langle \phi_i \phi_i | \hat{O} | \phi_i \phi_i \rangle \quad (2.33)$$

The functional in eqn. 2.33 has a maximum if the cosine is 1 and a minimum if the cosine is -1 :

$$\begin{aligned} \gamma_{\max} &= \alpha, & \alpha + \frac{1}{2}\pi, & \alpha + \pi, & \alpha + \frac{3}{2}\pi \\ \gamma_{\min} &= \alpha + \frac{1}{4}\pi, & \alpha + \frac{3}{4}\pi, & \alpha + \frac{5}{4}\pi, & \alpha + \frac{7}{4}\pi \end{aligned}$$

Now we build the matrix $D(U_k^{ab})_{\max}$ in order to see where the change of the functional due to a unitary rotation is maximal.

$$D(U_k^{ab})_{\max} = D(U_k^{ab}) - D(\phi) = A_{ab} + \sqrt{A_{ab}^2 + B_{ab}^2} \quad (2.34a)$$

$$D(U_k^{ab})_{\min} = D(U_k^{ab}) - D(\phi) = A_{ab} - \sqrt{A_{ab}^2 + B_{ab}^2} \quad (2.34b)$$

We obtain the orbital pair for which the unitary transformation leads to the maximal change from the matrix $D(U_k^{ab})_{\max}$. After application of this rotation to the orbitals we build $D(U_k^{ab})_{\max}$ again with the new orbitals. We iterate until all matrix elements in $D(U_k^{ab})_{\max}$ are lower than a given threshold.

The most popular criteria are Foster-Boys [3] and Pipek-Mezey [4], because of their \mathcal{N}^3 scaling with respect to the one-particle basis set. The Foster-Boys criterion is designed to minimize the distance of two electrons in the same orbital, which is equivalent to maximize the distance of the orbital centroids.

$$\begin{aligned} D(\phi) &= \sum_i \langle \phi_i \phi_i | (\mathbf{r}_1 - \mathbf{r}_2)^2 | \phi_i \phi_i \rangle = \sum_i \langle \phi_i \phi_i | \mathbf{r}_1^2 - 2\mathbf{r}_1 \mathbf{r}_2 + \mathbf{r}_2^2 | \phi_i \phi_i \rangle \\ &= 2 \sum_i \langle \phi_i | \mathbf{r}_1^2 | \phi_i \rangle - 2 \sum_i \langle \phi_i | \mathbf{r}_1 | \phi_i \rangle \langle \phi_i | \mathbf{r}_2 | \phi_i \rangle \\ &= 2 \sum_i \langle \phi_i | \mathbf{r}_1^2 | \phi_i \rangle - 2 \sum_i \langle \phi_i | \mathbf{r}_1 | \phi_i \rangle^2 \end{aligned} \quad (2.35)$$

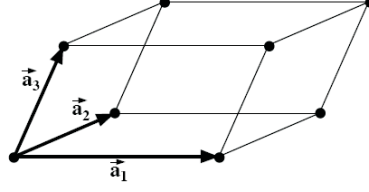


Figure 2.1: Parallelepiped spanned by the fundamental lattice vectors $\vec{a}_1, \vec{a}_2, \vec{a}_3$.

Since the first term is invariant to unitary transformations of the orbitals ϕ_i we can equivalently maximize the second term:

$$\max \left(\sum_i \langle \phi_i | \mathbf{r}_1 | \phi_i \rangle^2 \right) \quad (2.36)$$

For the Foster-Boys criterion we find:

$$A_{ab} = \langle \phi_a | \mathbf{r} | \phi_b \rangle^2 - \frac{1}{4} [\langle \phi_a | \mathbf{r} | \phi_a \rangle - \langle \phi_b | \mathbf{r} | \phi_b \rangle]^2 \quad (2.37a)$$

$$B_{ab} = \langle \phi_a | \mathbf{r} | \phi_b \rangle [\langle \phi_a | \mathbf{r} | \phi_a \rangle - \langle \phi_b | \mathbf{r} | \phi_b \rangle] \quad (2.37b)$$

We maximize the functional 2.33 by choosing the angle γ according to the suggestion of Edmiston and Ruedenberg [79] to be between 0 and 0.5π .

2.8 Periodic Systems

2.8.1 Lattice Translation Vectors

A 3-dimensional periodic array of points is completely defined by the three fundamental basis vectors $\vec{a}_1, \vec{a}_2, \vec{a}_3$ of the unit cell and the points within the reference cell \mathcal{C}_0 [80]. Further we define the cell \mathcal{C}_0 as parallelepiped spanned by the fundamental lattice vectors $\vec{a}_1, \vec{a}_2, \vec{a}_3$ at the origin $\vec{0}$.

$$\mathcal{C}_0 \subset \mathbb{R}^3$$

$$\mathcal{C}_0 := \{ \vec{r} \in \mathbb{R}^3 | \vec{r} = \alpha_1 \vec{a}_1 + \alpha_2 \vec{a}_2 + \alpha_3 \vec{a}_3 \text{ with } \alpha_1, \alpha_2, \alpha_3 \in [0, 1] \} \quad (2.38)$$

All other points of the crystal can be reached by a linear combination of the basis vectors:

$$\vec{r}_i = \vec{r}_0 + u_1 \vec{a}_1 + u_2 \vec{a}_2 + u_3 \vec{a}_3 \text{ with } u_1, u_2, u_3 \in \mathbb{Z} \quad (2.39)$$

where \vec{r}_0 is an arbitrary point in the reference cell \mathcal{C}_0 and \vec{r}_i is the translationally equivalent point in the cell \mathcal{C}_i . For reasons of convenience we introduce the set of lattice translation vectors \mathbb{T} as:

$$\mathbb{T} = \{ \vec{v} \in \mathbb{R}^3 | \vec{v} = u_1 \vec{a}_1 + u_2 \vec{a}_2 + u_3 \vec{a}_3 \text{ with } u_1, u_2, u_3 \in \mathbb{Z} \} \quad (2.40)$$

Since we have a unique translation operator for every translation vector in \mathbb{T} we introduce the set of translational operators $\hat{\mathbb{T}}$, where the translation operators are defined by their corresponding translational vector.

$$\hat{\mathbb{T}} = \{\hat{t}_{\vec{v}} | \vec{v} \in \mathbb{T}\} \quad (2.41)$$

A cell \mathcal{C}_λ is obtained by shifting the origin of \mathcal{C}_0 with the translation \hat{t}_λ :

$$\mathcal{C}_\lambda = \hat{t}_\lambda \mathcal{C}_0 \quad (2.42)$$

With this definition and eqn. 2.41 we can decompose \mathbb{R}^3 into a set of cells:

$$\mathbb{R}^3 = \bigcup_{\lambda} \hat{t}_\lambda \mathcal{C}_0 \quad (2.43)$$

According to the definition of the cells in eqn. 2.38 we see that neighboring cells are not disjoint. The intersection of the two sets is the boundary region. In order to avoid a multiple counting of the points on the borders, we introduce the weight factors of $\frac{1}{2}$ for a face centered point, $\frac{1}{4}$ for an edge centered point and $\frac{1}{8}$ for a corner point in a 3-dimensional lattice.

2.9 Finite-Cluster Approach

The total energy per cell E^{cell} of a polymer can be obtained by the difference of the energies of a supercell \mathcal{U}_n of n unit cells and a supercell \mathcal{U}_{n+1} of $n + 1$ unit cells. In the limit of an infinite n one obtains the exact energy per cell [38–40, 49, 81]. In order to avoid open-shell calculations, the dangling bonds are saturated by hydrogen atoms [38–40, 49, 81].

$$E^{\text{cell}} = \lim_{n \rightarrow \infty} [E(R\mathcal{U}_{n+1}R') - E(R\mathcal{U}_nR')] \quad (2.44)$$

In the difference the contributions from the saturating rests R and R' cancel each other for $n \rightarrow \infty$ and one obtains the desired result, the energy per cell. The expansion in eqn. 2.44 holds equally for the correlation energy of the polymer:

$$E_{\text{corr}}^{\text{cell}} = \lim_{n \rightarrow \infty} [E_{\text{corr}}(R\mathcal{U}_{n+1}R') - E_{\text{corr}}(R\mathcal{U}_nR')] \quad (2.45)$$

Eqn. 2.45 is very useful to obtain a benchmark correlation energy for the incremental correlation energy.

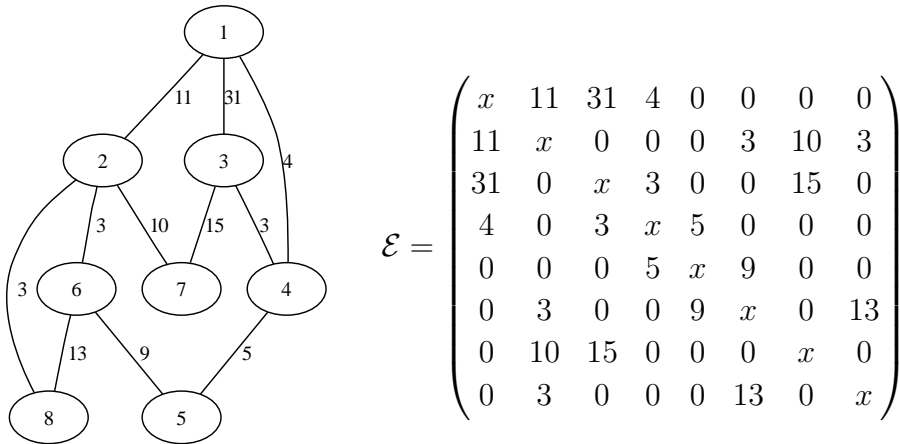


Figure 2.2: Example of a graph with 8 vertices and the representation matrix \mathcal{E} .

2.10 Graph Theory

An undirected Graph \mathcal{G} is determined by a set of vertices \mathbb{V} and a set of edges \mathbb{E} . The set \mathbb{E} is defined as a set of pairs of vertices. In the case of an edge-weighted unordered graph we may map every pair to an integer number as weight. Therefore we can represent \mathbb{E} by the adjacency matrix \mathcal{E} . The matrix element of \mathcal{E}_{ij} is zero iff the two corresponding vertices i and j are not connected (see figure 2.2). The numerical value of the element may represent the importance of the pair i and j . If it is possible to establish a path from any vertex to any other vertex of a graph, the graph is said to be connected; otherwise, the graph is disconnected [82].

2.10.1 METIS Graph-Partitioning [83]

A common problem in computer science is to partition the graph $\mathcal{G}(\mathbb{V}, \mathbb{E})$ into k disjoint subsets with the side condition that the sum of cut edge weights is minimal. Since the number of combinatorial possibilities grows very fast with the number of edges it is very convenient to transform the initial graph into a sequence of smaller graphs (coarsening). In the next step a sequence of bisections is performed until the desired number of parts is reached (initial partitioning). At the end the original graph is reconstructed (uncoarsening) in a stepwise fashion where the partitions are refined at each step in order to get a better result. Figure 2.3 demonstrates the various phases for a bisection of a graph in the METIS graph partitioning scheme.

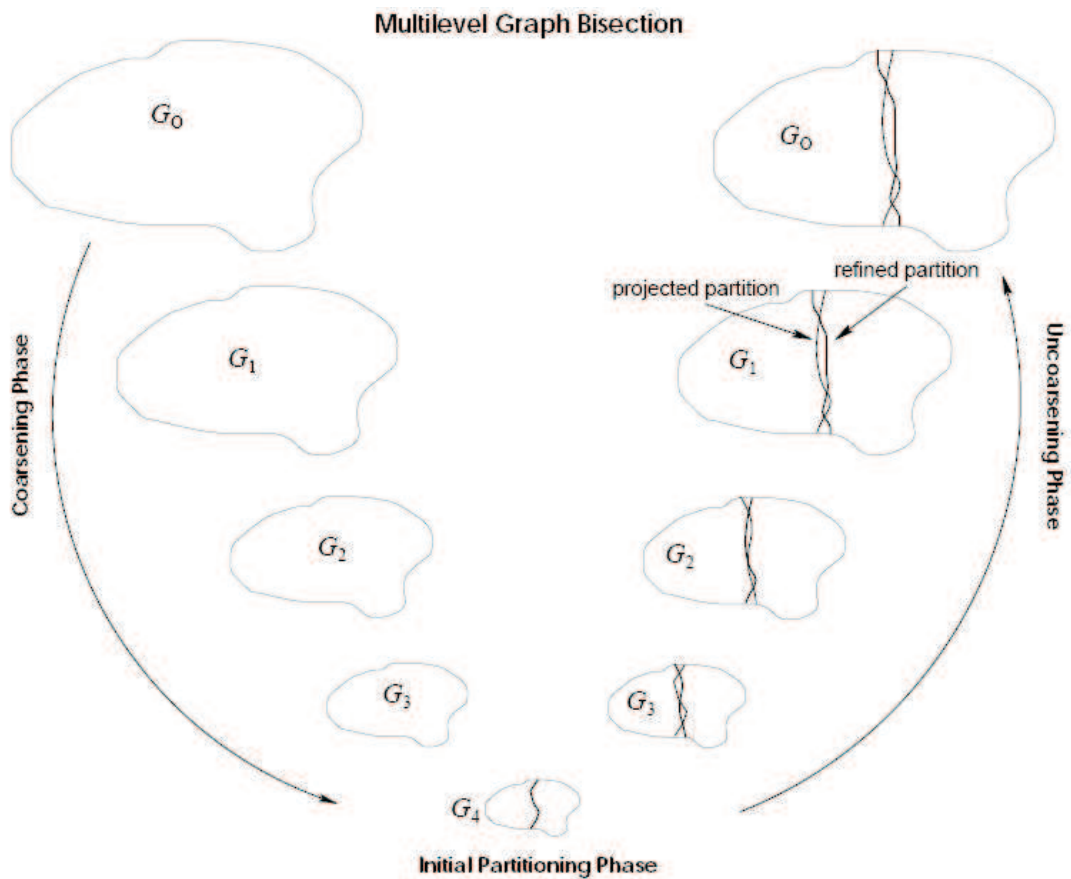


Figure 2.3: The various phases of the multilevel graph bisection. During the coarsening phase, the size of the graph is successively decreased; during the initial partitioning phase, a bisection of the smaller graph is computed; and during the uncoarsening phase, the bisection is successively refined as it is projected to the larger graphs. During the uncoarsening phase the light lines indicate projected partitions, and dark lines indicate partitions that were produced after refinement (figure and caption from [83]).

Chapter 3

The Incremental Scheme

3.1 The Incremental Expansion for a Molecule

Using localized molecular orbitals ϕ_w^{LMO} , we can calculate the correlation energy in the following way:

1. Divide the system under investigation into spatial parts and classify the ϕ_w^{LMO} according to these parts; e.g. in the case of four parts:

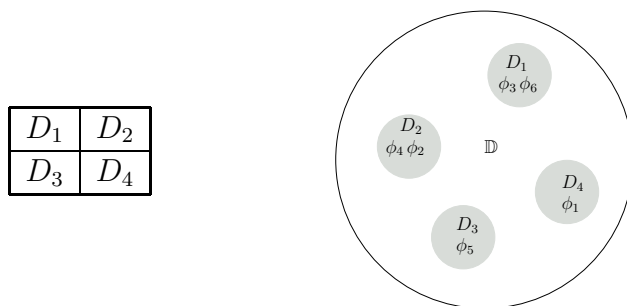


Figure 3.1: A system divided into four subsystems.

This yields a set of localized orbitals for every part. In this case we have the set of domains $\mathbb{D} = \{D_1, D_2, D_3, D_4\}$.

2. Set up the correlation energy as:

$$E_{\text{corr}} = \sum_i \Delta\varepsilon_i + \frac{1}{2!} \sum_{ij} \Delta\varepsilon_{ij} + \frac{1}{3!} \sum_{ijk} \Delta\varepsilon_{ijk} + \dots \quad (3.1)$$

with $i, j, k, \dots \in \mathbb{D}$, we require that i, j, k, \dots are pairwise disjoint and define the one-site increments as:

$$\Delta\varepsilon_i = \Delta\varepsilon(\{i\}) = \varepsilon_i = E_{\text{corr}}(i) \quad (3.2)$$

D_1	D_2
D_3	D_4

D_1	D_2
D_3	D_4

D_1	D_2
D_3	D_4

D_1	D_2
D_3	D_4

Figure 3.2: The four one-site increments in a system of 4 spatial parts.

and the two-site increments:

$$\Delta\varepsilon_{ij} = \Delta\varepsilon(\{i, j\}) = \varepsilon_{ij} - \Delta\varepsilon_i - \Delta\varepsilon_j \quad (3.3)$$

D_1	D_2
D_3	D_4

D_1	D_2
D_3	D_4

D_1	D_2
D_3	D_4

D_1	D_2
D_3	D_4

D_1	D_2
D_3	D_4

D_1	D_2
D_3	D_4

Figure 3.3: All possible two-site increments within a space of four one-site increments.

We can simplify eqn. 3.1 according to the index symmetry:

$$\Delta\varepsilon_{ij} = \varepsilon_{ij} - \Delta\varepsilon_i - \Delta\varepsilon_j = \varepsilon_{ji} - \Delta\varepsilon_j - \Delta\varepsilon_i = \Delta\varepsilon_{ji} \quad (3.4a)$$

$$\varepsilon_{ij} = \varepsilon_{ji} \quad \Rightarrow \quad \Delta\varepsilon_{ij} = \Delta\varepsilon_{ji} \quad (3.4b)$$

$$\varepsilon_{ijk} = \varepsilon_{jik} = \varepsilon_{kji} = \varepsilon_{kij} = \varepsilon_{ikj} = \varepsilon_{jki} \quad \text{and eqn. 3.4b} \quad (3.4c)$$

$$\Rightarrow \quad \Delta\varepsilon_{ijk} = \Delta\varepsilon_{jik} = \dots \quad (3.4d)$$

Therefore we can restrict the sums in our expansion and eliminate the prefactors:

$$E_{\text{corr}} = \sum_i \Delta\varepsilon_i + \sum_{i < j} \Delta\varepsilon_{ij} + \sum_{i < j < k} \Delta\varepsilon_{ijk} + \dots \quad (3.5)$$

This simplification saves a large amount of computer time especially for higher orders. According to eqn. 3.5 we identify the index set in a n -fold summation in eqn. 3.5 with the power set over \mathbb{D} of the cardinality n . This can be done because the restriction of the n -tuples in the summation leaves only one n -tuple for a given set of indices. Therefore we can use instead of the the restricted index set the power of the set of domains $\mathcal{P}(\mathbb{D})$ as new index set. Using these considerations, we can write the general increment as:

$$\begin{aligned} \Delta\varepsilon_{i_1 i_2 \dots i_t} &= \varepsilon_{i_1 i_2 \dots i_t} - \sum_{I_{t-1}} \Delta\varepsilon_{I_{t-1}} - \sum_{I_{t-2}} \Delta\varepsilon_{I_{t-2}} - \dots \\ &\quad - \sum_{I_2} \Delta\varepsilon_{I_2} - \sum_{I_1} \Delta\varepsilon_{I_1} \end{aligned} \quad (3.6)$$

where the index I_{t-v} is defined as the set of those elements of the power set of $\{i_1, i_2, \dots, i_t\}$ which have the cardinality $(t-v)$ [v runs from 1 to $(t-1)$]. Eqn. 3.6 defines a mapping $\mathcal{P}(\mathbb{D}) \mapsto \mathbb{R}$. A close look at eqn. 3.5 yields that we can use the definition of the power set in order to simplify the notation.

$$\begin{aligned}
 E_{\text{corr}} &= \sum_{\substack{\mathbb{X} \in \mathcal{P}(\mathbb{D}) \\ |\mathbb{X}| \leq \mathcal{O}}} \Delta \varepsilon_{\mathbb{X}} \\
 \mathbb{D} &:= \text{set of domains} \\
 \mathcal{P}(\mathbb{D}) &:= \text{power set of the set of domains} \\
 \mathcal{O} &:= \text{order of the expansion}
 \end{aligned} \tag{3.7}$$

where the one-site domains D_λ are defined as disjoint sets of occupied orbitals:

$$D_\lambda = \{\phi_a, \phi_b, \phi_c, \dots\} \tag{3.8}$$

$$D_\lambda \cap D_\mu = \emptyset \tag{3.9}$$

Alternatively we can represent the one-site domains by a set of vectors \mathcal{D}_λ :

$$\begin{aligned}
 \mathcal{D}_\lambda &= \bigcup_{\phi_a \in D_\lambda} \vec{R}_a \\
 \text{with } \phi_a &\mapsto \vec{R}_a = \begin{pmatrix} \langle \phi_a | x | \phi_a \rangle \\ \langle \phi_a | y | \phi_a \rangle \\ \langle \phi_a | z | \phi_a \rangle \end{pmatrix}
 \end{aligned} \tag{3.10}$$

The definition of the general increment $\Delta \varepsilon_{\mathbb{X}}$ reads in the short hand notation:

$$\Delta \varepsilon_{\mathbb{X}} = \varepsilon_{\mathbb{X}} - \sum_{\mathbb{Y} \in \mathcal{P}(\mathbb{X}) \wedge |\mathbb{Y}| < |\mathbb{X}|} \Delta \varepsilon_{\mathbb{Y}} \tag{3.11}$$

For closed shell systems it was found that the convergence of the series eqn. 3.7 is reasonably fast [35, 44, 49, 84–86]. Therefore we can truncate the series usually at low order ($|\mathcal{O}| \leq 4$).

3.1.1 Exactness of the Incremental Expansion

The incremental expansion is set up in a way that the exact result (of the given correlation method) is obtained if all terms in the expansion are considered.

A Simple Example

As a simple example we consider a system with 3 spatial parts as shown in figure 3.4 and calculate the incremental sum of eqn. 3.1 for this system.

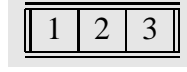


Figure 3.4: A test system with 3 spatial parts 1, 2, 3.

We have to calculate 3 one-site increments, $\Delta\varepsilon_1$, $\Delta\varepsilon_2$, $\Delta\varepsilon_3$, 3 two-site increments, $\Delta\varepsilon_{12}$, $\Delta\varepsilon_{13}$, $\Delta\varepsilon_{23}$ and one three-site increment $\Delta\varepsilon_{123}$. Using eqn. 3.1-3.3 we end up with:

$$\begin{aligned}
 E_{\text{corr}} &= \Delta\varepsilon_1 + \Delta\varepsilon_2 + \Delta\varepsilon_3 + \Delta\varepsilon_{12} + \Delta\varepsilon_{13} + \Delta\varepsilon_{23} + \Delta\varepsilon_{123} = \\
 &\underline{\varepsilon_1 + \varepsilon_2 + \varepsilon_3} + \underline{\varepsilon_{12} - \varepsilon_1 - \varepsilon_2} + \underline{\varepsilon_{13} - \varepsilon_1 - \varepsilon_3} + \underline{\varepsilon_{23} - \varepsilon_3 - \varepsilon_2} \\
 &+ \underline{\varepsilon_{123} - \varepsilon_{12} + \varepsilon_1 + \varepsilon_2} - \underline{\varepsilon_{13} + \varepsilon_1 + \varepsilon_3} - \underline{\varepsilon_{23} + \varepsilon_3 + \varepsilon_2} - \underline{\varepsilon_1 - \varepsilon_2 - \varepsilon_3} \\
 &= \varepsilon_{123}
 \end{aligned} \tag{3.12}$$

From eqn. 3.12 we recognize that at the end of the expansion the highest order term cancels exactly the contributions from the lower orders. Therefore we obtain the exact result within the applied correlation method at the highest order level.

The general case

We start with eqn. 3.7 and sum over all contributions up to the highest order.

$$E_{\text{corr}} = \sum_{\mathbb{X} \in \mathcal{P}(\mathbb{D})} \Delta\varepsilon_{\mathbb{X}} = \sum_{\mathbb{X} \in \mathcal{P}(\mathbb{D}) \wedge |\mathbb{X}| < |\mathbb{D}|} \Delta\varepsilon_{\mathbb{X}} + \Delta\varepsilon_{\mathbb{D}} \tag{3.13}$$

$\mathbb{D} :=$ set of domains

$\mathcal{P}(\mathbb{D}) :=$ power set of the set of domains

Now we substitute $\Delta\varepsilon_{\mathbb{D}}$ by the definition of the general increment eqn. 3.11.

$$E_{\text{corr}} = \sum_{\mathbb{X} \in \mathcal{P}(\mathbb{D}) \wedge |\mathbb{X}| < |\mathbb{D}|} \Delta\varepsilon_{\mathbb{X}} + \varepsilon_{\mathbb{D}} - \sum_{\mathbb{X} \in \mathcal{P}(\mathbb{D}) \wedge |\mathbb{X}| < |\mathbb{D}|} \Delta\varepsilon_{\mathbb{X}} = \varepsilon_{\mathbb{D}} \tag{3.14}$$

We can see immediately that the only remaining term is the exact correlation energy of the total system. This demonstrates that the incremental expansion is in principle exact.

3.1.2 Approximations to the Incremental Scheme

Due to large spatial distances of the included domains one may encounter nearly vanishing n -site increments. If the one-site domains of a given two-site domain are far apart, we find that due to the local nature of electron correlation, the corresponding incremental energy correction is small since the correlation energy of the two-site domain is given as:

$$\varepsilon_{ij} = \varepsilon_i + \varepsilon_j + \varepsilon(R_{ij}) \quad (3.15)$$

In the limit of an infinite distance R_{ij} between the two domains i and j the correction $\varepsilon(R_{ij})$ vanishes exactly and the increment $\Delta\varepsilon_{ij}$ in eqn. 3.16 becomes exactly zero:

$$\begin{aligned} \Delta\varepsilon_{ij} &= \varepsilon_{ij} - \Delta\varepsilon_i - \Delta\varepsilon_j \\ &= \varepsilon_i + \varepsilon_j + \varepsilon(R_{ij}) - \Delta\varepsilon_i - \Delta\varepsilon_j \\ &= \varepsilon(R_{ij}) = 0 \quad \text{for } R_{ij} = \infty \end{aligned} \quad (3.16)$$

This can be generalized to higher order increments (vide infra). Obviously, for an efficient implementation one should exploit this property which goes beyond the original definition of the incremental series. The implementation of such a truncation can be achieved by a distance threshold R_{min} . Since the higher order terms are usually smaller than the low-order terms, we can decrease the threshold R_{min} with increasing order (see chapter 6).

We can neglect the energy contribution of a n -site domain if we can form at least two subsets of this domain with a minimum distance larger than R_{min} .

We prove this rule by induction. We start with the case $n = 2$ which is proven in eqn. 3.16. Now we assume that the set of domains \mathbb{X} can be divided into two disjoint sets \mathbb{A} and \mathbb{B} with $\mathbb{A}, \mathbb{B} \neq \emptyset$ and where all distances between the elements of \mathbb{A} and the elements of \mathbb{B} are larger than R_{min} . With this requirement we find an expression for the correlation energy for $\varepsilon_{\mathbb{X}}$ according to eqn. 3.16. This can be done because the domains i and j in eqn. 3.16 were chosen freely. Therefore we can always evaluate $\varepsilon_{\mathbb{X}}$ by:

$$\varepsilon_{\mathbb{X}} = \varepsilon_{\mathbb{A}} + \varepsilon_{\mathbb{B}} + \varepsilon(R_{\mathbb{A}\mathbb{B}}) \quad (3.17)$$

We define the largest correction $\varepsilon(R_{\mathbb{A}\mathbb{B}})$ which is obtained from $\Delta\varepsilon_{\tilde{\mathbb{X}}}$ for the subsets $\tilde{\mathbb{X}}$ of \mathbb{X} with $\mathbb{A} \cap \tilde{\mathbb{X}} = \tilde{\mathbb{A}} \neq \emptyset$ and $\mathbb{B} \cap \tilde{\mathbb{X}} = \tilde{\mathbb{B}} \neq \emptyset$ as $\varepsilon(R_{min})$. Using this we can estimate the absolute value of an increment $\Delta\varepsilon_{\tilde{\mathbb{X}}}$:

$$|\Delta\varepsilon_{\tilde{\mathbb{X}}}| = |\varepsilon_{\tilde{\mathbb{X}}} - \varepsilon_{\tilde{\mathbb{A}}} - \varepsilon_{\tilde{\mathbb{B}}}| \leq |\varepsilon(R_{min})| \quad \forall \tilde{\mathbb{X}} \quad (3.18)$$

Before we start with the induction we give an example for the 3-body increment $\Delta\varepsilon_{ijk}$ with two sets of domains $\mathbb{A} = \{i, j\}$ and $\mathbb{B} = \{k\}$ with a certain distance R_{min}

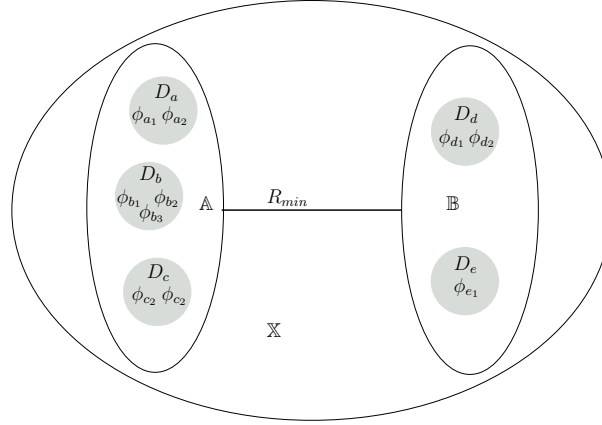


Figure 3.5: Pictorial view of the involved sets, if the set of domains \mathbb{X} is partitioned into two sets \mathbb{A} and \mathbb{B} with the minimal distance R_{min} between these sets.

between these domains. We can use eqn. 3.18 and eqn. 3.17 to approximate the terms with the mixed index combinations of eqn. 3.6.

$$\begin{aligned} |\Delta\varepsilon_{ik}| &\leq |\varepsilon(R_{min})| \\ |\Delta\varepsilon_{jk}| &\leq |\varepsilon(R_{min})| \end{aligned} \quad (3.19)$$

$$\varepsilon_{ijk} = \varepsilon_{ij} + \varepsilon_k + \varepsilon(R_{\{ij\},\{k\}}) \quad (3.20)$$

Inserting eqn. 3.20 into eqn. 3.6 and using the definition of the 2-site increments we obtain:

$$\begin{aligned} |\varepsilon_{ijk}| &= |\varepsilon_{ij} + \varepsilon_k + \varepsilon(R_{\{ij\},\{k\}}) \\ &\quad - (\varepsilon_{ij} - \Delta\varepsilon_i - \Delta\varepsilon_j + \Delta\varepsilon_{ik} + \Delta\varepsilon_{kj} + \Delta\varepsilon_i + \Delta\varepsilon_j + \Delta\varepsilon_k)| \\ &\leq 2 \cdot |\varepsilon(R_{min})| + |\varepsilon_{\{ij\},\{k\}}| \leq 3 \cdot |\varepsilon(R_{min})| \end{aligned} \quad (3.21)$$

where we used eqn. 3.18 in the last step to approximate $\varepsilon(R_{\{ij\},\{k\}})$. We generalize these considerations above by introducing the induction assumption:

$$\begin{aligned} |\Delta\varepsilon_{\mathbb{X}}| &\leq c \cdot |\varepsilon(R_{min})| \quad \text{if} \\ A &= \bigcup_{D_a \in \mathbb{A}} D_a \quad B = \bigcup_{D_b \in \mathbb{B}} D_b \\ |\vec{R}_a - \vec{R}_b| &> R_{min} \quad \forall \phi_a \in A, \phi_b \in B \\ \mathbb{A}, \mathbb{B} &\neq \emptyset, \mathbb{A} \cup \mathbb{B} = \mathbb{X} \end{aligned}$$

and $|\mathbb{X}| \leq n$. Now we write the $(n + 1)$ case as:

$$\Delta\varepsilon_{\mathbb{Y}} = \varepsilon_{\mathbb{Y}} - \sum_{\mathbb{I} \in \mathcal{P}(\mathbb{Y}) \wedge |\mathbb{I}| < |\mathbb{Y}|} \Delta\varepsilon_{\mathbb{I}} \quad (3.22)$$

with $|\mathbb{Y}| = n + 1$. Next we classify the sets \mathbb{I} in eqn. 3.22 according to the mixed indices X_m and not mixed indices X_{nm} :

$$\begin{aligned} X_m &:= \{\mathbb{I} \in \mathcal{P}(\mathbb{Y}) \mid |\mathbb{I}| < |\mathbb{Y}| \wedge \mathbb{I} \cap \mathbb{A} \neq \emptyset \wedge \mathbb{I} \cap \mathbb{B} \neq \emptyset\} \\ X_{nm} &:= \{\mathbb{I} \in \mathcal{P}(\mathbb{Y}) \setminus X_m \mid |\mathbb{I}| < |\mathbb{Y}|\} \end{aligned} \quad (3.23)$$

Then we split the summation over \mathbb{I} into two summations:

$$\Delta\varepsilon_{\mathbb{Y}} = \varepsilon_{\mathbb{Y}} - \sum_{\mathbb{I} \in X_m} \Delta\varepsilon_{\mathbb{I}} - \sum_{\mathbb{I} \in X_{nm}} \Delta\varepsilon_{\mathbb{I}} \quad (3.24)$$

We use eqn. 3.17 to evaluate $\varepsilon_{\mathbb{Y}}$ as:

$$\varepsilon_{\mathbb{Y}} = \varepsilon_{\mathbb{A}} + \varepsilon_{\mathbb{B}} + \varepsilon(R_{\mathbb{A}\mathbb{B}})$$

Since the cardinalities of the sets in X_m and X_{nm} are always smaller than the cardinality of \mathbb{Y} , we can use the induction assumption to evaluate the summation over the mixed indices X_m . An upper bound for the introduced error is

$$\sum_{\mathbb{I} \in X_m} |\Delta\varepsilon_{\mathbb{I}}| \leq \sum_{i=1}^{|X_m|} c_i \cdot |\varepsilon(R_{min})|$$

where the prefactors c_i are determined by the cardinality of the element of X_m . The summation over X_{nm} can be split into two summations ordered by the sets \mathbb{A} and \mathbb{B} :

$$\sum_{\mathbb{I} \in X_{nm}} \Delta\varepsilon_{\mathbb{I}} = \sum_{\mathbb{I} \in \mathcal{P}(\mathbb{A})} \Delta\varepsilon_{\mathbb{I}} + \sum_{\mathbb{I} \in \mathcal{P}(\mathbb{B})} \Delta\varepsilon_{\mathbb{I}}$$

There are no restrictions to the power sets because:

$$\begin{aligned} |\mathbb{A}| + |\mathbb{B}| &= |\mathbb{Y}| \wedge |\mathbb{A}|, |\mathbb{B}| > 0 \\ &\Rightarrow |\mathbb{A}|, |\mathbb{B}| < |\mathbb{Y}| \end{aligned}$$

Since the summations over the power sets of \mathbb{A} and \mathbb{B} are not restricted we can replace the summations by the exact results $\varepsilon_{\mathbb{A}}$ and $\varepsilon_{\mathbb{B}}$ respectively.

$$\sum_{\mathbb{I} \in X_{nm}} \Delta\varepsilon_{\mathbb{I}} = \varepsilon_{\mathbb{A}} + \varepsilon_{\mathbb{B}}$$

Using the considerations above we get for the absolute value of the increment $\Delta\varepsilon_{\mathbb{Y}}$:

$$\begin{aligned}
|\Delta\varepsilon_{\mathbb{Y}}| &= \left| \varepsilon_{\mathbb{A}} + \varepsilon_{\mathbb{B}} + \varepsilon(R_{\mathbb{A}\mathbb{B}}) - \sum_{\substack{\mathbb{I} \\ \mathbb{I} \in X_m}} \Delta\varepsilon_{\mathbb{I}} - \varepsilon_{\mathbb{A}} - \varepsilon_{\mathbb{B}} \right| \\
&\leq |\varepsilon(R_{\mathbb{A}\mathbb{B}})| + \sum_{i=1}^{|X_m|} c_i \cdot |\varepsilon(R_{min})| \\
&\leq \left(\sum_{i=1}^{|X_m|} c_i + 1 \right) \cdot |\varepsilon(R_{min})|
\end{aligned} \tag{3.25}$$

where we used eqn. 3.18 to estimate $|\varepsilon(R_{\mathbb{A}\mathbb{B}})|$. The factor c is given as

$$c = \left(\sum_{i=1}^{|X_m|} c_i + 1 \right)$$

which completes the proof, since it is independent of the distance R_{min} .

For high-order contributions it might be a problem that the prefactors are very large and therefore R_{min} has to be very large. However, for a fourth-order increment this is no serious problem, as we can see if we consider the worst case example. The cardinality of a power set can be calculated by $|\mathcal{P}(\{1, 2, \dots, n\})| = 2^n$. In order to obtain the cardinality of X_m we have to subtract the cardinalities of $\mathcal{P}(\mathbb{A})$ and $\mathcal{P}(\mathbb{B})$ from $\mathcal{P}(\mathbb{X})$. In the worst case this is given as:

$$\text{second-order: } |X_m| = 2^2 - 2^1 - 2^1 = 0$$

$$\text{third-order: } |X_m| = 2^3 - 2^1 - 2^2 = 2$$

Using this result we obtain for the second-order coefficient $c^{(2)} = 1$, for third-order coefficient $c^{(3)} = 3$ and for the fourth-order coefficient we have $c^{(4)} = 4 \cdot 3 + 4 \cdot 1 = 16$.

3.1.3 Energy Screening

Another way to reduce the total number of calculations is to calculate the energy increments with a lower level method and neglect all terms which are smaller than a given threshold at the low-level method.

$$E_{\text{corr}} = \sum_{\substack{\mathbb{X} \\ \mathbb{X} \in \mathcal{P}(\mathbb{D}) \wedge |\mathbb{X}| \leq \mathcal{O} \\ |\Delta\varepsilon_{\mathbb{X}}| > E_{\text{thres}}}} \Delta\varepsilon_{\mathbb{X}} \tag{3.26}$$

Here \mathbb{X} runs over all members of the power set of the set of the domains $\mathcal{P}(\mathbb{D})$, up to a certain cardinality as in eqn. 3.7.

3.1.4 Formal Scaling

The formal scaling of the incremental expansion in eqn. 3.1 is determined by the number of individual calculations \mathcal{N}_{calc} and the time for the individual calculations of the subsystems. The number of calculations is given as:

$$\mathcal{N}_{calc} = \sum_{i=1}^{\mathcal{O}} \binom{|\mathbb{D}|}{i} \quad (3.27)$$

with $|\mathbb{D}| \equiv$ number of domains. The total computational time t for incremental calculations can be evaluated as the sum over all single calculations:

$$t = \sum_i t_i \leq \binom{|\mathbb{D}|}{1} \cdot t_1^{max} + \binom{|\mathbb{D}|}{2} \cdot t_2^{max} + \dots + \binom{|\mathbb{D}|}{\mathcal{O}} \cdot t_{\mathcal{O}}^{max} \quad (3.28)$$

where t_i^{max} is the time for the most time consuming calculation at i -th order. Using CCSD theory we have :

$$t_i^{max} \propto \mathcal{N}_{\mathbb{O}}^2(i) \cdot \mathcal{N}_{\mathbb{V}}^4(i) \quad \mathcal{N}_{\mathbb{V}}(i) \gg \mathcal{N}_{\mathbb{O}}(i) \quad (3.29)$$

where $\mathcal{N}_{\mathbb{O}}(i)$ is the number of occupied orbitals in the domain corresponding to t_i^{max} and $\mathcal{N}_{\mathbb{V}}(i)$ the number of virtual orbitals in this domain. For the most time consuming step we have:

$$t_{\mathcal{O}}^{max} \propto \mathcal{N}_{\mathbb{O}}^2(\mathcal{O}) \cdot \mathcal{N}_{\mathbb{V}}^4(\mathcal{O}) \quad (3.30)$$

Using eqn. 3.27-3.30 we find an upper bound u for the computational time according to:

$$u \propto \mathcal{N}_{calc} \cdot \mathcal{N}_{\mathbb{O}}^2(\mathcal{O}) \cdot \mathcal{N}_{\mathbb{V}}^4(\mathcal{O}) \quad (3.31)$$

The dimensions $\mathcal{N}_{\mathbb{O}}(\mathcal{O})$ and $\mathcal{N}_{\mathbb{V}}(\mathcal{O})$ of the n -site domains will have an upper bound, independent of the size of the total system, if the size of the one-site increments is fixed and the incremental series is truncated at a given order. The largest $\mathcal{N}_{\mathbb{O}}(\mathcal{O})$ is fixed to a constant by the set of the one-site domains and the order of the expansion. Since $\mathcal{N}_{\mathbb{V}}(\mathcal{O})$ is determined by $\mathcal{N}_{\mathbb{O}}(\mathcal{O})$, we fix this dimension, too (section 4.1.1). Therefore we can rewrite eqn. 3.31 as:

$$u \propto \mathcal{N}_{calc} \cdot t_{\mathcal{O}}^{max} \quad \text{with } t_{\mathcal{O}}^{max} = \text{const} \quad (3.32)$$

If we further neglect all energy increments which are separated by a certain distance, according to eqn. 3.16 the number of calculations increases linearly with the system size.

$$\mathcal{N}_{calc} \approx \mathcal{N} \quad (3.33)$$

Consequently, it is possible to set up the incremental scheme as a linear scaling method. There are two important points to note here. Firstly we have to truncate the virtual space for the domains in order to get a constant virtual space $\mathcal{N}_v(\mathcal{O})$ and secondly we need a linear scaling MO-transformation for the incremental scheme. However, with the current implementation using the CCSD code in MOLPRO [87] we are limited to a formal \mathcal{N}^5 scaling if we use eqn. 3.16 to reduce the number of calculations.

$$N_{\mathcal{O}}(\mathcal{O}) = \text{const}, \quad \mathcal{N}_{calc} \propto \mathcal{N}, \quad \mathcal{N}_v^4 \equiv \mathcal{N}^4$$

The virtual space dependence enters completely into our scaling, because of the impossibility to truncate the virtual space within the MOLPRO CCSD code.

3.2 Derivation of the Correlation Energy for a Solid

The total correlation energy $E_{\text{corr}}^{\text{total}}$ of a system of n domains may be obtained by

$$\begin{aligned} E_{\text{corr}}^{\text{total}} = & \frac{1}{1!} \sum_I^n \Delta\varepsilon_I + \frac{1}{2!} \sum_{IJ}^n \Delta\varepsilon_{IJ} + \frac{1}{3!} \sum_{IJK}^n \Delta\varepsilon_{IJK} + \dots \\ & + \frac{1}{|\{I, J, K, \dots, T\}|!} \sum_{IJK\dots T}^n \Delta\varepsilon_{IJK\dots T} + \dots + \frac{1}{n!} \sum_{IJK\dots T\dots N}^n \Delta\varepsilon_{IJK\dots T\dots N} \end{aligned} \quad (3.34)$$

where the indices I, J, K, \dots, T run over all domains. In a periodic system the correlation energy is infinite, therefore one refers to the correlation energy per cell $E_{\text{corr}}^{\text{cell}}$ which is defined as:

$$E_{\text{corr}}^{\text{cell}} = \lim_{n \rightarrow \infty} \frac{E_{\text{corr}}^{\text{total}}(n)}{n} \quad (3.35)$$

If we have a supercell of n cells, we can divide the system into n equal parts:

$$E_{\text{corr}}^{\text{total}}(n) = \sum_{i=1}^n E_{\text{corr}}^i(n) \quad (3.36)$$

Inserting eqn. 3.36 into eqn. 3.35 we obtain:

$$E_{\text{corr}}^{\text{cell}} = \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n E_{\text{corr}}^i(n)}{n} = \lim_{n \rightarrow \infty} \frac{n \cdot E_{\text{corr}}^{\text{cell}}(n)}{n} = \lim_{n \rightarrow \infty} E_{\text{corr}}^{\text{cell}}(n) \quad (3.37)$$

The n dependence in eqn. 3.37 refers to the infinity of the system. Now we rewrite eqn. 3.34:

$$E_{\text{corr}}^{\text{total}}(n) = \sum_{I=1}^n \left[\frac{1}{1!} \Delta\varepsilon_I + \frac{1}{2!} \sum_{J=1}^n \Delta\varepsilon_{IJ} + \frac{1}{3!} \sum_{J,K=1}^n \Delta\varepsilon_{IJK} + \dots \right] \quad (3.38)$$

If we compare the right hand side of eqn. 3.36 with eqn. 3.38 we find the incremental energy per cell as:

$$E_{\text{corr}}^{\text{cell}}(n) = \frac{1}{1!} \Delta \varepsilon_I + \frac{1}{2!} \sum_{J=1}^n \Delta \varepsilon_{IJ} + \frac{1}{3!} \sum_{J,K=1}^n \Delta \varepsilon_{IJK} + \dots \quad (3.39)$$

taking the limit $n \rightarrow \infty$ we get:

$$E_{\text{corr}}^{\text{cell}} = \frac{1}{1!} \Delta \varepsilon_I + \frac{1}{2!} \sum_{J=1}^{\infty} \Delta \varepsilon_{IJ} + \frac{1}{3!} \sum_{J,K=1}^{\infty} \Delta \varepsilon_{IJK} + \dots \quad (3.40)$$

At this point we use the index symmetry in eqn. 3.39 in order to reduce the computational effort drastically:

$$E_{\text{corr}}^{\text{cell}} = \Delta \varepsilon_I + \frac{1}{2} \sum_{J=1}^{\infty} \Delta \varepsilon_{IJ} + \frac{1}{3} \sum_{J>K, K=1}^{\infty} \Delta \varepsilon_{IJK} + \dots \quad (3.41)$$

The prefactors of $\frac{1}{\mathcal{O}}$ can also be dropped if we sum over translationally equivalent increments only once, since we have n translationally equivalent n -site increments.

3.3 Incremental Correlation Energy for Small Domains

Sometimes it might be advantageous to divide the reference cell into smaller domains. If we treat the problem straightforward, the incremental expansion reads:

$$E_{\text{corr}}^{\text{cell}} = \frac{1}{1!} \sum_{I=1}^l \Delta \varepsilon_I + \frac{1}{2!} \sum_{I=1}^l \sum_{J=1}^{\infty} \Delta \varepsilon_{IJ} + \frac{1}{3!} \sum_{I=1}^l \sum_{J=1}^{\infty} \sum_{K=1}^{\infty} \Delta \varepsilon_{IJK} + \dots \quad (3.42)$$

where I runs over all domains in the reference cell, J, K run over all domains, l is the number of one-site domains in the reference cell and the set \mathbb{D} is ordered in a way that the first l elements are located in the reference cell. We have to consider two cases for a second-order increment:

1. both indices are in the reference cell
2. one index is in a different cell

for a third-order increment we have already 4 cases:

1. all indices are in the reference cell
2. two indices are the reference cell and the other one in the environment
3. one index is in the reference cell and the other two in the same cell of the environment
4. one index is in the reference cell and the other two in different cells of the environment

Since all these cases have different prefactors, it's quite tedious to treat higher order increments.

3.3.1 Expansion of the Fragmental Energies in a Further Expansion

To derive a general expression for the incremental energy we expand the correlation energies of the fragments $\varepsilon_{\mathbb{X}}$ of eqn. 3.41 in a further expansion. The one-site increment ε_I reads:

$$\varepsilon_I = \frac{1}{1!} \sum_{(a)_I} \Delta\varepsilon_{(a)_I} + \frac{1}{2!} \sum_{(ab)_I} \Delta\varepsilon_{(ab)_I} + \frac{1}{3!} \sum_{(abc)_I} \Delta\varepsilon_{(abc)_I} + \dots \quad (3.43)$$

The index I restricts the index space of a, b, c, \dots to the cell I . In the short hand notation we get:

$$\begin{aligned} \varepsilon_I &= \sum_{\mathbb{Y} \in \overline{\mathcal{P}(\mathbb{X}_I)}} \Delta\varepsilon_{\mathbb{Y}} \\ \varepsilon_{IJ} &= \sum_{\mathbb{Y} \in \overline{\mathcal{P}(\mathbb{X}_{IJ})}} \Delta\varepsilon_{\mathbb{Y}} \end{aligned} \quad (3.44)$$

where \mathbb{X}_I is the subset of the domains located in the cell I and \mathbb{X}_{IJ} is the subset of the domains in the cells I and J . For the two-site increments we get in this case:

$$\Delta\varepsilon_{IJ} = \sum_{\mathbb{Y} \in \overline{\mathcal{P}(\mathbb{X}_{IJ})}} \Delta\varepsilon_{\mathbb{Y}} - \sum_{\mathbb{Y} \in \overline{\mathcal{P}(\mathbb{X}_I) \cup \mathcal{P}(\mathbb{X}_J)}} \Delta\varepsilon_{\mathbb{Y}} = \sum_{\substack{\mathbb{Y} \in \overline{\mathcal{P}(\mathbb{X}_{IJ})} \\ [\mathcal{P}(\mathbb{X}_I) \cup \mathcal{P}(\mathbb{X}_J)]}} \Delta\varepsilon_{\mathbb{Y}} \quad (3.45)$$

Since the higher order increments have a similar structure, we introduce a short hand notation for the set of indices in the last term in order to simplify the notation:

$$\overline{\mathcal{P}(\mathbb{X}_{IJ})} = \overline{\mathcal{P}(\mathbb{X}_{IJ})} \setminus [\mathcal{P}(\mathbb{X}_I) \cup \mathcal{P}(\mathbb{X}_J)] \quad (3.46)$$

The general index set $\overline{\mathcal{P}(\mathbb{X}_{\mathbb{K}})}$ is defined as:

$$\overline{\mathcal{P}(\mathbb{X}_{\mathbb{K}})} = \overline{\mathcal{P}(\mathbb{X}_{\mathbb{K}})} \setminus \bigcup_{\substack{\lambda \in \overline{\mathcal{P}(\mathbb{K})} \\ |\lambda| < |\mathbb{K}|}} \overline{\mathcal{P}(\mathbb{X}_{\lambda})} \quad (3.47)$$

where \mathbb{K} denotes a set of cells. Note that the definition of the $\overline{\mathcal{P}(\mathbb{X}_{IJ})}$ in eqn. 3.46 is equivalent to the general definition in eqn. 3.47, since $\overline{\mathcal{P}(\mathbb{X}_I)} = \overline{\mathcal{P}(\mathbb{X}_I)}$. We note that

the sets $\overline{\mathcal{P}(\mathbb{X}_{\mathbb{K}})}$ and $\overline{\mathcal{P}(\mathbb{X}_{\mathbb{K}'})}$ are disjoint for different sets \mathbb{K} and \mathbb{K}' . Using eqn. 3.47 we can write the general form of eqn. 3.45 as:

$$\Delta\varepsilon_{\mathbb{K}} = \sum_{\substack{\mathbb{Y} \\ \mathbb{Y} \in \overline{\mathcal{P}(\mathbb{X}_{\mathbb{K}})}}} \Delta\varepsilon_{\mathbb{Y}} \quad (3.48)$$

We prove eqn. 3.48 by induction. We assume that eqn. 3.48 is true for all lower orders ($n - 1$ condition). This is equivalent to that it is true for all subsets \mathbb{L} with $|\mathbb{L}| < |\mathbb{K}|$. Now we substitute this condition into the general definition of an increment eqn. 3.11 and obtain:

$$\begin{aligned} \Delta\varepsilon_{\mathbb{K}} &= \sum_{\substack{\mathbb{Y} \\ \mathbb{Y} \in \overline{\mathcal{P}(\mathbb{X}_{\mathbb{K}})}}} \Delta\varepsilon_{\mathbb{Y}} - \sum_{\substack{\mathbb{Y} \\ \mathbb{Y} \in \overline{\mathcal{P}(\mathbb{K})} \\ |\mathbb{Y}| < |\mathbb{K}|}} \Delta\varepsilon_{\mathbb{Y}} = \sum_{\substack{\mathbb{Y} \\ \mathbb{Y} \in \overline{\mathcal{P}(\mathbb{X}_{\mathbb{K}})}}} \Delta\varepsilon_{\mathbb{Y}} - \sum_{\substack{\mathbb{Y} \\ \mathbb{Y} \in \overline{\mathcal{P}(\mathbb{K})} \\ |\mathbb{Y}| < |\mathbb{K}|}} \sum_{\substack{\mathbb{L} \\ \mathbb{L} \in \overline{\mathcal{P}(\mathbb{X}_{\mathbb{Y}})}}} \Delta\varepsilon_{\mathbb{L}} \\ &= \sum_{\substack{\mathbb{Y} \\ \mathbb{Y} \in \overline{\mathcal{P}(\mathbb{X}_{\mathbb{K}})}}} \Delta\varepsilon_{\mathbb{Y}} - \sum_{\substack{\mathbb{Y} \\ \mathbb{Y} \in \mathbb{A}}} \Delta\varepsilon_{\mathbb{Y}} = \sum_{\substack{\mathbb{Y} \\ \mathbb{Y} \in \overline{\mathcal{P}(\mathbb{X}_{\mathbb{K}})} \setminus \mathbb{A}}} \Delta\varepsilon_{\mathbb{Y}} \end{aligned} \quad (3.49)$$

Due to the fact that the sets in the double summation in eqn. 3.49 are disjoint, we can replace the double summation by a summation over a unified index set \mathbb{A} ,

$$\mathbb{A} = \bigcup_{\substack{\mathbb{Y} \\ \mathbb{Y} \in \overline{\mathcal{P}(\mathbb{K})} \\ |\mathbb{Y}| < |\mathbb{K}|}} \overline{\mathcal{P}(\mathbb{X}_{\mathbb{Y}})} \quad (3.50)$$

If we insert the definition of \mathbb{A} into the last sum in eqn. 3.49 we see immediately that we sum over the index set $\overline{\mathcal{P}(\mathbb{X}_{\mathbb{K}})}$ defined in eqn. 3.47, which is the desired result. We start the induction at second order which is proven in eqn. 3.46.

3.3.2 Separate Treatment of the Translational Symmetry

The goal of using a second expansion for the inner cell correlation energies is to separate the translational symmetry from the index symmetry. Essentially we avoid a separate treatment of distinct cases. Since we can use eqn. 3.48 to simplify the expressions, we get the same terms as in the straightforward expansion eqn. 3.42. The only difference is that we regrouped the terms according to the translational symmetry.

Collecting this we end up with:

$$\begin{aligned}
 E_{\text{corr}}^{\text{cell}} = & \frac{1}{1!} \left[\sum_{\substack{\mathbb{Y} \\ \mathbb{Y} \in \overline{\mathcal{P}(\mathbb{X}_I)}}} \Delta \varepsilon_{\mathbb{Y}} \right] + \frac{1}{2!} \sum_J \left[\sum_{\substack{\mathbb{Y} \\ \mathbb{Y} \in \overline{\mathcal{P}(\mathbb{X}_{IJ})}}} \Delta \varepsilon_{\mathbb{Y}} \right] \\
 & + \frac{1}{3!} \sum_{JK} \left[\sum_{\substack{\mathbb{Y} \\ \mathbb{Y} \in \overline{\mathcal{P}(\mathbb{X}_{IJK})}}} \Delta \varepsilon_{\mathbb{Y}} \right] + \dots
 \end{aligned} \tag{3.51}$$

In order to get eqn. 3.51 equivalent to the straight forward expansion, we have to restrict the index sets \mathbb{Y} to a certain order \mathcal{O} ($|\mathbb{Y}| \leq \mathcal{O}$). The index symmetry of the inner expansion is present in the index space $\overline{\mathcal{P}(\mathbb{X}_{\mathbb{K}})}$ completely. For the outer expansion we can remove this index symmetry by restricting the summation. We want to point out that the prefactors do not cancel completely in this case, since we restrict the index I to be in the reference cell.

$$\begin{aligned}
 E_{\text{corr}}^{\text{cell}} = & \left[\sum_{\substack{\mathbb{Y} \\ \mathbb{Y} \in \overline{\mathcal{P}(\mathbb{X}_I)}}} \Delta \varepsilon_{\mathbb{Y}} \right] + \frac{1}{2} \sum_J \left[\sum_{\substack{\mathbb{Y} \\ \mathbb{Y} \in \overline{\mathcal{P}(\mathbb{X}_{IJ})}}} \Delta \varepsilon_{\mathbb{Y}} \right] \\
 & + \frac{1}{3} \sum_{J>K} \left[\sum_{\substack{\mathbb{Y} \\ \mathbb{Y} \in \overline{\mathcal{P}(\mathbb{X}_{IJK})}}} \Delta \varepsilon_{\mathbb{Y}} \right] + \dots
 \end{aligned} \tag{3.52}$$

In the next step we use the translational symmetry to cancel the prefactors completely. Since the indices I, J, \dots run over cells, we have exactly two equal terms due to translation (e.g. \mathbb{Y} in $\overline{\mathcal{P}(\mathbb{X}_{IJ})}$ and \mathbb{Y} in $\overline{\mathcal{P}(\mathbb{X}_{IJ'})}$). In general the prefactor of the sum over I, J, \dots is always equal to the inverse of the number of translationally equivalent increments. Therefore we leave this prefactors out and sum only over translationally non-equivalent increments.

$$\begin{aligned}
 E_{\text{corr}}^{\text{cell}} = & \left[\sum_{\substack{\mathbb{Y} \\ \mathbb{Y} \in \overline{\mathcal{P}(\mathbb{X}_I)}}} \Delta \varepsilon_{\mathbb{Y}} \right] + \sum_{J'} \left[\sum_{\substack{\mathbb{Y} \\ \mathbb{Y} \in \overline{\mathcal{P}(\mathbb{X}_{IJ'})}}} \Delta \varepsilon_{\mathbb{Y}} \right] \\
 & + \sum_{J'>K'} \left[\sum_{\substack{\mathbb{Y} \\ \mathbb{Y} \in \overline{\mathcal{P}(\mathbb{X}_{IJK'})}}} \Delta \varepsilon_{\mathbb{Y}} \right] + \dots
 \end{aligned} \tag{3.53}$$

Here the restriction to translational different terms is denoted by the primes in the summation.

3.4 Incremental Expansion for non-Disjoint Cells

We derived eqn. 3.52 with the assumption that the cells are disjoint. In order to explore the full crystallographic symmetry of the crystal we have to allow face, edge, or corner shared domains. For this purpose, we derive eqn. 3.53 for disjoint cells again with a different ansatz. In the next step we extend the approach to non-disjoint unit cells. We write the correlation energy of a finite system of n cells and the set of domains \mathbb{D} as:

$$\begin{aligned}
E_{\text{corr}}^{\text{total}} = & \frac{1}{1!} \sum_{i'} \Delta \varepsilon_{i'} + \frac{1}{2!} \sum_{i'j} \Delta \varepsilon_{i'j} + \frac{1}{3!} \sum_{i'jk} \Delta \varepsilon_{i'jk} + \dots \\
& + \frac{1}{|\{i', j, k, \dots, t\}|!} \sum_{i'j\dots t} \Delta \varepsilon_{i'jk\dots t} + \dots + \frac{1}{|\mathbb{D}|!} \sum_{i'jk\dots t\dots z} \Delta \varepsilon_{i'jk\dots t\dots z} \quad (3.54) \\
& i', j, k, \dots, t, \dots, z \in \mathbb{D} \quad |\{i', j, k, \dots, t, \dots, z\}| = |\mathbb{D}|
\end{aligned}$$

In the next step we reorder the summation over i' according to the chosen cells. This means we divide the index set of i' into disjoint subsets. Using the definition of the one-site domains of eqn. 3.9 we define the set of the one-site domains \mathbb{C}_λ within the cell \mathcal{C}_λ as:

$$\mathbb{C}_\lambda = \{D_\lambda \in \mathbb{D} | \vec{R}_a \in \mathcal{C}_\lambda \forall \phi_a \in D_\lambda\} \quad (3.55)$$

where we used the definition of a cell and the mapping:

$$\phi_a \mapsto \vec{R}_a := \begin{pmatrix} \langle \phi_a | x | \phi_a \rangle \\ \langle \phi_a | y | \phi_a \rangle \\ \langle \phi_a | z | \phi_a \rangle \end{pmatrix} = \begin{pmatrix} x_a \\ y_a \\ z_a \end{pmatrix} \quad (3.56)$$

We note that translational symmetry with respect to the domains has to be required, in order to regroup the summations with respect to the cells.

$$\begin{aligned}
E_{\text{corr}}^{\text{total}} = & \sum_{\lambda=0}^{n-1} \left[\frac{1}{1!} \sum_{i \in \mathbb{C}_\lambda} \Delta \varepsilon_i + \frac{1}{2!} \sum_{i \in \mathbb{C}_\lambda} \sum_j \Delta \varepsilon_{ij} + \frac{1}{3!} \sum_{i \in \mathbb{C}_\lambda} \sum_{jk} \Delta \varepsilon_{ijk} + \dots \right. \\
& \left. + \frac{1}{|\{j, k, \dots, t\}|!} \sum_{i \in \mathbb{C}_\lambda} \sum_{jk\dots t} \Delta \varepsilon_{ijk\dots t} + \dots + \frac{1}{|\mathbb{D}|!} \sum_{i \in \mathbb{C}_\lambda} \sum_{jk\dots t\dots z} \Delta \varepsilon_{ijk\dots t\dots z} \right] \quad (3.57)
\end{aligned}$$

We get the energy per cell as:

$$\begin{aligned}
E_{\text{corr}}^{\text{cell}} = & \frac{1}{1!} \sum_{i \in \mathbb{C}_\lambda} \Delta \varepsilon_i + \frac{1}{2!} \sum_{i \in \mathbb{C}_\lambda} \sum_j \Delta \varepsilon_{ij} + \frac{1}{3!} \sum_{i \in \mathbb{C}_\lambda} \sum_{jk} \Delta \varepsilon_{ijk} + \dots \\
& + \frac{1}{|\{j, k, \dots, t\}|!} \sum_{i \in \mathbb{C}_\lambda} \sum_{jk\dots t} \Delta \varepsilon_{ijk\dots t} + \dots + \frac{1}{|\mathbb{D}|!} \sum_{i \in \mathbb{C}_\lambda} \sum_{jk\dots t\dots z} \Delta \varepsilon_{ijk\dots t\dots z} \quad (3.58)
\end{aligned}$$

Note that eqn. 3.58 yields the exact energy per cell of eqn. 3.42 of an infinite ideal solid if we take the limit $n \rightarrow \infty$. Since the limit $n \rightarrow \infty$ is equivalent to $|\mathbb{D}| \rightarrow \infty$ we get infinite summations over j, k, \dots in this case.

3.4.1 The Prefactors for Disjoint Cells

As a first case we analyze the prefactors of disjoint cells. The generalization to non-disjoint cells can be based upon the general considerations in this section, as we will see later on. In the first step we classify the indices of a given t -site domain d_μ by translation vectors.

$$d_\mu \in \mathcal{P}(\mathbb{D}) \text{ with } |d_\mu| = t$$

In order to do this classification we define the translational equivalence class \mathbb{E}_T of a domain d_μ with respect to the reference cell \mathcal{C}_0 according to:

$$\mathbb{E}_T := \text{all translations where at least one index of the } n\text{-site domain is in the reference cell}$$

$$\mathbb{E}_T = \{\hat{t} \in \hat{\mathbb{T}} \mid \hat{t} \circ d_\mu = d'_\mu \text{ with } \mathbb{X}_{\hat{t}} = d'_\mu \cap \mathcal{C}_0 \neq \emptyset\} \quad (3.59)$$

where we leave the representation of $\hat{t} \circ d_n$ unspecified. It will be defined in the general section on symmetry (section 3.6). With the translations of eqn. 3.59 we are able to define the sets $\mathbb{X}_{\hat{t}_\lambda}$ as set of domains within the reference cell. Now we are in the position to analyze the prefactors for an increment with respect to index and translational symmetry. If we start with the first element of the first set $\mathbb{X}_{\hat{t}_1}$ we have $(t-1)$ summation indices j, k, \dots in eqn. 3.58 (for the t -th order summation) since we fix i to the first domain in the cell \mathcal{C}_0 . For the other domains in $\mathbb{X}_{\hat{t}_1}$ we do the same considerations and obtain a total factor of $|\mathbb{X}_{\hat{t}_1}| \cdot (t-1)!$ for \hat{t}_1 as translation. This considerations hold for all other translations in \mathbb{E}_T , too. Therefore we get the total prefactor f_t for a t -site increment as:

$$f_t = \sum_{i \in \mathbb{E}_T} |\mathbb{X}_i| \cdot (t-1)! \quad (3.60)$$

Since we classified all indices of the t -site domain by translation vectors, we have:

$$\sum_{i \in \mathbb{E}_T} |\mathbb{X}_i| = t \quad (3.61)$$

Inserting eqn. 3.60 and eqn. 3.61 into the incremental expansion 3.58 we see that the prefactors $\frac{1}{t!}$ cancel with the translational symmetry and the index symmetry. We obtain the same result as in eqn. 3.53.

3.4.2 The Prefactors for non-Disjoint Cells

At this point we consider the case of symmetric cells like in figure 3.6:

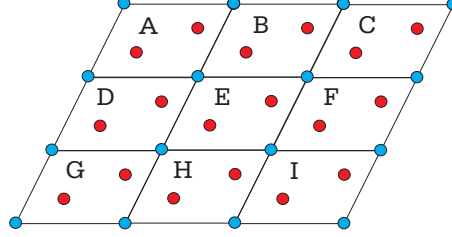


Figure 3.6: A 3×3 supercell built from a unit cell with atoms/centers of charge at the corners.

In this case we start again from eqn. 3.54 and split the sums over i' into a number of summations. But this time we order the terms according to single one-site domains:

$$\begin{aligned}
 E_{\text{corr}}^{\text{total}} &= \frac{1}{1!} \Delta \varepsilon_1 + \frac{1}{2!} \sum_j \Delta \varepsilon_{1j} + \frac{1}{3!} \sum_{jk} \Delta \varepsilon_{1jk} + \dots \\
 &+ \frac{1}{1!} \Delta \varepsilon_2 + \frac{1}{2!} \sum_j \Delta \varepsilon_{2j} + \frac{1}{3!} \sum_{jk} \Delta \varepsilon_{2jk} + \dots \\
 &+ \frac{1}{1!} \Delta \varepsilon_3 + \frac{1}{2!} \sum_j \Delta \varepsilon_{3j} + \frac{1}{3!} \sum_{jk} \Delta \varepsilon_{3jk} + \dots \\
 &\vdots
 \end{aligned} \tag{3.62}$$

This means we split up the incremental expansion for the whole system into a set of expansions: one for every one-site increment (every point in figure 3.6). At this stage we can easily introduce weight factors due to the symmetry of the cell in figure 3.6. We just multiply every expansion on these special points by their weight factor in the cell (e.g. $4 \cdot \frac{1}{4}$ for an edge domain). This is to insert a proper one according to the definition of the cell. Clearly, this leaves our total expansion unchanged. But now we are able to get the contributions for shared domains to the energy per cell. Collecting all contributions for a single unit cell using the weight factors for face, edge or corner

shared cells and taking the limit $n \rightarrow \infty$ we obtain:

$$\begin{aligned}
E_{\text{corr}}^{\text{cell}} = & \frac{1}{8} \sum_{i \in \mathbb{C}_c} \left[\frac{1}{1!} \Delta \varepsilon_i + \frac{1}{2!} \sum_j \Delta \varepsilon_{ij} + \frac{1}{3!} \sum_{jk} \Delta \varepsilon_{ijk} + \dots \right] \\
& + \frac{1}{4} \sum_{i \in \mathbb{C}_e} \left[\frac{1}{1!} \Delta \varepsilon_i + \frac{1}{2!} \sum_j \Delta \varepsilon_{ij} + \frac{1}{3!} \sum_{jk} \Delta \varepsilon_{ijk} + \dots \right] \\
& + \frac{1}{2} \sum_{i \in \mathbb{C}_f} \left[\frac{1}{1!} \Delta \varepsilon_i + \frac{1}{2!} \sum_j \Delta \varepsilon_{ij} + \frac{1}{3!} \sum_{jk} \Delta \varepsilon_{ijk} + \dots \right] \\
& + \sum_{i \in \mathbb{C}_{in}} \left[\frac{1}{1!} \Delta \varepsilon_i + \frac{1}{2!} \sum_j \Delta \varepsilon_{ij} + \frac{1}{3!} \sum_{jk} \Delta \varepsilon_{ijk} + \dots \right]
\end{aligned} \tag{3.63}$$

Here we partitioned the index set \mathbb{C}_0 of the reference cell into pairwise disjoint subsets:

$$\mathbb{C}_0 = \mathbb{C}_c \cup \mathbb{C}_e \cup \mathbb{C}_f \cup \mathbb{C}_{in}$$

with

\mathbb{C}_c one-site increments at the corners of the reference cell

\mathbb{C}_e one-site increments at the edges of the reference cell

\mathbb{C}_f one-site increments at the faces of the reference cell

\mathbb{C}_{in} all other one-site increments in the reference cell (the red points in figure 3.6)

Now we analyze the prefactor of every increment analogous to the previous considerations according to eq 3.60. The summation over the translational classes does not immediately yield t , because we shared domains. For a face-shared domain we get 2 translations, for an edge-shared domain we get 4 and for a corner-shared domain we get 8 translations in a 3-dimensional lattice. These factors cancel with the additional prefactors in eqn. 3.63 to one and we get the same result as in the case of disjoint cells.

3.5 Incremental Expansion for a Multi-reference Case

The straightforward extension of the incremental scheme to the multi-reference case reads:

$$E_{\text{corr}} = E_{\text{corr}}^{\text{active}} + \sum_i \Delta \varepsilon_i + \frac{1}{2!} \sum_{ij} \Delta \varepsilon_{ij} + \frac{1}{3!} \sum_{ijk} \Delta \varepsilon_{ijk} + \dots \tag{3.64}$$

where $E_{\text{corr}}^{\text{active}}$ is the correlation energy of the active space using the whole virtual space \mathbb{V} (for a truncation of \mathbb{V} consider 3.5.1). This means we have to build all excitations

from our active space into \mathbb{V} (up to the given excitation level of the method in use), and keep all other electrons frozen. The one-site increments are redefined as:

$$\Delta\varepsilon_i = \varepsilon_i - E_{\text{corr}}^{\text{active}}$$

$$\Delta\varepsilon_{ij} = \varepsilon_{ij} - \Delta\varepsilon_i - \Delta\varepsilon_j - E_{\text{corr}}^{\text{active}}$$

Since active space is included in every calculation, we introduced the additional terms in the expansion to ensure that we count the active space contribution to the correlation energy only once.

As a special case we discuss a single electron in a single orbital. In this case we have

$$E_{\text{corr}}^{\text{active}} = 0$$

because there is no "self-correlation". Thus the expansion eqn. 3.64 reduces to the standard incremental expansion eqn. 3.1. This can be used to expand the RCCSD correlation energy in open shell calculations. The difference to the conventional incremental expansion is that we have the active electron in every calculation.

3.5.1 Truncation of the Virtual Space

In order to speed up the calculation time, a truncation of the virtual space is desired. This can be achieved by a small modification of the $\Delta\varepsilon_i$ in eqn. 3.64.

$$\Delta\varepsilon'_i = \varepsilon_i - E_{\text{corr}}^{\text{active}}(\mathbb{V}_i)$$

$$\Delta\varepsilon'_{ij} = \varepsilon_{ij} - \Delta\varepsilon'_i - \Delta\varepsilon'_j - E_{\text{corr}}^{\text{active}}(\mathbb{V}_{ij})$$

$E_{\text{corr}}^{\text{active}}(\mathbb{V}_i) :=$ correlation energy of the active electrons with the subspace \mathbb{V}_i of \mathbb{V}

This modification ensures that the calculation is speeded up, even if the coupling of the virtual space with the active electrons is large.

3.6 Treatment of Symmetry

The treatment of symmetry is very important, since it can reduce the number of calculations significantly. Within the framework of the incremental scheme we can do this using the concept of equivalence classes. First we introduce the representation of a local orbital as a vector according to the center of charge eqn. 3.56 and define the set of symmetry operators \mathbb{G} of the point group \mathcal{G} . A symmetry operator \hat{O} transforms a vector \vec{v} into its symmetry equivalent vector \vec{v}' .

$$\hat{O}\vec{v} = \vec{v}'$$

Now we define an equivalence class \mathbb{W} of vectors:

$$\mathbb{W}_{\vec{v}_a} = \{\vec{v}'_a \in \mathbb{R}^3 | \hat{O}\vec{v}_a = \vec{v}'_a \text{ for all } \hat{O} \in \mathbb{G}\} \quad (3.65)$$

Since a domain D_λ usually contains a set of orbitals we need equivalence classes for sets of vectors. In order to get an analogous definition of eqn. 3.65 for a set of vectors \mathcal{D}_λ , we define the action of a symmetry operator on a set of vectors by the action of the symmetry operator on all elements of the set:

$$\hat{O}\mathcal{D}_\lambda = \mathcal{D}'_\lambda = \{\hat{O}\vec{v}_a | \vec{v}_a \in \mathcal{D}_\lambda\} \quad (3.66)$$

Going one step further we define the action of a symmetry operator on a set of sets of vectors as:

$$\hat{O}\{\mathcal{D}_{\lambda_1}, \dots, \mathcal{D}_{\lambda_n}\} = \{\mathcal{D}'_{\lambda_1}, \dots, \mathcal{D}'_{\lambda_n}\} = \{\hat{O}\mathcal{D}_{\lambda_1}, \dots, \hat{O}\mathcal{D}_{\lambda_n}\} \quad (3.67)$$

In order to have a clear notation we introduce $\tilde{\mathbb{D}}$ analogous to the set of all domains as set of all \mathcal{D}_λ . Using the definition eqn. 3.66 we can define the equivalence class \mathbb{E} for the representation of an arbitrary domain \mathbb{X} as ($\mathbb{X} \in \mathcal{P}(\tilde{\mathbb{D}})$):

$$\mathbb{E}_\mathbb{X} = \{\mathbb{X}' \in \mathcal{P}(\tilde{\mathbb{D}}) | \hat{O}\mathbb{X} = \mathbb{X}' \text{ for all } \hat{O} \in \mathbb{G}\} \quad (3.68)$$

For exactly symmetric local orbitals we find that the energy increments of domains within the same equivalence class are equal. For approximately symmetric orbitals they should be equal up to a given accuracy. We can reduce the number of calculations by calculating only one increment per equivalence class.

3.6.1 Symmetric One-Site Domains

In the above discussion we implicitly assumed that the one-site domains are disjoint and fulfill the symmetry requirements. This leads to some restrictions of the one-site increments. Firstly the one-site domains must be symmetry adapted to fulfill eqn. 3.66. Therefore we define the set of all vectors $\tilde{\mathbb{O}}$ which map to an occupied orbital in \mathbb{O} .

$$\tilde{\mathbb{O}} = \{\vec{R}_1, \vec{R}_2, \dots, \vec{R}_n\}$$

The symmetry condition can be fulfilled by dividing $\tilde{\mathbb{O}}$ into equivalence classes according to eqn. 3.65. In the second step we find the non-redundant part $\bar{\mathbb{O}}$ of the system by choosing one element of every equivalence class $\mathbb{W}_{\vec{v}_a}$. Since we want to have compact domains we choose the element of the $\mathbb{W}_{\vec{v}_a}$ where the distance to an element of $\bar{\mathbb{O}}$ is smallest. If we have the non-redundant part of the molecule we can use the symmetry operations of the group to construct the complete system again. Therefore we get automatically symmetric domains if we divide the non-redundant part and then apply the

symmetry operations to get the equivalent ones. In order to make sure that the domains are disjoint, we have to introduce another restriction, as we can see from the following example:

$$\begin{aligned}
\mathcal{D} &= \{\vec{v}_1, \vec{v}_2\} \\
\hat{O}\vec{v}_1 &= \vec{v}_1 \\
\hat{O}\vec{v}_2 &= \vec{v}'_2 \\
\hat{O}\mathcal{D} &= \{\vec{v}_1, \vec{v}'_2\} \\
\mathcal{D} \cap \mathcal{D}' &= \{\vec{v}_1, \vec{v}_2\} \cap \{\vec{v}_1, \vec{v}'_2\} \neq \emptyset
\end{aligned} \tag{3.69}$$

To ensure that one-site domains are disjoint after generating the symmetry equivalent one-site domains, we introduce the concept of site symmetry. We map to every point in $\bar{\mathbb{O}}$ the set of symmetry operators \mathbb{S} which leave this point invariant.

$$\mathbb{S}_{\vec{R}_i} = \{\hat{O}_j \in \mathbb{G} | \hat{O}_j \vec{R}_i = \vec{R}_i\} \tag{3.70}$$

Using eqn. 3.70 we can divide $\bar{\mathbb{O}}$ into a set of disjoint subsets \mathbb{T}_k with:

$$\mathbb{T}_k = \{\vec{R}_i \in \bar{\mathbb{O}} | \mathbb{S}_{\vec{R}_k} = \mathbb{S}_{\vec{R}_i}\} \tag{3.71}$$

Now we construct the domains \mathcal{D} within the sets \mathbb{T}_k and apply the symmetry operators to generate symmetry adapted one-site domains.

3.7 Error Analysis in the Incremental Expansion

A closer look at eq. 3.1 reveals that the sum of the energy increments at low order is included for several times at higher orders. In a system of $|\mathbb{D}|$ domains we have $\binom{|\mathbb{D}|}{2}$ two-body increments. A special one-site increment is included in the summation over the second-order for $(|\mathbb{D}| - 1)$ times. In general the prefactor p_{st} for the sum over the s -site increments included in the t -site increments can be determined according to:

$$p_{st} = \binom{(|\mathbb{D}| - s)}{(t - s)} \text{ with } t > s \tag{3.72}$$

The summation over a given order yields a constant error e . Considering eq. 3.72 we end up in a large accumulation of the error for higher order increments, since the prefactors increase quite fast and thus give a large weight to the errors of the sums in low orders.

We checked this propagation of errors in the incremental expansion by a numerical study. The results are displayed in figure 3.7. They are based on 2000 sets of uniformly

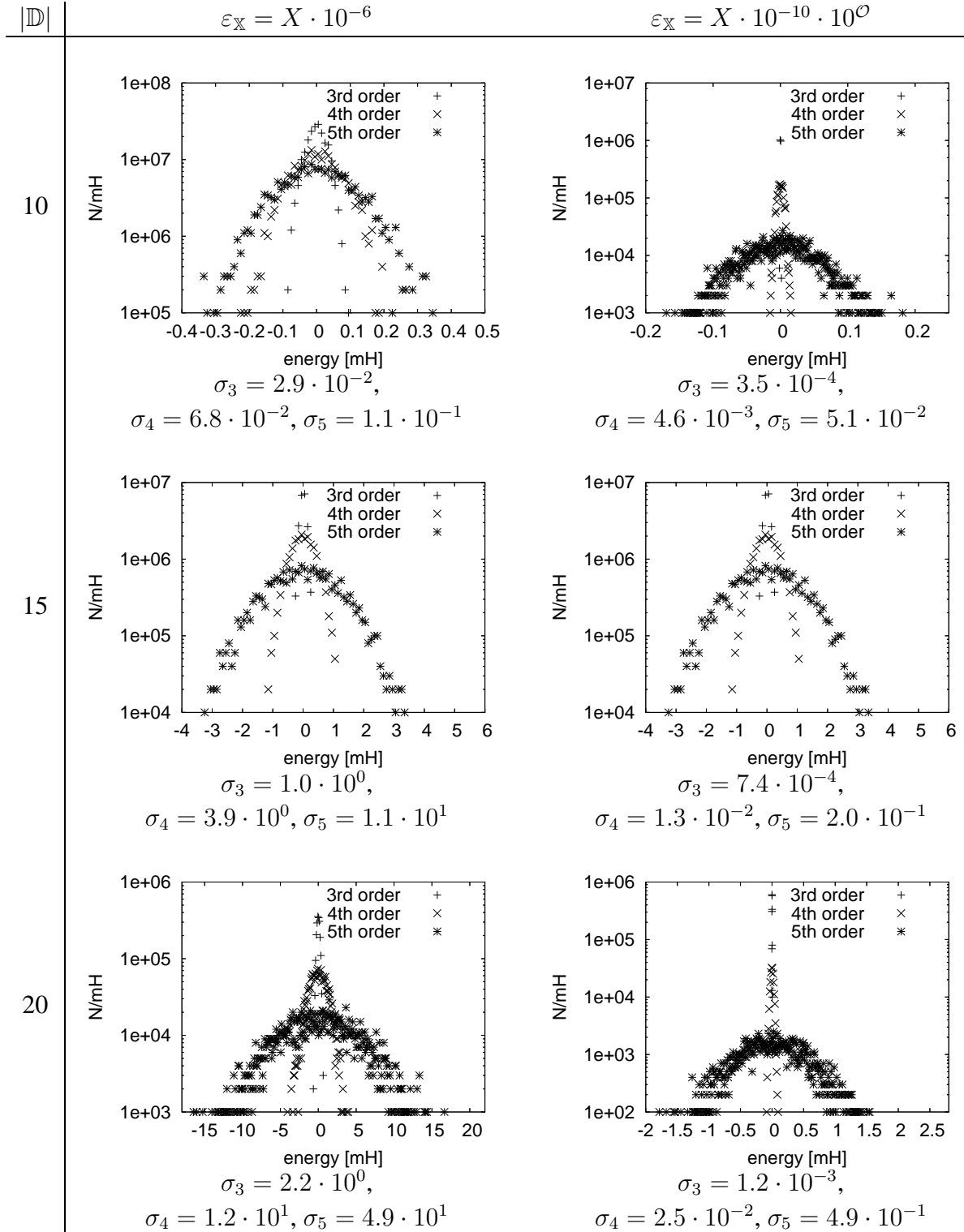


Figure 3.7: Error distribution of the incremental series for 10, 15 and 20 domains. The errors on the left are fixed to a constant value and are order-dependent on the right (see text). The standard deviation [mH] at i -th order is given as σ_i .

distributed random numbers X in the interval $[-0.5; 0.5]$. The left side of figure 3.7 was generated by multiplying the random number X by a constant factor of 10^{-6} (we use $X \cdot 10^{-6}$ for $\varepsilon_{\mathbb{X}}$ instead of the CCSD correlation energies). This models errors introduced by CCSD energies which are converged to a threshold of 10^{-6} Hartree. It can be seen that the errors increase with increasing number of domains and broader distributions around the exact results are obtained for increasing expansion order. Clearly, we can iterate the CCSD solutions of the fragments further and thus get more accurate correlation energies and smaller errors in the incremental expansion. However, one should note that at high order such a strategy might become quite expensive. Since the sums over the computationally cheap low-order increments are multiplied by large factors we have to iterate the corresponding correlation energies to a higher accuracy than those more expensive ones for higher orders. This strategy is simulated on the right hand side of figure 3.7 ($\varepsilon_{\mathbb{X}} = X \cdot 10^{-10} \cdot 10^{\mathcal{O}}$). The errors especially for larger numbers of domains $|\mathbb{D}|$ can be decreased significantly, although the correlation energies of the subsystems resulting from the unification of the largest number of domains (i.e. 20) are converged to a lower accuracy (i.e. 10^{-5} Hartree) than on the left hand side. Thus order-dependent convergence thresholds are an efficient way to obtain accurate incremental energies while reducing the cost of the most time consuming calculations to obtain high-order increments.

There are two different types of errors in the expansion at higher orders. The first one is the error caused by the multiplication of the low-order errors by large factors and the second one is the error caused by the summation over all increments of higher order. Note that in practical calculations both of these errors are reduced by using the truncation threshold R_{\min} . The first error is reduced, because the weight factors are lowered if we use only a subset of the higher-order increments. The second error is reduced, because we sum only over a smaller set of increments. Note that the truncation threshold R_{\min} introduces an error, too (chapter 6). Figure 3.8 shows the distribution of the incremental error for third and fourth order with 20 domains using an order dependent energy threshold for the fragmental CCSD energies. Both distributions are compact and therefore do not cause a loss of accuracy.

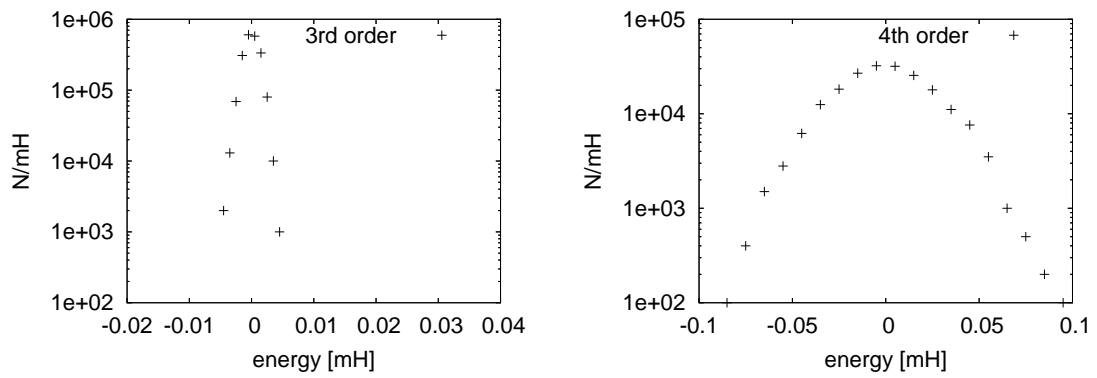


Figure 3.8: Error distribution of the incremental series with 20 domains using artificial order-dependent energies of $(\varepsilon_{\mathbb{X}} = X \cdot 10^{-10} \cdot 10^{\mathcal{O}})$.

Chapter 4

Automatization of the Incremental Scheme

In order to set up a fully automatized incremental scheme we have to construct the one-site domains automatically. After this step everything else is determined by set theory.

4.1 Obtaining Groups of Occupied Orbitals

The occupied orbitals are localized with a Foster-Boys procedure [3] using the algorithm of Edmiston and Ruedenberg [79]. For these localized orbitals we build the centers of charge from the diagonal elements of the dipole integrals in MO-basis.

$$\phi_a \mapsto \vec{R}_a := \begin{pmatrix} \langle \phi_a | x | \phi_a \rangle \\ \langle \phi_a | y | \phi_a \rangle \\ \langle \phi_a | z | \phi_a \rangle \end{pmatrix} = \begin{pmatrix} x_a \\ y_a \\ z_a \end{pmatrix}$$

Using the mapping of the Foster-Boys orbitals to their centers of charge, we are able to map our set of occupied orbitals \mathbb{O} onto a set of vectors. From this set of vectors we build the distance matrix D of all vector pairs. In the next step we construct the connectivity matrix C according to:

$$C_{ij} = \begin{cases} 10^8, & \text{if } D_{ij} \leq t_{con} \wedge \frac{w}{D_{ij}} \geq 10^8 \\ \frac{w}{D_{ij}}, & \text{if } D_{ij} \leq t_{con} \wedge \frac{w}{D_{ij}} < 10^8 \\ 0, & \text{if } D_{ij} > t_{con} \end{cases} \quad (4.1)$$

where t_{con} is a distance threshold and w is a constant stretching factor of 10^4 . The factor of 10^8 enters as an approximation of infinity in the regime of 32 bit integers. Since

METIS graph partitioning [83] needs the number of parts in advance, we introduce the domain size parameter (dsp) to control this number. Due to the fast increase of the calculation time it is more convenient to control the size of the domains than a given number of domains. We calculate the number of parts parameter (nop) for METIS according to:

$$\text{nop} = \frac{|\mathbb{O}|}{\text{dsp}} \quad (4.2)$$

Furthermore, since the graph partitioning requires integer variables we do a type cast from double to integer. The graph partitioning is done with the condition that the sum of the cut edge weights is minimal. According to the definition of our connectivity matrix C this forces close-lying orbitals to be in one domain. At the end we obtain local disjoint subsets of occupied orbitals. In the symmetric case we do the partitioning in the sets \mathbb{T}_k as discussed in the previous section.

4.1.1 Excitation Spaces for One-Site Domains

Our goal is to obtain a virtual space \mathbb{V}_{ϕ_a} for every occupied orbital ϕ_a :

$$\phi_a \mapsto \{\tilde{\phi}_i^{AO}\} \quad (4.3)$$

This is analogous to introducing excitation domains as in the well established local MPn methods of Pulay et al. [8] or Werner et al. [10]. Our virtual space is spanned by a set of projected atomic orbitals (PAOs) $\{\tilde{\phi}_i^{AO}\}$. The set of PAOs is constructed according to [8, 20]:

$$|\tilde{\phi}_i^{AO}\rangle = |\phi_i^{AO}\rangle - \sum_j^{occ} \langle \phi_j^{MO} | \phi_i^{AO} \rangle |\phi_j^{MO}\rangle \quad (4.4)$$

In local orbitals we recognize that an atomic orbital is important if its center is close to the center of charge of the MO. Guided by the spatial decay of the localized occupied orbitals we can restrict the excitation space of an occupied orbital according to eqn. 4.5.

$$\int (\phi_a^{MO} - \bar{\phi}_a^{MO})^2 d\tau \leq t_{dens} \quad (4.5)$$

We use the AO-representation of $\bar{\phi}_a^{MO}$ to find the most important AO-functions in ϕ_a^{MO} .

$$\begin{aligned} \phi_a^{MO} &= \sum_i c_i \phi_{ai}^{AO}(\vec{r}, \vec{\lambda}) \\ \bar{\phi}_a^{MO} &= \sum_i c_{h_i} \phi_{ah_i}^{AO}(\vec{r}, \vec{\lambda}) \end{aligned} \quad (4.6)$$

where $\vec{\lambda}$ is the shift vector of the AO-function. Since we usually have a set of AO-basis functions on the same center, we define the set $\mathbb{E}_{\vec{v}}$:

$$\mathbb{E}_{\vec{v}} = \{\phi^{AO}(\vec{r}, \vec{\lambda}) | \vec{\lambda} = \vec{v}\} \quad (4.7)$$

According to eqn. 4.7 we can divide the set of AO-functions into disjoint subsets. We can find the important AO-functions in ϕ_a^{MO} by successively using the sets $\mathbb{E}_{\vec{v}}$ where \vec{v} is closest to the center of charge of ϕ_a^{MO} until eqn. 4.5 is fulfilled. In order to get the ordering of the sets $\mathbb{E}_{\vec{v}}$ according to the distance to the center of charge of the a-th local MO we define the ordering relation as:

$$\left| \vec{R}_a - \vec{v}_s \right| > \left| \vec{R}_a - \vec{v}_t \right| \text{ for } s > t \quad (4.8)$$

Now we find the smallest n where eqn. 4.9 is fulfilled.

$$\int \left(\phi_a^{MO} - \sum_{i=1}^n \sum_{h \in \mathbb{E}_{\vec{v}_i}} \langle \phi_h^{AO} | \phi_a^{MO} \rangle \phi_h^{AO} \right) d\tau \leq t_{dens} \quad (4.9)$$

With this procedure we obtain a $\bar{\phi}_a^{MO}$. From eqn. 4.4 we identify the mapping:

$$\phi_i^{AO} \mapsto \tilde{\phi}_i^{AO} \quad (4.10)$$

The representation of $\bar{\phi}_a^{MO}$ contains a set of AO-functions which can be mapped to their corresponding PAOs according to eqn. 4.10. Thus we find a local excitation space \mathbb{V}_{ϕ_a} for the orbital ϕ_a^{MO} .

In order to obtain a local excitation space \mathbb{V}_i for our one-site domain i we have to unify the sets of PAOs which correspond to the occupied orbitals in the domain.

$$\mathbb{V}_i = \bigcup_{\phi_a \in D_i} \mathbb{V}_{\phi_a} \quad (4.11)$$

4.1.2 Construction of the n -Site Domains

The n -site domains are constructed using simple set theory. $\mathbb{O}_{D_1 D_2}$ is defined as the union of the occupied orbitals of D_1 (\mathbb{O}_{D_1}) with the occupied orbitals of D_2 (\mathbb{O}_{D_2}). The same holds for the virtual orbitals in D_1 and D_2 ($\mathbb{V}_{D_1 D_2}$). In general occupied and virtual spaces for the n -site domains are constructed according to:

$$\mathbb{O}_{\mathbb{X}} = \bigcup_{\lambda \in \mathbb{X}} \mathbb{O}_{\lambda} \quad (4.12)$$

$$\mathbb{V}_{\mathbb{X}} = \bigcup_{\lambda \in \mathbb{X}} \mathbb{V}_{\lambda} \quad (4.13)$$

$\mathbb{X} = \{D_{i_1}, D_{i_2}, D_{i_3}, \dots, D_{i_n}\} := \text{subset of } \mathbb{D} \text{ with cardinality } n$

$$\mathbb{X} \in \mathcal{P}(\mathbb{D}) \wedge |\mathbb{X}| = n$$

4.1.3 Obtaining Correlation Energies

In order to calculate the energies with a standard quantum chemistry package we have to account for the linear dependencies and the non-orthogonality in the PAO space. For this purpose we use a linear transformation which includes symmetric orthogonalization:

$$D^{-\frac{1}{2}}U^\dagger C^T S^{AO} C U D^{-\frac{1}{2}} = \mathbb{1} \quad (4.14)$$

with the MO coefficient matrix C , the overlap matrix in AO basis S^{AO} , the diagonal matrix $D = U^\dagger C^T S^{AO} C U$ and the Matrix \tilde{U} which diagonalizes $\tilde{S} = C^T S^{AO} C$. The matrix U is obtained by restricting \tilde{U} to those eigenvectors which correspond to an eigenvalue greater than 10^{-10} . Now we build the new MO matrix within the localized occupied orbitals and with the new orthogonalized linear independent PAOs. All occupied orbitals which are not in the set $\mathbb{O}_{\mathbb{K}}$ are frozen, and all virtual orbitals which are not in $\mathbb{V}_{\mathbb{K}}$ are deleted in the following CCSD calculation. Finally we obtain the total CCSD correlation energy according to eqn. 3.1.

4.2 Truncation by Distance

Up to now we constructed all possible combinations of one-site domains for a given order of the expansion. In order to get the correct scaling with respect to the number of calculations we introduce a truncation threshold according to section 3.1.2. We can reduce the question of the importance of a given n -site increment for the total energy to a question of graph connectivity. Let \mathbb{X} be a n -site domain, $\mathbb{V}_{\mathbb{X}}$ be a set of vectors with:

$$\mathbb{V}_{\mathbb{X}} = \{\vec{R}_a | \phi_a \in \bigcup_{\lambda \in \mathbb{X}} D_\lambda\} \quad (4.15)$$

Further we define the graph $\mathcal{G}(\mathbb{V}_{\mathbb{X}}, \mathbb{E})$, where the threshold R_{min} defines the adjacency of $\mathcal{G}(\mathbb{V}_{\mathbb{X}}, \mathbb{E})$:

$$\mathbb{E} = \{\{\vec{x}, \vec{y}\} | \vec{x}, \vec{y} \in \mathbb{V}_{\mathbb{X}} \wedge 0 < |\vec{x} - \vec{y}| < R_{min}\} \quad (4.16)$$

We neglect all n -site domains ($n > 1$) for which $\mathcal{G}(\mathbb{V}_{\mathbb{X}}, \mathbb{E})$ is disconnected. We note that eqn. 3.16 requires compact one-site domains, i.e. the orbitals in the one-site domains must be close in space.

Furthermore we implemented a dynamic distance threshold for the different orders \mathcal{O}_i of the expansion, in order to have more degrees of freedom for the truncation. Since high order increments are usually smaller than low order increments we use a distance truncation according to $\frac{f}{\mathcal{O}_i}$. In this case f is an adjustable parameter and we test the performance of this dynamic screening for several values of f .

Chapter 5

Implementation

5.1 Interfaces

We built interfaces to DALTON 2.0, MOLPRO 2002.6 and MOLCAS 6.4 quantum chemistry packages. For reasons of simplicity we modified the source code of the corresponding programs to obtain the overlap integrals in AO basis, the dipole integrals in AO basis and the MO coefficient matrix from a previous SCF calculation. We print the necessary numbers in double precision into the standard output and extract the data via an extraction class (e.g. `Molpro_extractor.H`, `Molcas_Overlap_Dipole_Interface.H/Molcas_6_4_INPORB_Interface.H` and for DALTON `Overlap_Dipol_Interface.H/MOPUN_Interface.H`). In MOLPRO calculations it is necessary to add some additional statements in order to get the desired information (see appendix B.3 for an example). Since MOLCAS uses the same format for the CASSCF orbital file, we could just rename the file and use the extraction module for the SCF orbitals in this case, too. Furthermore we built interfaces to the correlation codes of DIESEL [77] for, MR-CISD MR-ACPF, MR-AQCC and MR-CEPA(0), to the CCSD and the property modules of DALTON [88] and to the CCSD and RCCSD codes in MOLPRO [89].

5.2 Foster-Boys Localization

We generated three stand alone Foster-Boys programs, for MOLCAS, DALTON and MOLPRO respectively (`Molcas_6_4_FB`, `Dalton_FB`, `Molpro_FB`). This was necessary, since we have to use different extraction classes for the different codes. These modules produce an output with information of the convergence and store a file for the MO-coefficients which are read in by the corresponding server in the incremental calculation. The performance of the localization procedure is optimized by a reduced matrix multiplication in the transformation of the MO-coefficient matrix and in the

AO-MO transformation of the one particle integrals analogous to eqn. 5.2.

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \cdot \begin{pmatrix} E & F \\ G & H \end{pmatrix} = \begin{pmatrix} AE + BG & AF + BH \\ CE + DG & CF + DH \end{pmatrix} \quad (5.1)$$

for $D = \mathbb{1}$ and F, G a zero matrix we get:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \cdot \begin{pmatrix} E & 0 \\ 0 & \mathbb{1} \end{pmatrix} = \begin{pmatrix} AE & B \\ CE & D \end{pmatrix} \quad (5.2)$$

Eqn. 5.2 gives a sketch of the simplification which can be used to reduce the time for the successive 2×2 matrix multiplications drastically.

5.3 Parallelization

The parallelization is achieved by a server/client structure. It can be run with a script on the GRIDENGINE queuing system [90]. Before the clients are started, a file called Servers_HOSTNAME has to be generated. This file is copied to all slaves. It contains the name of the server which is necessary to establish the connection to the server via socket++. Using the GRIDENGINE framework has the advantage that a resource management by hand is not necessary, since the GRIDENGINE checks for available computers and starts the processes automatically if one or more nodes are free. Furthermore it is not necessary to start all slaves simultaneously, and therefore the job can start as soon as a node is available. Further slaves can start as soon as more nodes are free.

5.3.1 The Server

The server reads all information from an input file via cin (e.g. server < infile > outfile). Since we have many codes with different options we decided to write special servers for the different codes/correlation modules. The necessary input can be read in the first few lines of the corresponding source code as a variable declaration with comments.

The server collects all data and builds the symmetry adapted one-site domains. In the next step it builds all possible n -site domains up to a given order and divides them into equivalence classes. After the construction of this data structure the connection protocol is started, which is the real server. If a client connects, the server reads the job identification, the error information and the correlation energy of the last job. The correlation energy is arranged in the order of the job identification. If an error occurs the job is terminated. The server builds the MO-matrix for a given calculation on the

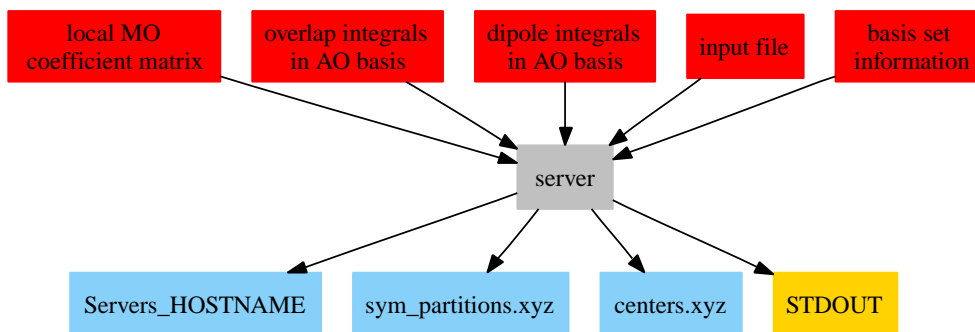


Figure 5.1: A sketch of the essential input data files for the incremental calculation. The output files include the information of the centers of charge (`centers.xyz`), the symmetry adapted redundant one-site domains (`sym_partitions.xyz`) and the node of the server (`Servers_HOSTNAME`). The output is written to standard output (`STDOUT`).

fly and sends it together with the job identification to the client. If no more jobs are available, the server sends the end signal to the client. After all jobs have run properly the server calculates the incremental correlation energy and ends.

5.3.2 The Client

The client connects to the server, sends the energy/error-wrapper to the server and asks for work. If there are still jobs to do, the client gets the necessary information as job identification and the MO-coefficient matrix by a wrapper class and writes the input file for the correlation code. In the next step it calls the correlation code by a system call. After the correlation calculation it extracts the correlation energy as well as the error information with an extraction class from the output of the correlation calculation. At this point it connects to the server again and starts the procedure again.

5.3.3 The Wrapper

The wrapper classes are data classes which can be written to a stream and constructed from a stream. Furthermore they contain an initial constructor from a set of data classes, which ensures the correct initialization. Therefore it is very convenient to use wrapper classes for data transfer in a stream-based parallel framework like `socket++`.

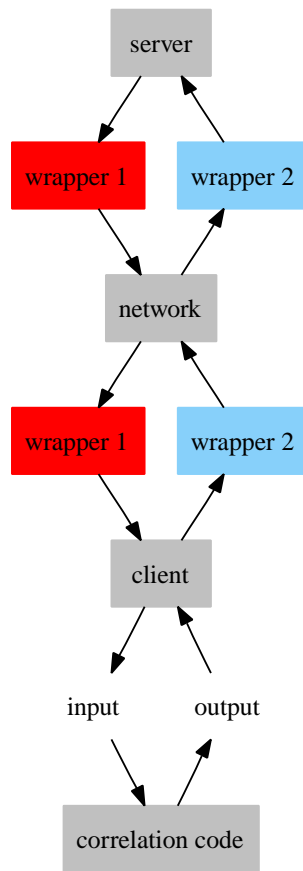


Figure 5.2: Data flow within the server/client structure.

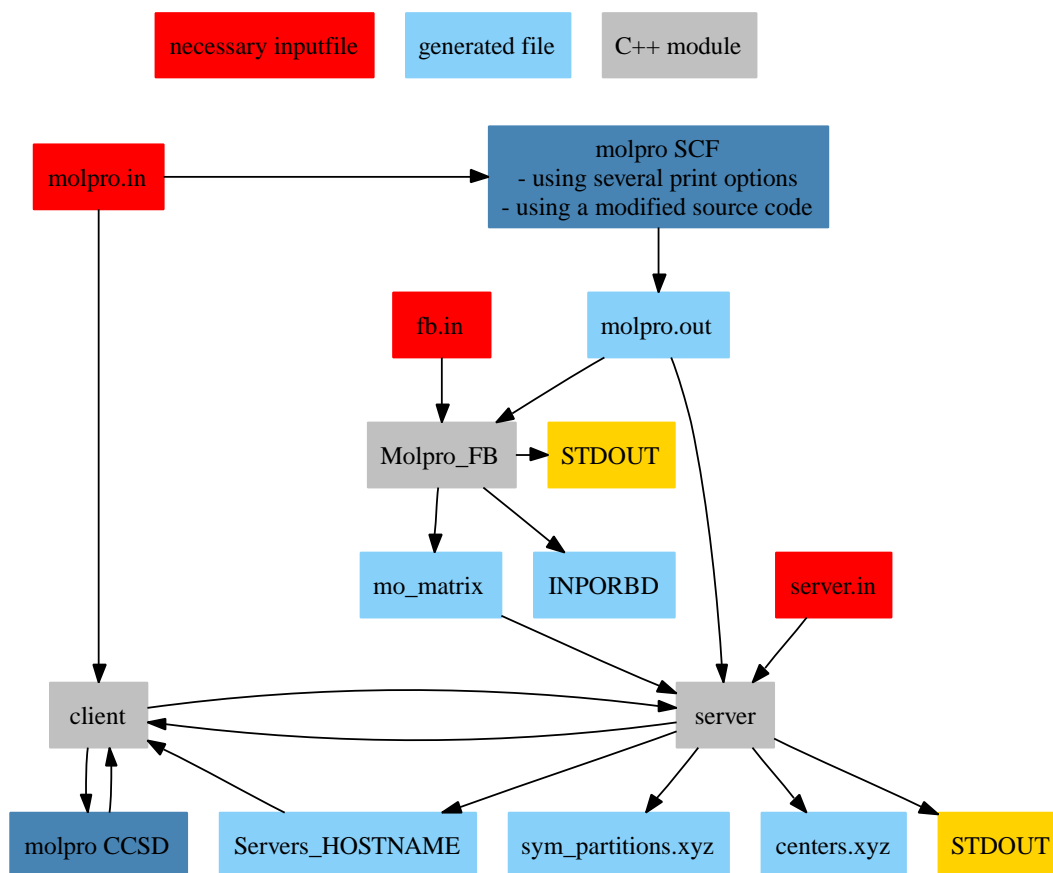


Figure 5.3: Data flow for an incremental calculation within the MOLPRO environment.

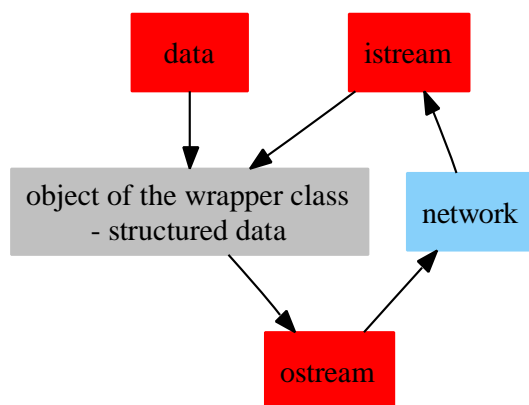


Figure 5.4: Visualization of the concept of a wrapper class as data container for the network transfer.

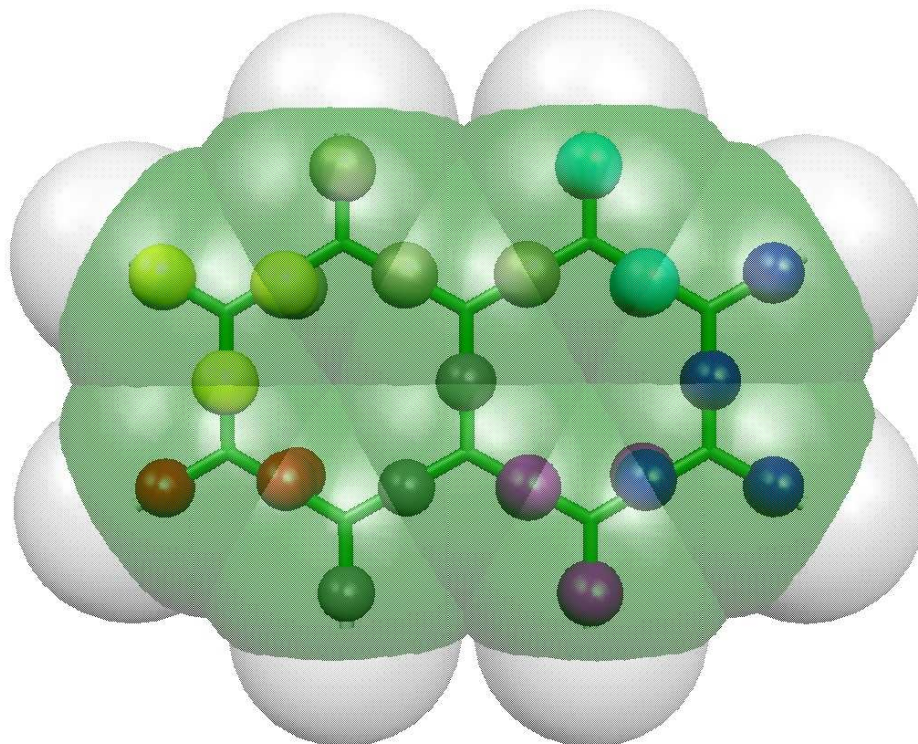


Figure 5.5: Centers of charge for naphthalene colored by their membership to the one-site domains.

5.4 Visualization of the Domains

The server writes a file called `sym_partitions.xyz`. It is formatted as a usual xyz-file for atoms and contains the centers of charge for the domains. The grouping to different domains is done by fictitious atom labels. This enables us to check the performance of the partitioning into one-site domains visually by standard programs like MOLDEN [91] or MOLEKEL [92, 93]. Figure 5.5 and 5.6 present the domains of naphthalene, where every sphere represents an occupied orbital by the center of charge. The double bonds can be seen clearly in figure 5.6 by the two spheres between two carbon atoms. For sigma bonds we find one sphere between two carbons or between a carbon and a hydrogen respectively. Note that the location of the centers of charge matches very well to the chemical intuition of the locations of the electrons in the molecule. The other resonance structure might be obtained by choosing a different threshold for the localization procedure. The locality of the one-site domains can be seen by the color of the centers of charge.

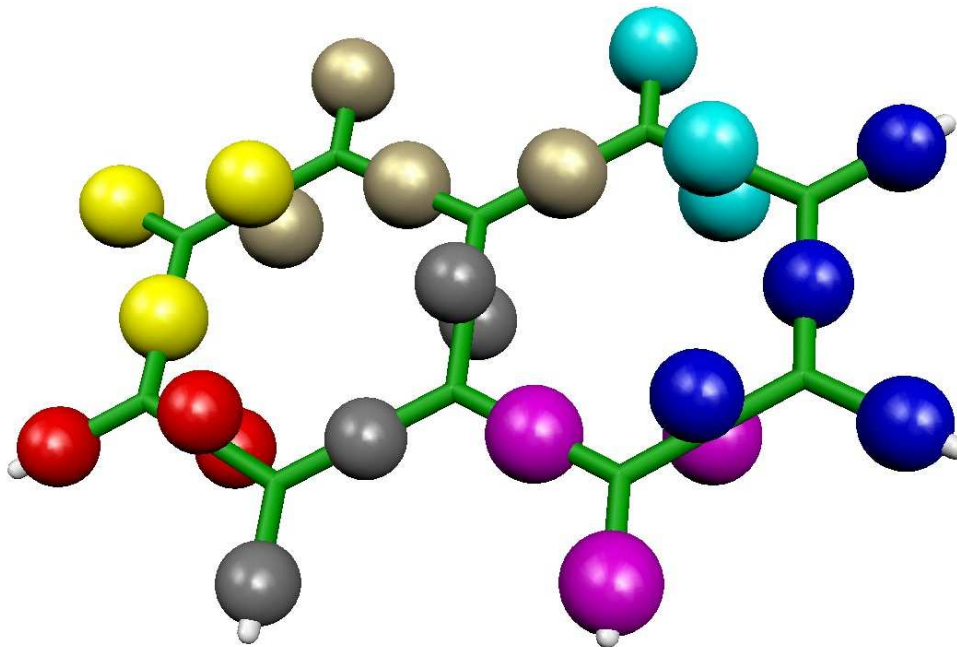


Figure 5.6: Centers of charge for naphthalene colored by their membership to the one-site domains.

5.5 Symmetry for Molecules

We use eqn. 3.68 to treat the symmetry of the n -site domains. As representation of the n -site domains we can use a set of vectors. This can be done using the STL vector or set classes. Both classes are optimized for performance and do not contain all possible access structures. The STL classes contain only those, which are efficient for a certain data structure. For reasons of convenience we swap between the different classes if some special features are desired. The basic ingredients for the treatment of symmetry are sets of symmetry operators and sets of vectors. In order to use the efficient set<T> of the STL we have to introduce a binary < relation. Since these are not defined mathematically for a matrix or a vector we define them artificially. The set<T> of the STL checks for the equality of two objects a , b of a class by the < operation of the class T according to:

$$(a < b) == \mathbf{false} \quad \wedge \quad (b < a) == \mathbf{false} \\ \Rightarrow a = b$$

This has to be done with respect to the numerical accuracy of matrix multiplication or matrix vector multiplication e.g. we have to introduce a threshold for the comparison. The source code for a specialization of the matrix class is given as an example:

```
1  template<>
2  bool Matrix<double>::operator < (Matrix<double> const &a) const
3  {
4  const double threshold=1E-12;
5  for (unsigned int i=0; i< (*this).size(); i++)
6  {
7  for (unsigned int j=0; j<(*this)[i].size(); j++)
8  {
9  if ( (*this)[i][j] < a[i][j] &&
10  (abs((*this)[i][j]-a[i][j])>threshold) )
11  {
12  return true;
13  }
14  else if ( ((*this)[i][j]>a[i][j]) &&
15  (abs((*this)[i][j]-a[i][j])>threshold) )
16  {
17  return false;
18  }
19  }
20 }
21 return false;
22 }
```

The implementation of the < operator for the class SortableVectorR3 is analogous to the implementation of the matrix class.

5.5.1 Generating the Operators of the Point Group

According to eqn. 3.68 we need all operators of the point group. This can be done by storing all information in a convenient way, or by generation of the operators from some basic operators. We decided to combine both possibilities in order to have no limiting restrictions. A tabulated point group may be called by the name of the point group. All other groups can be constructed by a list of operator names. The generation of the operators is implemented in the class All_Symmetry_Operations_Generator.


```

1  class All_Symmetry_Operations_Generator :
2  public set<Matrix<double> >
3  //
4  All_Symmetry_Operations_Generator(Point_Group pg);

```

The class Point_Group contains a set of generators and the name of the point group in Schoenflies notation.

```

1  class Point_Group
2  //
3      Point_Group(string point_group_symbol);
4  // Schoenflies nomenclature
5      Point_Group(string point_group_symbol,
6      vector<string> generators);
7  //
8      string _point_group_symbol;
9      vector<string> _generators;

```

In order to obtain a matrix representation of a symmetry operator we use the class Symmetry_element. It contains all information for the generation of the transformation matrix from a string. Note that the vector<T> of the STL is a one-dimensional data container without any mathematical property of a vector.

```

1  class Symmetrie_element : public Matrix<double>
2  //
3  Symmetrie_element(string symmetrie_Element);

```

The generation of the complete set of the symmetry operators is performed in two steps. First we generate the matrix representations of all symmetry elements in the class Point_Group and insert them into the inherited set<Matrix<double>> of an object of the class All_Symmetry_Operations_Generator. In the second step we multiply all members of this set with each other and insert all products into the set of the class. The constructor terminates, if the number of products is equal to the size of the object.

5.5.2 Symmetry Adapted One-Site Domains

The adaption of the one-site increments to symmetry can be done as follows: first we reduce the set of vectors to a set of non-redundant vectors.

```

1  class Non_Redundant : public set<SortableVectorR3>
2  //
3  Non_Redundant(
4      set<SortableVectorR3> coordinates,
5      All_Symmetry_Operations_Generator asog);

```

The class `Non_Redundant` groups the vectors into classes according to eqn. 3.65 and selects one vector of every group with the side condition that the selected vectors are close in space. The selected vectors are inserted into the `Non_Redundant` object during the construction. Next we classify the vectors by site symmetry according to eqn. 3.71 and do the partitioning into the domains. These steps are encapsulated in the class `Sym_Partitioning`.

```

1  class Sym_Partitioning : public vector<set<SortableVectorR3> >

```

In order to get all partitions, we apply the symmetry operators of the group onto every domain and get the set of redundant one-site domains. The set of the redundant one-site domains is represented by the class `Set_of_Redundant_domains`. Note that the data structure in the code is equivalent to the data structure of the mathematical representation:

$$\tilde{\mathbb{D}} = \{\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_n\}$$

$$\mathcal{D}_1 = \{\vec{R}_a, \vec{R}_b, \dots, \vec{R}_q\}$$

```

1  class Set_of_Redundant_domains :
2  public vector<vector<SortableVectorR3> >

```

5.5.3 Symmetry Classified n -Site Domains

For reasons of convenience we do the permutation management in C1 symmetry. Since every domain is associated with a large correlation calculation we did not observe a performance loss for this step. The class `Correlation_Energy_Container` is used to construct all combinations of one-site domains up to a given order. Furthermore it stores a double precision number for every permutation. The class `Incremental_Addresses` constructs all sets of vectors associated with the combinations of the one-site domains.

```

1  class Incremental_Addresses :
2  public vector<vector<vector<SortableVectorR3> > >

```

The first dimension refers to the order of the expansion, the second to the number of domains of a given order and the last one to the orbitals in a domain. Finally the symmetry classification is done by the class `Incremental_Addresses_equivalence_classes`. It uses the symmetry operators of the class to find the symmetry equivalent domains and stores them as `Equivalence_Class`.

```
1 class Incremental_Addresses_equivalence_classes :
2   public vector<vector<Equivalence_Class> >
```

Here the class `Equivalence_Class` contains all symmetry equivalent n -site domains as well as the energy and the truncation variable. We decided to generate several versions of the classes `Incremental_Addresses_equivalence_classes` and `Equivalence_Class` in order to keep the old versions of the code while generalizing the data container (templating the class).

```
1 class Equivalence_Class : public
2   map<Vector_Of_SortableVectorR3, unsigned int>
3   //
4   double _energy;
5   bool _calculate;
```

The truncation variable is a boolean which is set to true if the increment screening via the class `Adjustable_Truncation` yields the energy increment of the domain as significant and false otherwise. The server loops over all significant equivalence classes and writes the energy to them. The mapping to the orbitals of a domain is done according to the centers of charge:

$$\vec{R}_a \mapsto \phi_a$$

At the end the energies of the equivalence classes are back-transformed to C1 symmetry and written to an object of the class `Correlation_Energy_Container`. This enables us to use the same expansion class for C1 symmetry in the symmetric case, too. Again this is no significant loss of performance since the number of calculations dominates the total cpu time.

5.6 Symmetry for Periodic Systems

In order to treat all possible 3-dimensional space groups we build an interface to extract the matrix representation of the symmetry elements from the standard representation string of the International Tables for Crystallography [94]. For a convenient usage we downloaded all symmetry elements in the standard representation from the Bilbao crystallographic server [95]. This enables us to treat all translational groups, if the Foster-Boys procedure yields local symmetric orbitals within the desired accuracy.

The treatment of symmetry in periodic systems is somewhat different, since we cannot construct all symmetry operations of the group due to the infinite number of symmetry operations. We decided to construct all symmetry operations within the reference cell first, and then perform a predefined set of translations, in order to recover all symmetry operations within a given supercell. All other steps of the calculation are similar to the treatment of molecular symmetry.

5.7 Algorithm to Obtain the Virtual Space

The virtual space truncation as discussed in section 4.1.1 theoretically, was implemented according to the algorithm below.

$$\int (\phi_a^{MO} - \bar{\phi}_a^{MO})^2 d\tau \leq t_{dens} \quad (5.3)$$

Since we are using local orbitals, we can obtain $\bar{\phi}_i^{MO}$ in eqn. 5.3 in the following way:

- a. Initialize all coefficients of $\bar{\phi}_a^{MO}$ to zero
- b. Go to the center of charge \vec{R}_a of the a -th MO
- c. Find the AO-basis functions centered on \vec{P} closest to \vec{R}_a which have not been used so far
- d. Set all coefficients of $\bar{\phi}_a^{MO}$ which have an AO-function on \vec{P} to those of ϕ_a^{MO}
- e. End, if $\int (\phi_a^{MO} - \bar{\phi}_a^{MO})^2 d\tau$ is below t_{dens}
- f. Go to c.

With this procedure we obtain a $\bar{\phi}_a^{MO}$ which contains a set of AO-functions. These AO-functions can be mapped to their corresponding PAOs according to eqn. 4.10. Thus we find a local excitation space \mathbb{V}_{ϕ_a} for the orbital ϕ_a^{MO} .

In order to obtain a local excitation space \mathbb{V}_i for our one-site domain i we have to unify the sets of PAOs which correspond to the occupied orbitals in the domain. These mapping steps are performed in the class `Sym_Increment`. At this point we note that the virtual space is not symmetric in the current implementation. But this can be easily achieved by construction of the virtual space for the non-redundant part of the molecule and using the symmetry operators to obtain the symmetry-adapted virtual space. Since the large applications were done with MOLPRO we could not truncate the virtual space anyhow and for the applications to small molecules with DALTON or MOLCAS we were not able to save a significant part of the virtual space.

Chapter 6

Applications

6.1 Molecules

The geometries were obtained by optimization with the Becke-Perdew86 [96,97] gradient-corrected exchange-correlation functional (BP86) using density fitting and the SVP basis set of polarized double- ζ quality in the TURBOMOLE 5.6 [98] quantum chemistry package. Stationary points were characterized by force constant calculations.

6.1.1 Hydrocarbon Compounds

Table 6.1 shows the convergence behavior of the incremental expansion for the hydrocarbon compounds in figure 6.1. For molecules of this type we can truncate the expansion of the correlation energy at third-order while recovering the correlation energy almost exactly. An expansion up to second-order increments is already a good guess for the correlation energy whereas the first-order energy has a large error (for the chosen domain size). We emphasize that the error in the total correlation energy at third order is lower than 1 kcal/mol for sigma-bonded chains as well as for aromatic or conjugated compounds.

6.1.2 Transition Metal/Actinide Compounds

Transition metal complexes are a very important class of compounds in organic and inorganic chemistry. A huge number of modern homogeneous catalysts belong to this type of molecules. Since these compounds usually assemble a nearly spherical shape, they challenge local correlation methods. In table 6.2 we see that the convergence of the incremental series for the molecules in figure 6.2 depends on the type of the ligands. For the molybdenum fluorine complex we obtain very fast convergence. In

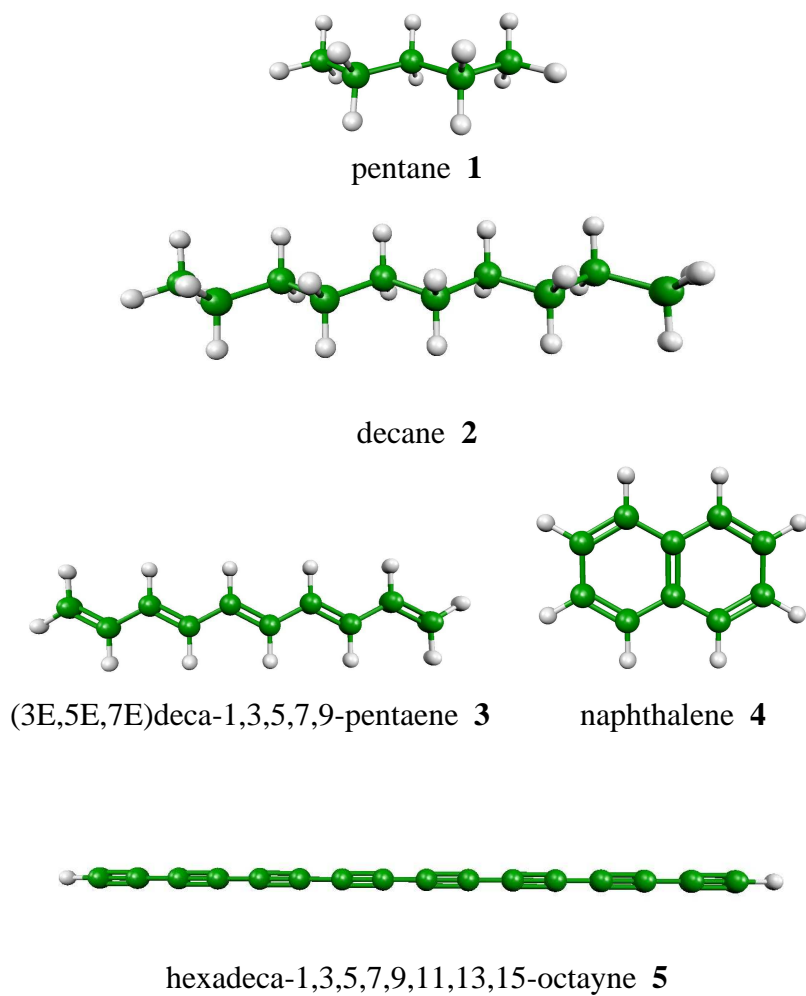


Figure 6.1: RI-BP86/SVP optimized structures of some hydrocarbon molecules.

system	order i	i -th order correction [au]	$E_{\text{corr}}(i)$ [au]	error [kcal/mol]	% E_{corr}
pentane^a 1					
	1	-0.661436	-0.661436	89.89	82.20
	2	-0.145661	-0.807097	-1.52	100.30
	3	0.002399	-0.804697	-0.01	100.00
	4	0.000020	-0.804677	0.00	100.00
exact CCSD			-0.804677		
decane^a 2					
	1	-1.257413	-1.257413	204.12	79.45
	2	-0.331197	-1.588610	-3.71	100.37
	3	0.005823	-1.582787	-0.05	100.01
	4	0.000084	-1.582704	0.00	100.00
exact CCSD			-1.582702		
alkene^a 3					
	1	-1.052959	-1.052959	229.75	74.20
	2	-0.374002	-1.426960	-4.94	100.55
	3	0.007689	-1.419271	-0.12	100.01
	4	0.000203	-1.419069	0.01	100.00
exact CCSD			-1.419088		
naphthalene^a 4					
	1	-0.906710	-0.906710	278.87	67.11
	2	-0.461387	-1.368097	-10.65	101.26
	3	0.016755	-1.351341	-0.14	100.02
	4	0.000467	-1.350875	0.15	99.98
exact CCSD			-1.351119		
alkyne^a 5					
	1	-1.399036	-1.399036	395.63	68.93
	2	-0.650050	-2.049085	-12.28	100.96
	3	0.018755	-2.030331	-0.52	100.04
	4	0.000753	-2.029578	-0.04	100.00
exact CCSD			-2.029509		

Table 6.1: Comparison of the incremental energies with the full CCSD calculations for the hydrocarbons in figure 6.1.

^a dsp=3. All calculations in the 6-31G** basis set of Pople and coworkers [99, 100].

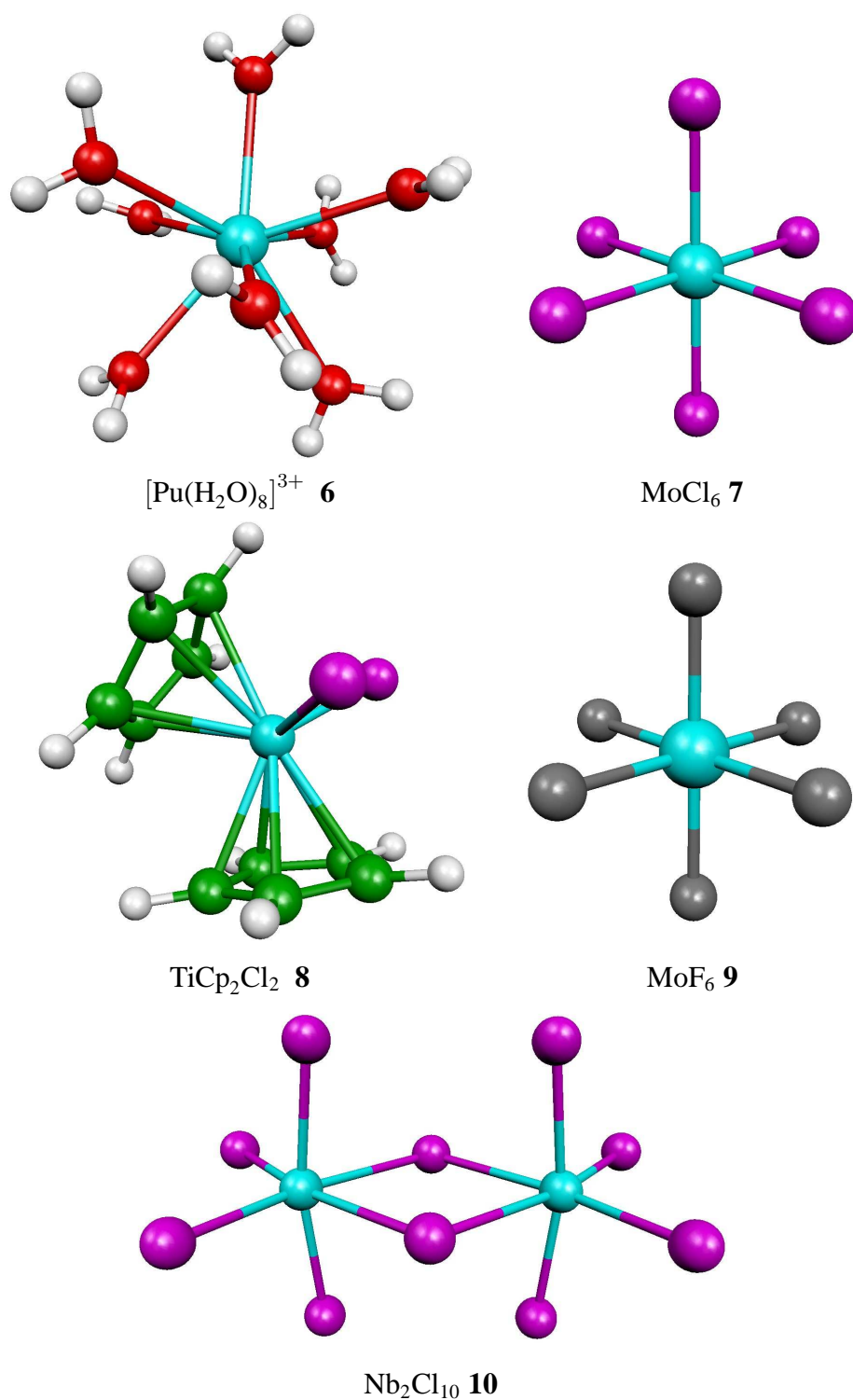
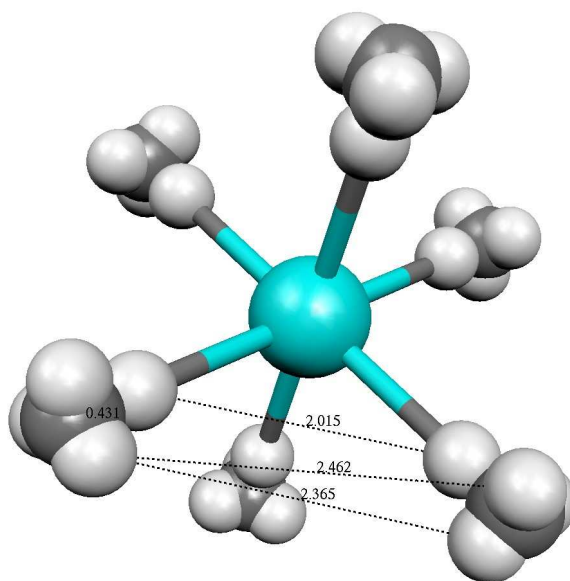


Figure 6.2: TiCp_2Cl_2 **8**, MoF_6 **9**, MoCl_6 **7** and $\text{Nb}_2\text{Cl}_{10}$ **10** were optimized with the RI-BP86/SVP in TURBOMOLE. $[\text{Pu}(\text{H}_2\text{O})_8]^{3+}$ was taken from [101].

the other cases we need fourth order increments in order to achieve convergence to 0.1 % of the correlation energy. We note that in these cases the sum of the fourth-order contributions is still quite large. The source of this behavior is that individual energy increments are still quite large at this order. In the case of $\text{Nb}_2\text{Cl}_{10}$, MoCl_6 and TiCp_2Cl_2 we find still large contributions of the adjacent fourth order increments ($\approx 10^{-4}$ Hartree). In MoF_6 for example the fourth order increments are in the order of 10^{-5} Hartree (see appendix C.1). Therefore we think that the large polarizability of the Cl^- ligands is responsible for the slow convergence behavior. This is supported by the fact, that the centers of charge in the MoF_6 complex are closer on each F^- and closer to the neighboring fluorine atoms than the corresponding centers of charge in MoCl_6 (figure 6.3 and 6.4). The larger distance of the centers of charge on a Cl atom is the reason why it is easier to polarize Cl^- than F^- . This effect can be seen in the incremental energies of MoF_6 and MoCl_6 . Our incremental calculations on TiCp_2Cl_2 account for 99.99% of the correlation energy at fourth-order. This behavior does not change significantly if different domains are chosen, as we can see from table 6.5. Furthermore we can see from this table that the convergence of the incremental expansion is good also for small domain sizes. We want to point out that it is possible to obtain accurate correlation energies before the spanned space of the orbitals in the domains approaches the full space.

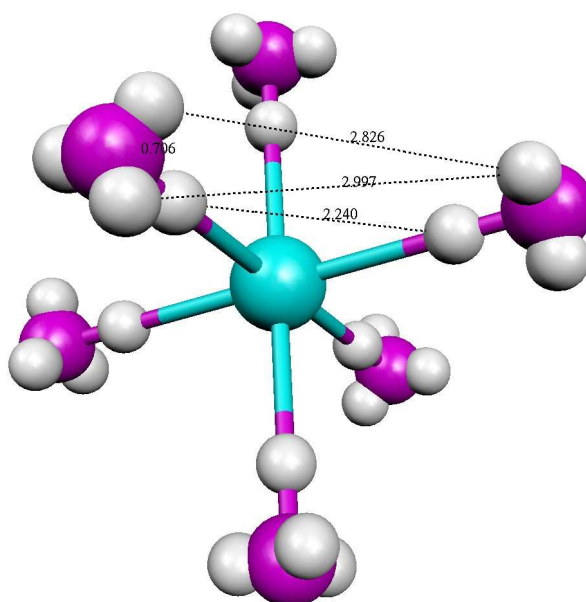
Again we find that the first-order approximation yields an inaccurate correlation energy (for the chosen domain size), the second-order is a good guess and the third-order has a small error, whereas we obtain excellent agreement at fourth-order level. The convergence of the incremental correlation energy for the $[\text{Pu}(\text{H}_2\text{O})_8]^{3+}$ **6** complex is quite fast. The second-order yields the correlation energy within an error of only -0.82 kcal/mol. The third-order level of increments yields the correlation energy almost exactly. This convergence behavior does not change significantly if we increase the basis set from double- ζ to triple- ζ , as we can see from table 6.4. Note that the canonical CCSD calculation cannot be done in the larger basis with the same code on the same machines. Therefore we conclude that the incremental scheme provides a way to reduce the hardware requirements of a large calculation, without significant loss of accuracy.

For the gold halogenide in figure 6.3 we find a fast convergence of the incremental series, too. The expansion up to third-order yields almost the exact correlation energy in this case (table 6.3).



MoF₆ 9

Figure 6.3: Centers of charge of MoF₆. The white spheres represent the centers of charge.



MoCl₆ 7

Figure 6.4: Centers of charge of MoCl₆. The white spheres represent the centers of charge.

system	order i	i -th order correction [au]	$E_{\text{corr}}(i)$ [au]	error [kcal/mol]	% E_{corr}
<hr/>					
MoF ₆ ^{a†}					
	1	-1.214064	-1.214064	24.62	96.87
	2	-0.043326	-1.257390	-2.57	100.33
	3	0.004799	-1.252591	0.44	99.94
	4	-0.000805	-1.253396	-0.06	100.01
exact CCSD			-1.253299		
<hr/>					
MoCl ₆ ^{a†}					
	1	-0.983501	-0.983501	46.86	92.94
	2	-0.090874	-1.074375	-10.17	101.53
	3	0.021664	-1.052711	3.43	99.48
	4	-0.006767	-1.059478	-0.82	100.12
exact CCSD			-1.058176		
<hr/>					
TiCp ₂ Cl ₂ ^{a#}					
	1	-1.360373	-1.360373	305.25	73.66
	2	-0.541651	-1.902023	-34.64	102.99
	3	0.067304	-1.834719	7.60	99.34
	4	-0.011983	-1.846702	0.08	99.99
exact CCSD			-1.846825		
<hr/>					
Nb ₂ Cl ₁₀ ^{b#}					
	1	-1.648100	-1.648100	78.78	92.92
	2	-0.153063	-1.801164	-17.27	101.55
	3	0.035460	-1.765703	4.98	99.55
	4	-0.009367	-1.775070	-0.90	100.08
exact CCSD			-1.773638		
<hr/>					
[Pu(H ₂ O) ₈] ³⁺ 6 ^{a#}					
	1	-1.887521	-1.887521	31.53	97.41
	2	-0.051540	-1.939062	-0.82	100.07
	3	0.001256	-1.937806	-0.03	100.00
	4	0.000059	-1.937747	0.01	100.00
exact CCSD			-1.937762		

Table 6.2: Comparison of the incremental energies with the full CCSD calculations for the molecules in figure 6.2.

^a dsp=3, ^b dsp=4

[†] 6-31G* basis set of Pople and coworkers [99, 100], [#] cc-pVDZ basis set of Dunning [102]

Mo, Ti, Nb, ECP28MWB [103, 104] Pu, ECP83MWB [105]

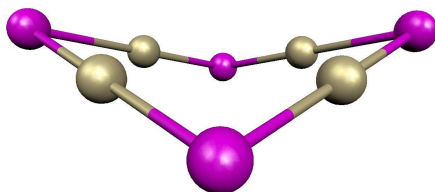


Figure 6.5: RI-BP86/SVP optimized structure of Au_4Cl_4 **11** (D_{2d}).

system	order i	i -th order correction [au]	$E_{\text{corr}}(i)$ [au]	error [kcal/mol]	% E_{corr}
Au_4Cl_4	1	-0.924218	-0.924218	84.98	87.22
	2	-0.141936	-1.066154	-4.09	100.61
	3	0.006637	-1.059517	0.08	99.99
	4	-0.000065	-1.059582	0.04	99.99
exact CCSD			-1.059642		

Table 6.3: Comparison of the incremental energies with the full CCSD calculations for Au_4Cl_4 **11** in figure 6.5.

(dsp=2, 9 domains, core=36, 6-31G** basis set of Pople and coworkers [99, 100])

order i	i -th order correction	E_{corr} [au]
1	-2.239121	-2.239121
2	-0.059162	-2.298283
3	0.001435	-2.296848

Table 6.4: The $[\text{Pu}(\text{H}_2\text{O})_8]^{3+}$ complex of figure 6.2 in a triple- ζ basis (H,cc-pVTZ sp; O,cc-pVTZ spd; Pu,ECP83MWB [105]). We note that the standard calculation of this type is already infeasible with the same code on the same computer.

order i	i-th order correction [au]	$E_{\text{corr}}(i)$ [au]	error [kcal/mol]	% E_{corr}
dsp=3				
$t_{\text{con}}=2.5$				
9 domains				
1	-1.360373	-1.360373	305.25	73.66
2	-0.541651	-1.902023	-34.64	102.99
3	0.067304	-1.834719	7.60	99.34
4	-0.011983	-1.846702	0.08	99.99
dsp=4				
$t_{\text{con}}=3.5$				
8 domains				
1	-1.421872	-1.421872	266.66	76.99
2	-0.473666	-1.895537	-30.57	102.64
3	0.059733	-1.835805	6.92	99.40
4	-0.010099	-1.845904	0.58	99.95
dsp=6				
$t_{\text{con}}=3.5$				
5 domains				
1	-1.524918	-1.524918	202.00	82.57
2	-0.346342	-1.871260	-15.33	101.32
3	0.028312	-1.842948	2.43	99.79
4	-0.004008	-1.846956	-0.08	100.01
exact CCSD		-1.846825		

Table 6.5: Comparison of the convergence for different dsp and t_{con} in the case of TiCp_2Cl_2 .

6.2 Intermolecular Interactions

6.2.1 Water Clusters

For the correlation energy of systems with intermolecular interactions such as water clusters we find that the incremental scheme performs even better than for the hydrocarbon compounds or transition metal complexes discussed before and in reference [85]. In the 6-31G** basis one obtains for $(\text{H}_2\text{O})_8$ (figure 6.6) already 99.99% of the correlation energy at second-order. The convergence behavior is similar for the calculations in the larger cc-pVTZ basis where the full calculation is infeasible on less than 1.35 GB machines (table 6.6). Comparing the two calculations in the 6-31G** basis we find that the convergence is slightly slower for the 11 domains. But the error is lower than 1 kcal/mol at second-order of the incremental expansion for this case too. This shows that the result depends on the choice of the domains, but not as much as may be expected. So we conclude that the incremental scheme is quite robust with respect to the choice of the one-site domains.

In table 6.7 we compare for $(\text{H}_2\text{O})_8$ the convergence behavior of the incremental scheme with respect to the density parameter t_{dens} of eqn. 4.5. Naturally the parameter affects the convergence behavior of the series, if the excitation space is restricted too rigorously. With proper values of this parameter we can still obtain fast convergence in the series as we can see for the last two examples in table 6.7.

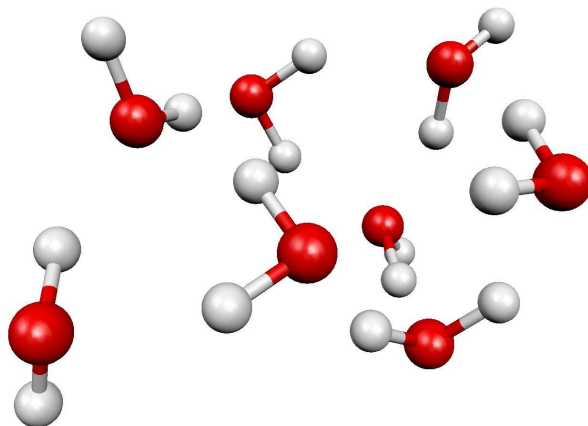


Figure 6.6: RI-BP86/SVP optimized structure of a set of eight water molecules.

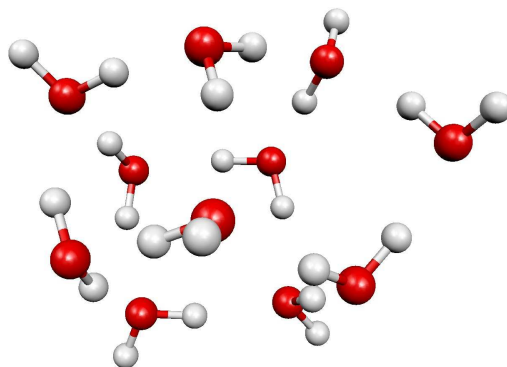


Figure 6.7: $(\text{H}_2\text{O})_{11}$ cluster taken from Bulusu et al. [106].

basis	order i	i -th order correction [au]	$E_{\text{corr}}(i)$ [au]	error [kcal/mol]	% E_{corr}
6-31G** ^a	1	-1.648410	-1.648410	35.61	96.67
	2	-0.056524	-1.704934	0.14	99.99
	3	-0.000207	-1.705142	0.01	100.00
	4	-0.000007	-1.705149	0.00	100.00
exact CCSD			-1.705151		
6-31G** ^b	1	-1.441864	-1.441864	165.21	84.56
	2	-0.264439	-1.706303	-0.72	100.07
	3	0.001113	-1.705190	-0.02	100.00
	4	0.000044	-1.705147	0.00	100.00
exact CCSD			-1.705151		
cc-pVTZ ^a	1	-2.244489	-2.244489		
	2	-0.077810	-2.322298		
	3	0.000111	-2.322188		

Table 6.6: Comparison of the incremental energies for the $(\text{H}_2\text{O})_8$ -cluster in figure 6.6 with the full CCSD calculations and comparison of the basis set effect.

^a 8 domains, core=0

^b 11 domains, core=0

Table 6.8 and 6.9 show for $(\text{H}_2\text{O})_{11}$ (figure 6.7) in 6-31G** basis the performance of an energy screening procedure using a dynamic distance threshold $R_{\text{min}}(\mathcal{O})$ and

(H ₂ O) ₈	order i	i -th order correction [au]	$E_{\text{corr}}(i)$ [au]	error [kcal/mol]	% E_{corr}
density threshold 0.1					
	1	-1.573940	-1.573940	32.17	96.85
	2	-0.053823	-1.627763	-1.60	100.16
	3	-0.000571	-1.628335	-1.96	100.19
exact CCSD			-1.625212		
density threshold 0.01					
	1	-1.575988	-1.575988	30.89	96.97
	2	-0.053820	-1.629808	-2.88	100.28
	3	0.004666	-1.625142	0.04	100.00
exact CCSD			-1.625212		
density threshold 0.001					
	1	-1.578597	-1.578597	29.25	97.13
	2	-0.046443	-1.625040	0.11	99.99
	3	-0.000053	-1.625093	0.07	99.99
exact CCSD			-1.625212		

Table 6.7: Convergence behavior of the incremental scheme with respect to the density parameter t_{dens} of eqn. 4.5. (dsp=5, 8 domains, core=0, calculation in 6-31G* basis set of Pople and coworkers [99] at the RI-BP86/SVP geometry.)

$N_{\text{calc}}/\text{total } N_{\text{calc}}$	f	order i	i-th order correction [au]	$E_{\text{corr}}(i)$ [au]	error [kcal/mol]	% E_{corr}
11/11	12	1	-2.235023	-2.235023	33.35	97.68
16/55		2	-0.049996	-2.285019	1.98	99.86
27/165		3	-0.003239	-2.288258	-0.05	100.00
0/330		4	0.000000	-2.288258	-0.05	100.00
54/561						
11/11	16	1	-2.235023	-2.235023	33.35	97.68
41/55		2	-0.052651	-2.287674	0.32	99.98
31/165		3	-0.000246	-2.287920	0.16	99.99
53/330		4	-0.000443	-2.288363	-0.12	100.01
136/561						
11/11	∞	1	-2.235023	-2.235023	33.35	97.68
55/55		2	-0.052874	-2.287897	0.18	99.99
165/165		3	-0.000260	-2.288157	0.01	100.00
330/330		4	-0.000025	-2.288182	0.00	100.00
561/561						

Table 6.8: Performance of the approximation of the incremental scheme with respect to a dynamic distance threshold $R_{\text{min}}(\mathcal{O}_i) = \frac{f}{\mathcal{O}_i}$ for the CCSD/6-31G** energy of $(\text{H}_2\text{O})_{11}$. (11 domains, core=11)

an energy threshold, respectively. The number of calculations can be reduced significantly, if we use the above distance truncation, as we can see from table 6.8. If we compare the reduction of calculations according to an energy threshold from table 6.9 with the performance of the truncation based on graph theory, we see that they behave quite similar. Note that we used the exact energies to simulate the energy screening. If approximate values are used it would be necessary to use a lower threshold in order to avoid discarding too many relevant contributions due to the approximation error. Table 6.10 presents the saving of the cpu time for $(\text{H}_2\text{O})_{11}$. For this example we find that we can reduce the calculation time for the incremental calculation to 5-25 % of the full incremental calculation. Furthermore the ratio between the number of calculations of the distance approximated incremental calculation and the full incremental calculation gives an upper bound to the computational time compared to the time for the full incremental calculation.

$$\frac{\mathcal{N}_{\text{calc}}(R_{\text{min}}, \mathcal{O})}{\mathcal{N}_{\text{calc}}(\mathcal{O})} > \frac{\text{cpu-time}(R_{\text{min}}, \mathcal{O})}{\text{cpu-time}(\mathcal{O})} \quad (6.1)$$

This is in general true, because $R_{min}(\mathcal{O})$ removes more high-order contributions, which are more time consuming to evaluate than the low-order contributions.

E_{thres}	ncal	E_{corr}	error	% E_{corr}
10^{-8}	513	-2.288182	0.00	100.00
10^{-7}	269	-2.288179	0.00	100.00
10^{-6}	130	-2.288167	0.01	100.00
10^{-5}	70	-2.287968	0.13	99.99
10^{-4}	37	-2.286425	1.10	99.92

Table 6.9: Performance of the approximation scheme eqn. 3.26 for the CCSD/6-31G** energy of the $(\text{H}_2\text{O})_{11}$ cluster due to an energy selection. For simplicity we used the exact incremental energies to estimate the individual contributions of the increments.

f	∞	16	12
% Ncalc	100	24	10
% cpu	100	19	5

Table 6.10: Performance of the dynamic distance threshold $R_{min}(\mathcal{O}_i) = \frac{f}{\mathcal{O}_i}$ with respect to the computational saving for $(\text{H}_2\text{O})_{11}$.

6.2.2 π - π /CH- π -Interactions

π - π -interactions are very important, since they can also affect the structure of DNA and proteins. The benzene dimer chosen here as a model of π - π -interactions was studied by several groups [107–113]. Another important intermolecular interaction is the CH- π -interaction [114–116]. For our purposes we have chosen the indole methane complex as studied by Ringer et al. [116].

For the intermolecular interactions between two benzene molecules (figure 6.8) we obtain at third-order level almost the exact CCSD energy (table 6.11). For the CH- π -interaction in figure 6.9 we obtain a reasonable energy at third-order level and almost the exact CCSD energy at fourth-order level. We point out that we have fast convergence for the compact 6-31G** basis as well as for the diffuse 6-31++G** basis set.

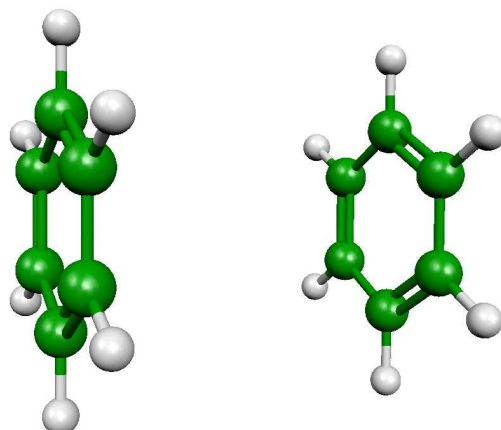


Figure 6.8: Sandwich structure of the benzene dimer at a benzene-benzene distance of 3.75 Å using the monomer C-H and C-C distances of Gauss and Stanton [117].

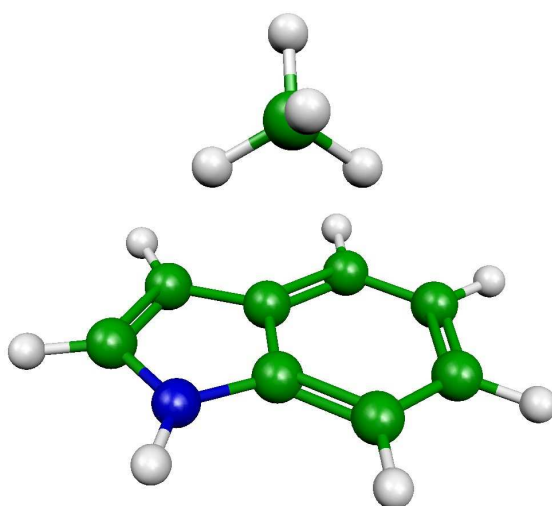


Figure 6.9: Indole-methane complex optimized by Ringer et al. [116].

system	order i	i -th order correction [au]	$E_{\text{corr}}(i)$ [au]	error [kcal/mol]	% E_{corr}
<hr/>					
$(\text{C}_6\text{H}_6)_2^a$					
	1	-1.100083	-1.100083	345.59	66.64
	2	-0.573180	-1.673263	-14.09	101.36
	3	0.022393	-1.650871	-0.04	100.00
	4	0.000185	-1.650686	0.08	99.99
<hr/>					
exact CCSD			-1.650808		
<hr/>					
indole-methane complex ^b					
	1	-1.012639	-1.012639	274.42	69.84
	2	-0.459677	-1.472315	-14.03	101.54
	3	0.022988	-1.449327	0.39	99.96
	4	-0.000546	-1.449874	0.05	99.99
<hr/>					
exact CCSD			-1.449955		

Table 6.11: Comparison of the incremental energies with the full CCSD calculations for the molecules in figure 6.8 and 6.9.

^a 6-31G** basis set of Pople and coworkers [99, 100] (10 domains, core=12)

^b 6-31++G** basis set of Pople and coworkers [99, 100] (8 domains, core=10)

6.2.3 The Auophilic Attraction

In order to check the performance of the presented approach for the auophilic intermolecular interaction we chose the test molecule $\text{Au}_2(\text{PH-C}_2\text{H}_2\text{-S})_2$ in figure 6.10 which was studied previously by Mendizabal and Pyykkö [118] ($d^{10}-d^{10}$ -interaction). Compared to the other systems in this study we find a relatively slow convergence. We still have a considerably large error for both domain sizes in table 6.12 at the third-order level e.g. the introduced error is 2 and 1 kcal/mol for the small and the large domains respectively. The convergence is somewhat faster for larger domains. At fourth-order level we are very close to the exact CCSD energy for both domain sizes. We note that an analysis of the single energy contributions analogous to table 6.9 shows that we could reduce the number of calculations significantly without loss of accuracy.

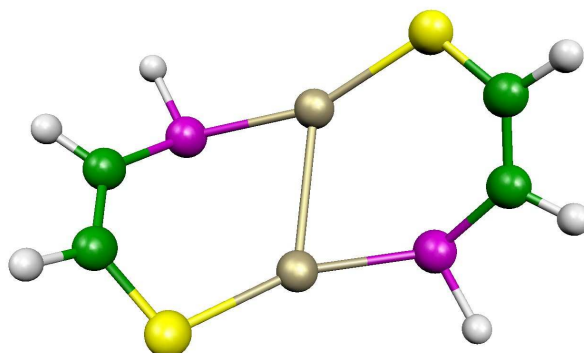


Figure 6.10: MP2 optimized structure of $\text{Au}_2(\text{PH-C}_2\text{H}_2\text{-S})_2$ using the 6-31G** basis set of Pople and coworkers [99, 100] for P,C,H,S and ECP60MDF/(8s6p5d)/[7s3p4d] [119] for Au. In order to polarize the d-shell of the gold atoms we added two f-type polarization functions ($\alpha_f = 0.20, 1.19$) [118].

$\text{Au}_2(\text{PH-C}_2\text{H}_2\text{-S})_2$					
order i	i-th order correction [au]	$E_{\text{corr}}(i)$ [au]	error [kcal/mol]	% E_{corr}	
dsp 3					
1	-0.904022	-0.904022	317.17	64.14	
2	-0.553462	-1.457485	-30.13	103.41	
3	0.051170	-1.406315	1.98	99.78	
4	-0.003270	-1.409585	-0.07	100.01	
dsp 5					
1	-1.111903	-1.111903	186.72	78.89	
2	-0.318708	-1.430611	-13.27	101.50	
3	0.022591	-1.408020	0.91	99.90	
4	-0.001396	-1.409416	0.03	100.00	
exact CCSD		-1.409467			

Table 6.12: Comparison of the incremental energies with the full CCSD calculations for $\text{Au}_2(\text{PH-C}_2\text{H}_2\text{-S})_2$ in figure 6.10 using the 6-31G** basis set of Pople and coworkers [99, 100] for P,C,H,S and ECP60MDF/(8s6p5d)/[7s3p4d] [119] for Au. (dsp=3, 10 domains, core=32; dsp=5, 6 domains, core=32).

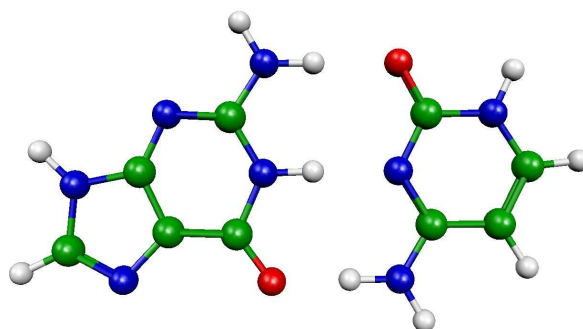


Figure 6.11: RI-BP86/SVP optimized guanine-cytosine base pair.

6.2.4 DNA Base Pair

The accurate calculation of the guanine-cytosine base pair in figure 6.11 is very important to benchmark lower level methods like force fields or DFT, since highly accurate wavefunction-based methods are not applicable to large DNA molecules. From a theoretical point of view these systems are interesting because of their great number of π -electrons. Since a localization of conjugated π -orbitals yields usually relatively extended localized orbitals, we checked the performance of our local approach for this case. The results of the incremental calculations are given in table 6.13. We find that a truncation according to $\frac{16}{\mathcal{O}_i}$ yields quite accurate results for this system. At third-order level we obtain 100.11% of the correlation energy whereas we obtain 100.03 % at fourth-order level. This corresponds to an absolute error of only -0.46 kcal/mol with respect to the exact CCSD energy. Note that the full CCSD calculation needs 2.5 GB of RAM and 18.5 GB of disk space. The incremental calculations however took at most 1.4 GB of RAM and 7.4 GB of disk space.

This example demonstrates again the power of the automatic distance truncation, because the ratio $\frac{N_{calc}(R_{min}, \mathcal{O})}{N_{calc}(\mathcal{O})}$ is 0.17, i.e. 2095 CCSD calculations for $f=16$ at fourth-order are avoided.

6.2.5 Reaction Pathways of the 4-*exo*/5-*endo* Cyclization

The convergence of the incremental RCCSD correlation energies was checked for a chemical example taken from a collaboration with the experimental group of Gansäuer at the University of Bonn. The geometries were obtained by RI-BP86/TZVP geometry optimizations with the TURBOMOLE 5.6 program package [98]. The stationary points were characterized by a force constant analysis.

Figure 6.12 shows a schematic view of the competing 4-*exo* and 5-*endo* cyclizations.

guanine-cytosine base pair				
order i	i-th order correction [au]	$E_{\text{corr}}(i)$ [au]	error [kcal/mol]	% E_{corr}
1	-1.916622	-1.916622	592.65	66.99
2	-0.979382	-2.896004	-21.92	101.22
3	0.031816	-2.864189	-1.96	100.11
4	0.002382	-2.861807	-0.46	100.03
exact CCSD		-2.861067		

Table 6.13: Comparison of the incremental energies with the full CCSD calculations for the guanine-cytosine dimer in figure 6.11 using the 6-31G** basis set of Pople and coworkers [99, 100]. (dsp=3, 16 domains, core=19)

Figure 6.13 shows the convergence of the incremental RCCSD correlation energies of eqn. 3.64 for a set of intermediates in the 4-*exo*/5-*endo* cyclization. The complete data is given in the appendix C.3. The first-order incremental correlation energy has a large error whereas the second-order correlation energies are all in between 4.4 - 6.4 kcal/mol. This corresponds to a non-linear error of ca. 2 kcal/mol which is still too large for chemical accuracy. At third-order level we recover more than 99.9 % of correlation energy which corresponds to a largest error of 0.2 kcal/mol. The fourth-order corrections do not improve the third-order energies significantly, because the third-order was already sufficient to obtain the desired accuracy. This is demonstrated by figure 6.13, since the spread of the errors decreases very rapidly with increasing incremental order.

Comparing the relative incremental energies of the different intermediates in table 6.14 with respect to the order we find a better accuracy of the energies, because of a beneficial error cancellation. The first order incremental CCSD energies leads to inaccurate results for all points in table 6.14. The resulting error is about 2 kcal/mol for the second-order level and at third-order level we observe accurate relative CCSD energies. This finding was expected because the absolute energies were already accurate at third order level (cf. figure 6.13). The summed incremental CCSD/cc-pV(n)Z (n=D,T) correlation energies are given in appendix C.3. Comparing the energy contributions for a given order with respect to the basis set, we did not observe a large sensibility.

A comparison of the performance of the different quantum chemical methods in table 6.14 for the educts and the products yields similar energies for all applied post-HF methods. DFT agrees for the educt and the 4-*exo* product, whereas the 5-*endo* product is 3-4 kcal/mol higher in energy. Due to the fact that the HF results do not agree with the other methods, we find that electron correlation plays an important role in the

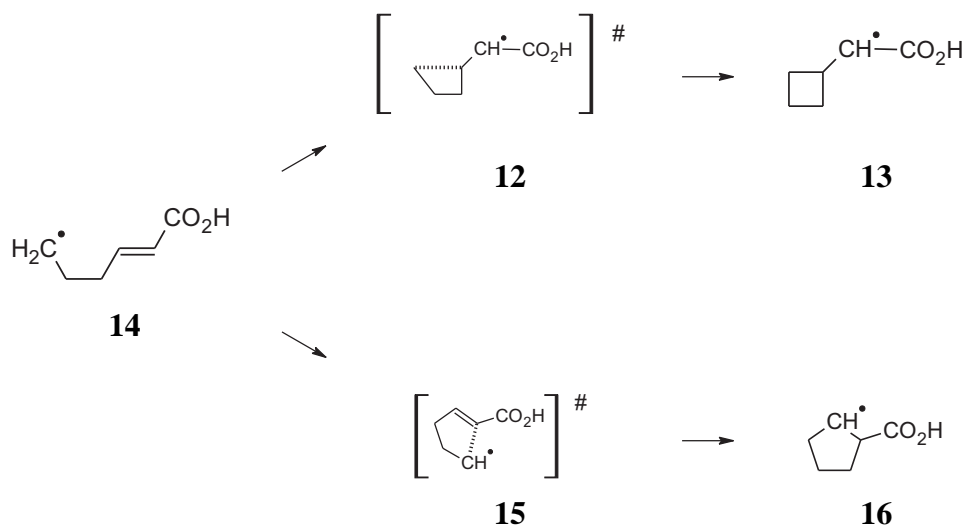


Figure 6.12: Intermediates of the 4-*exo*/5-*endo* cyclization.

quantitative description of the 4-*exo*/5-*endo* cyclization. Comparing the incremental results for the transition states, with those obtained by the MP2 method, we find in this case that the cheap second-order incremental expansion is much closer to the CCSD results (cc-pVDZ, approximate cc-pVTZ) as well as to the DFT result. Comparing the relative CCSD energies of the transition states with the DFT results we find a quite large discrepancy. From the 3-*exo* cyclization we would expect that the perturbative triples correction is quite important and shifts the coupled cluster results somewhat towards the DFT result [120]. Another aspect is the single point approximation in the CCSD calculations, which might also be in the order of a few kcal/mol. Since the CCSD/cc-pVDZ geometry optimizations were already too time consuming we were not able to check how the relative CCSD energies are affected by the usage of the DFT geometries for the CCSD calculations. Finally we conclude that the relative energies at second-order level are fairly accurate and agree better with the relative CCSD results than the MP2 results do.

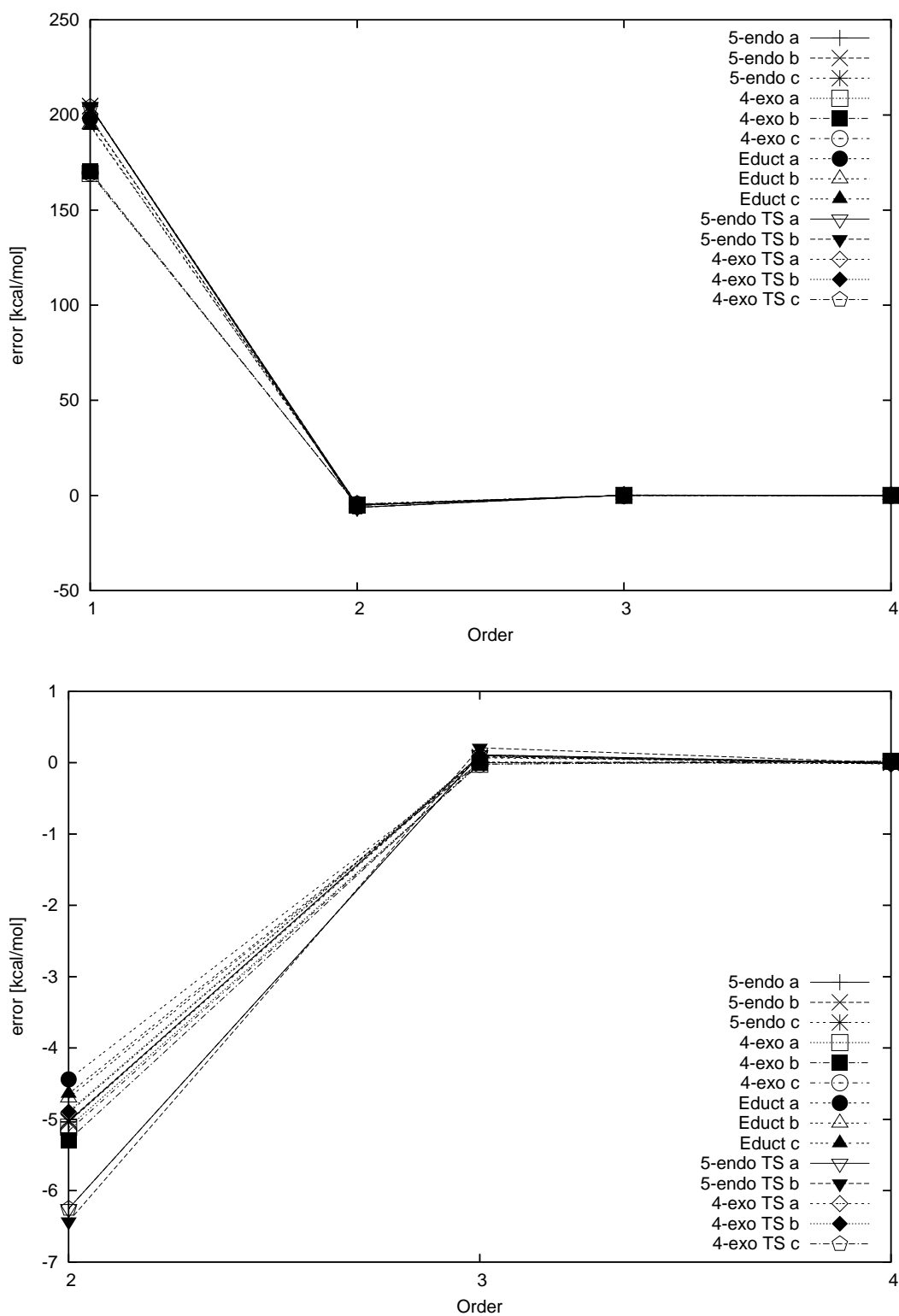


Figure 6.13: Convergence of the incremental correlation energies for intermediates of the 4-*exo* and 5-*endo* cyclization of **14**. In the second chart the first order was omitted in order to get a meaningful scale for the error.

method	Incremental CCSD							CCSD	RI-CC2	RI-MP2	RI-BP86	HF
	cc-pVDZ				cc-pVTZ			cc-pVDZ	TZVPP	TZVPP	TZVP	cc-pVTZ
basis												
molecule/order	1	2	3	4	1	2	3					
16 a	-7.9	-17.8	-17.4	-17.5	-5.3	-17.0	-16.6	-17.5	-18.2	-18.1	-14.2	-11.1
16 b	-7.1	-17.1	-16.7	-16.8	-4.7	-16.4	-16.1	-16.8	-17.6	-17.5	-14.0	-10.9
16 c	-6.9	-17.1	-16.7	-16.7	-4.6	-16.4	-16.1	-16.7	-17.6	-17.5	-14.0	-10.9
14 a	6.3	3.3	3.1	3.1	9.6	3.3	3.1	3.1	3.3	3.4	2.8	4.0
14 b	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
14 c	-0.3	0.4	0.3	0.3	-0.4	0.4	0.3	0.3	0.4	0.5	0.2	0.2
13 a	-25.4	-0.3	0.1	0.2	-32.0	0.1	0.3	0.1	-0.4	0.9	-1.3	6.7
13 b	-25.6	-1.5	-1.0	-0.9	-31.4	-0.9	-0.4	-1.0	-1.7	-0.4	-1.1	6.7
13 c	-26.7	-1.8	-1.4	-1.3	-33.1	-1.4	-1.1	-1.3	-2.0	-0.6	-2.6	5.1
TS 12 a	22.8	19.1	19.4	19.3	25.5	18.4	18.8	19.3	17.8	23.4	11.6	27.9
TS 12 b	23.7	19.6	19.9	19.8	26.9	19.3	19.7	19.8	18.1	24.1	12.2	30.1
TS 12 c	17.2	17.4	17.6	17.5	16.9	16.9	17.1	17.5	15.7	21.4	9.9	26.5
TS 15 a	30.7	19.6	21.3	21.2	31.5	19.6	21.1	21.2	20.1	26.1	17.1	33.5
TS 15 b	30.7	19.2	21.2	21.0	31.8	19.4	21.2	21.0	20.0	26.0	17.8	33.6

Table 6.14: Performance of different quantum chemical methods for the 4-*exo*/5-*endo* reaction energies relative to **14 b**. The geometries were optimized at RI-BP86/TZVP level. The ZPE was approximately included by the RI-BP86/TZVP value.

a,b,c are different conformers of the corresponding molecule

6.3 Symmetric Systems

6.3.1 A Water Cluster

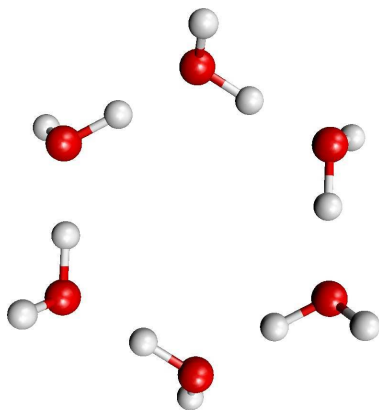


Figure 6.14: RI-BP86/SVP optimized structure of $(\text{H}_2\text{O})_6$ (C_3).

As a first example we choose the water aggregate of figure 6.14 as studied by Laasonen et al. [121], Kozmutza et al. [122], Upadhyay et al. [123] and Kim et al. [124]. The structure was optimized in C_3 symmetry with TURBOMOLE [98] using the RI-BP86/SVP method. We applied different truncation parameters f and different basis sets for this example. The results are given in tables 6.16-6.18. First we note that the convergence of the incremental scheme is very fast for all applied basis sets (6-31G** [99], aug-cc-pVDZ [102, 125] and cc-pVTZ [102]). The convergence behavior is not affected by the usage of the approximate symmetric Foster-Boys orbitals. Furthermore the convergence is not affected by the usage of the diffuse aug-cc-pVDZ basis set, if we compare the aug-cc-pVDZ results with the 6-31G** results. The usage of the cc-pVTZ basis does not change the convergence of the incremental series for this water cluster either. The errors increase a little, if we use the dynamic distance truncation $\frac{f}{\mathcal{O}_i}$ (see also section 4.2). We find that the errors are below 1 kcal/mol with respect to the canonical CCSD results at second-order level, for all chosen values of f (14, 16, 18, 22, 26, ∞) in all applied basis sets (6-31G**, aug-cc-pVDZ, cc-pVTZ). Note that the approximation by the parameter f has an effect on the total convergence of the incremental series. If we neglect incremental contributions of far distant domains, we can only get a finite accuracy. This can be seen if we compare the full incremental calculations up to fourth-order with the approximated calculations. The errors decrease up to a certain digit and are not improved further if the order of the expansion is increased. They improve, if the distance threshold is set to a larger value (table 6.18).

f	n calc	$\frac{\text{n calc}}{\text{total n calc}}$ [%]	error [kcal/mol] ^a
∞	20	36	0.00
16	10	18	-0.10
14	8	14	-0.10

Table 6.15: Comparison of the error introduced by the incremental expansion and saving with respect to symmetry and the distance truncation by $\frac{f}{\sigma_i}$. The total number of calculations in C_1 symmetry is $\sum_{i=1}^4 \binom{6}{i} = 56$.

^a the error of the incremental CCSD/aug-cc-pVDZ results at fourth-order of table 6.17.

For the $(\text{H}_2\text{O})_6$ cluster this limitation is not serious, because the error distribution of the CCSD method itself [68] is much larger than the error introduced by the distance truncation. On the other hand it is always possible to use a larger f in order to include more terms. The comparison of the introduced error due to the approximate treatment of symmetry and truncation due to the distance in table 6.15 shows that the introduced error is much smaller than the intrinsic CCSD error. Therefore we conclude that it is convenient and economic to use the symmetry for this class of systems. An interesting result is found in table 6.16 for the entries $f = 14$ and $f = 16$, since the sums up to third-order are equal up to the last digit whereas the sums up to second-order are not. We find the same for the analogous entries in table 6.17. An analysis of the single contributions yields that for $f = 14$ two site increments were neglected which were still included for the $f = 16$ case. This can be seen in the following example:

$$f = 14$$

$$\begin{aligned} \Delta\varepsilon_{01} &= e_{01}, \quad \Delta\varepsilon_{02} = e_{02}, \quad \Delta\varepsilon_{12} = 0 \\ \Delta\varepsilon_{012} &= e_{012} - e_{01} - e_{02} - e_0 - e_1 - e_2 \end{aligned} \quad (6.2)$$

$$f = 16$$

$$\begin{aligned} \Delta\varepsilon_{01} &= e_{01}, \quad \Delta\varepsilon_{02} = e_{02}, \quad \Delta\varepsilon_{12} = e_{12} \\ \Delta\varepsilon_{012} &= e_{012} - e_{01} - e_{02} - e_{12} - e_0 - e_1 - e_2 \end{aligned} \quad (6.3)$$

now we subtract the corresponding terms of the second-order sum for $f = 14$ from those of the second-order sum for $f = 16$:

$$e_{01} + e_{02} + 0 - (e_{01} + e_{02} + e_{12}) = -e_{12} \quad (6.4)$$

6-31G**	order i	i -th order correction [au]	$E_{\text{corr}}(i)$ [au]	error [kcal/mol]	% E_{corr}
$f = \infty$	1	-1.222751	-1.222751	22.77	97.12
	2	-0.036096	-1.258847	0.12	99.99
	3	-0.000168	-1.259015	0.01	100.00
	4	-0.000015	-1.259030	0.00	100.00
exact CCSD			-1.259031		
$f = 16$	1	-1.222751	-1.222751	22.77	97.12
	2	-0.036016	-1.258767	0.17	99.98
	3	-0.000149	-1.258916	0.07	99.99
	4	-0.000189	-1.259105	-0.05	100.01
exact CCSD			-1.259031		
$f = 14$	1	-1.222751	-1.222751	22.77	97.12
	2	-0.035307	-1.258058	0.61	99.92
	3	-0.000857	-1.258916	0.07	99.99
	4	-0.000189	-1.259105	-0.05	100.01
exact CCSD			-1.259031		

Table 6.16: Convergence of the incremental CCSD/6-31G** energies for a C_3 symmetric $(\text{H}_2\text{O})_6$ aggregate. Different values for the truncation parameters f were applied to check the convergence behavior of the incremental series with respect to a dynamic distance truncation.

If we do the same for the third-order summations we obtain:

$$\begin{aligned}
 & e_{012} - e_{01} - e_{02} - e_0 - e_1 - e_2 \\
 & - (e_{012} - e_{01} - e_{02} - e_{12} - e_0 - e_1 - e_2) = e_{12} \quad (6.5)
 \end{aligned}$$

Since the two terms in eqn. 6.4 and eqn. 6.5 cancel each other exactly in the summation, we get the same result at third-order level for both values of f . The sums for the fourth-order expansions are equal in both cases, because the same terms were dropped at fourth-order level and some special terms such as $\Delta\varepsilon_{123}$ were dropped at third-order level.

$$\begin{aligned}
 \Delta\varepsilon_{0123}^{14} - \Delta\varepsilon_{0123}^{16} &= \Delta\varepsilon_{012}^{14} - \Delta\varepsilon_{123}^{14} - \Delta\varepsilon_{12}^{14} \\
 &- (\Delta\varepsilon_{012}^{16} - \Delta\varepsilon_{123}^{16} - \Delta\varepsilon_{12}^{16}) = \Delta\varepsilon_{123}^{16} - \Delta\varepsilon_{123}^{14} \quad (6.6)
 \end{aligned}$$

aug-cc-pVDZ	order i	i -th order correction [au]	$E_{\text{corr}}(i)$ [au]	error [kcal/mol]	% E_{corr}
$f = \infty$	1	-1.360122	-1.360122	29.27	96.68
	2	-0.046486	-1.406608	0.10	99.99
	3	-0.000126	-1.406733	0.02	100.00
	4	-0.000025	-1.406759	0.00	100.00
exact CCSD			-1.406760		
$f = 16$	1	-1.360122	-1.360122	29.27	96.68
	2	-0.046323	-1.406445	0.20	99.98
	3	-0.000127	-1.406572	0.12	99.99
	4	-0.000343	-1.406914	-0.10	100.01
exact CCSD			-1.406760		
$f = 14$	1	-1.360122	-1.360122	29.27	96.68
	2	-0.045199	-1.405320	0.90	99.90
	3	-0.001252	-1.406572	0.12	99.99
	4	-0.000343	-1.406914	-0.10	100.01
exact CCSD			-1.406760		

Table 6.17: Convergence of the incremental CCSD/aug-cc-pVDZ energies for a C_3 symmetric $(\text{H}_2\text{O})_6$ aggregate. Different values for the truncation parameters f were applied to check the convergence behavior of the incremental series with respect to a dynamic distance truncation.

We discuss this special case in detail, because it is very unlikely to get exactly the same numbers without numerical noise at first glance.

Kozmutza et al. [122] analyze the main contributions in their water cluster based on MP2 data. An analogous analysis can be done within the framework of the incremental scheme at CCSD-level for n -body interactions, if the domains are chosen physically (every water molecule as a one-site domain). We point out that the incremental scheme is not limited to calculate the correlation energy, it can also provide a way to analyze the system by n -body interactions.

cc-pVTZ	order i	i -th order correction [au]	$E_{\text{corr}}(i)$ [au]	error [kcal/mol]	% E_{corr}
$f = \infty$	1	-1.599444	-1.599444	30.13	97.09
	2	-0.047846	-1.647290	0.11	99.99
	3	-0.000147	-1.647437	0.01	100.00
	4	-0.000021	-1.647458	0.00	100.00
exact CCSD			-1.647460		
$f = 26$	1	-1.599444	-1.599444	30.13	97.09
	2	-0.047846	-1.647290	0.11	99.99
	3	-0.000147	-1.647437	0.01	100.00
	4	-0.000015	-1.647452	0.01	100.00
exact CCSD			-1.647460		
$f = 22$	1	-1.599444	-1.599444	30.13	97.09
	2	-0.047846	-1.647290	0.11	99.99
	3	-0.000147	-1.647437	0.01	100.00
	4	-0.000015	-1.647452	0.01	100.00
exact CCSD			-1.647460		
$f = 18$	1	-1.599444	-1.599444	30.13	97.09
	2	-0.047846	-1.647290	0.11	99.99
	3	-0.000138	-1.647428	0.02	100.00
	4	-0.000024	-1.647451	0.01	100.00
exact CCSD			-1.647460		
$f = 16$	1	-1.599444	-1.599444	30.13	97.09
	2	-0.047716	-1.647161	0.19	99.98
	3	-0.000138	-1.647299	0.10	99.99
	4	-0.000282	-1.647581	-0.08	100.01
exact CCSD			-1.647460		

Table 6.18: Convergence of the incremental CCSD/cc-pVTZ energies for a C_3 symmetric $(\text{H}_2\text{O})_6$ aggregate. Different values for the truncation parameters f were applied to check the convergence behavior of the incremental series with respect to a dynamic distance truncation.

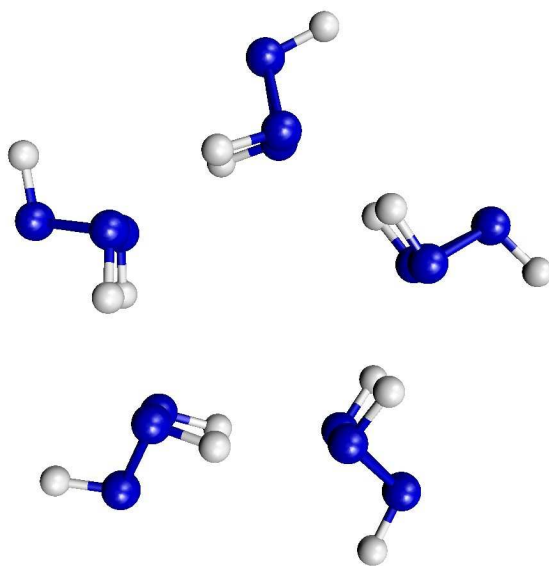


Figure 6.15: RI-BP86/SVP optimized structure of $(\text{N}_3\text{H}_3)_5$ (C_5).

6.3.2 Circular *Cis*, *Trans*-Cyclotriazine Cluster

The molecular structure of N_3H_3 in figure 6.15 was evaluated by several groups [126–129]. Since we want to check the performance of the incremental scheme with respect to symmetry, we choose the circular *cis*-, *trans*-cyclotriazine pentamer (C_5) as studied by Song et al. [130]. The geometry was optimized in C_5 symmetry using the RI-BP86/SVP method in TURBOMOLE 5.6 [98]. The stationary point was characterized as a minimum, by analyzing the Hessian matrix. Since the single molecules are rather large with 18 electrons per molecule to be correlated, if the 1s orbitals of N are considered as frozen core, we choose a dsp value which divides every molecule into two parts. The symmetry adaption for these parts was done according to the procedure in section 3.6.1. The focus of our partitioning was on the potential accuracy of the incremental scheme and on the efficiency, not on a physical basis. In order to analyze the n -body interactions of the molecules on the basis of n -body increments it might be more convenient to use complete N_3H_3 molecules as one-site increments. Note that such an analysis is similar to the analysis of n -body contributions by means of the supermolecular Møller-Plesset perturbation theory discussed by G. Chalasinski et al. [131, 132].

The convergence for the chosen partitioning is somewhat slower than for the symmetric water cluster before, but we still get accurate correlation energies at third-order level (table 6.19). Since we divided the N_3H_3 molecules into two parts, it is clear that the convergence is slower, because we have to account for the intramolecular corre-

6-31G**	order i	i -th order correction [au]	$E_{\text{corr}}(i)$ [au]	error [kcal/mol]	% E_{corr}
$f = \infty$	1	-2.118245	-2.118245	356.90	78.83
	2	-0.571134	-2.689380	-1.50	100.09
	3	0.002430	-2.686950	0.03	100.00
	4	-0.000055	-2.687005	0.00	100.00
exact CCSD			-2.686998		
$f = 16$	1	-2.118245	-2.118245	356.90	78.83
	2	-0.569853	-2.688098	-0.69	100.04
	3	-0.000308	-2.688406	-0.88	100.05
	4	0.000000	-2.688406	-0.88	100.05
exact CCSD			-2.686998		
$f = 14$	1	-2.118245	-2.118245	356.90	78.83
	2	-0.569853	-2.688098	-0.69	100.04
	3	0.001430	-2.686668	0.21	99.99
	4	0.000000	-2.686668	0.21	99.99
exact CCSD			-2.686998		

Table 6.19: Convergence of the incremental CCSD/6-31G** energies for a C_5 symmetric $(\text{N}_3\text{H}_3)_5$ aggregate. Different values for the truncation parameters f were applied to check the convergence behavior of the incremental series with respect to a dynamic distance truncation. (dsp=4, 10 domains, core=15)

lation corrections at second-order level. Due to the distance truncation the systematic convergence of the incremental series is slightly affected, as in the case of the water clusters. If we compare the obtained accuracy with respect to the approximate treatment of symmetry, we conclude that we have fast convergence for this system, too.

6.4 Cluster Compounds

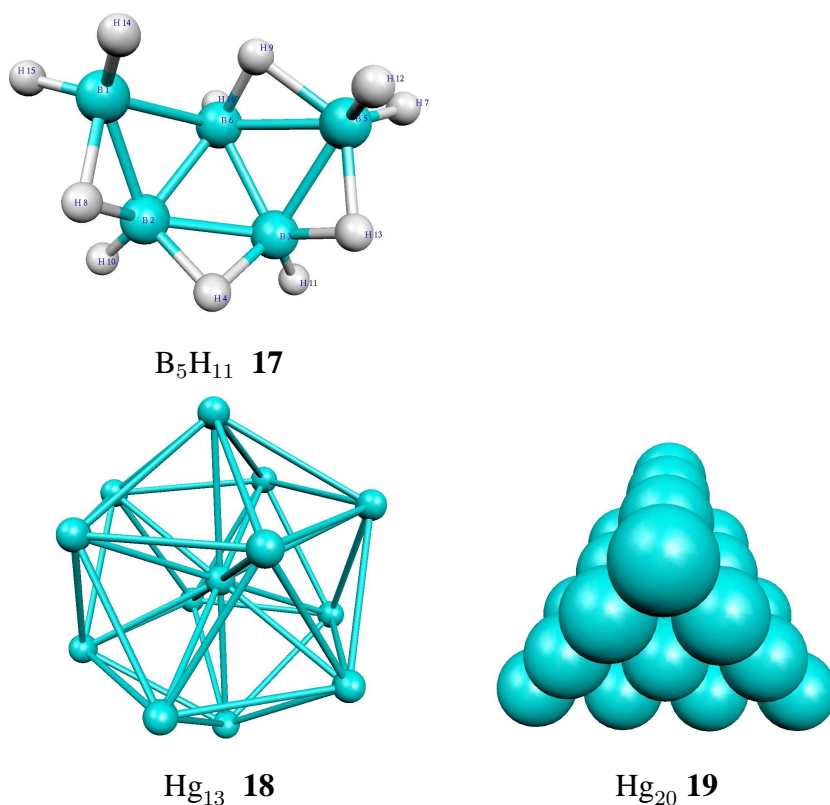


Figure 6.16: RI-BP86/SVP optimized structures of B_5H_{11} **17** (C_1), Hg_{13} **18** (I_h), and Hg_{20} **19** (T_d).

6.4.1 Boron Hydrides

For borate-anions like $B_nH_n^{2-}$ it was previously shown that the efficiency of local correlation methods is limited due to the delocalization of the electrons in the boron cage [53]. Therefore we choose the neutral borane cluster B_5H_{11} (**17** in figure 6.16) in order to check the performance of the proposed approach for this kind of systems. Tab. 6.20 shows the convergence of the incremental CCSD calculations for B_5H_{11} for two different partitionings into domains (i.e. 4 and 6 domains) with respect to the order

of the expansion. The convergence is a bit slower for smaller domains (dsp=2) compared to larger ones (dsp=3). If we compare the accuracy for the different orders in the incremental expansion, we find again that the first-order is not sufficient whereas the second-order is already a good guess with an accuracy of $\sim 1-3\%$. For the third-order level we get almost the exact correlation energy with errors less than 0.1 % or 0.4 kcal/mol. Whereas the fourth-order result for dsp=3 merely demonstrates that numerical errors are below 10^{-6} Hartree and the result of the standard CCSD calculation is recovered, the dsp=2 case shows that the series can indeed be truncated at low order, e.g. at third-order compared to the highest possible order of 6.

domains	order i	i -th order correction [au]	$E_{\text{corr}}(i)$ [au]	error [kcal/mol]	% E_{corr}
dsp=2, (6 domains) ^a	1	-0.412949	-0.412949	123.04	67.81
	2	-0.210519	-0.623468	-9.06	102.37
	3	0.015055	-0.608413	0.38	99.90
	4	-0.000591	-0.609004	0.01	100.00
dsp=3, (4 domains) ^a	1	-0.459470	-0.459470	93.85	75.44
	2	-0.155728	-0.615198	-3.88	101.01
	3	0.006282	-0.608916	0.07	99.98
	4	-0.000107	-0.609022	0.00	100.00
exact CCSD			-0.609022		

Table 6.20: Comparison of the incremental correlation energies with the full CCSD results for the B_5H_{11} -cluster (**17** in figure 6.16).

^a6-31G** basis sets of Pople and coworkers [99, 100]

6.4.2 Mercury Clusters

As a second example we choose two different mercury clusters, the icosahedral mercury cluster Hg_{13} (**18** in figure 6.16) and the tetrahedral mercury cluster Hg_{20} (**19** in figure 6.16). The convergence of the incremental expansion is fast for both systems as well as for both domain sizes, cf. tab. 6.21. Again we find a slightly faster convergence of the incremental series with the larger dsp value. The fourth-order result for

dsp=3 and Hg₁₃ again corresponds to the value obtained from standard CCSD calculations and indicates that the errors due to the incremental expansion are below 10⁻⁶ Hartree. Since a larger dsp leads to more occupied orbitals in a single domain, the time for the single CCSD calculations will increase as well as the corresponding memory and disk-space requirements. However, the number of calculations needed to reach a certain order and thus accuracy is reduced.

system, domains	order i	i -th order correction [au]	$E_{\text{corr}}(i)$ [au]	error [kcal/mol]	% E_{corr}
Hg₁₃ 18^a					
dsp=2, (5 domains) $t_{\text{con}}=8.0$	1	-0.330521	-0.330521	49.27	80.80
	2	-0.083257	-0.413777	-2.97	101.16
	3	0.004910	-0.408868	0.11	99.96
	4	-0.000174	-0.409042	0.00	100.00
dsp=3, (4 domains) $t_{\text{con}}=8.0$	1	-0.341241	-0.341241	42.55	83.42
	2	-0.070848	-0.412088	-1.91	100.74
	3	0.003080	-0.409009	0.02	99.99
	4	-0.000034	-0.409043	0.00	100.00
exact CCSD			-0.409043		
Hg₂₀ 19^a					
dsp=2, (10 domains) $t_{\text{con}}=6.9$	1	-0.497706	-0.497706	88.00	78.02
	2	-0.149857	-0.647564	-6.03	101.51
	3	0.010127	-0.637437	0.32	99.92
	4	-0.000556	-0.637992	-0.03	100.01
dsp=3, (6 domains) $t_{\text{con}}=8.0$	1	-0.516375	-0.516375	76.29	80.94
	2	-0.127074	-0.643449	-3.45	100.86
	3	0.005649	-0.637800	0.09	99.98
	4	-0.000144	-0.637944	0.00	100.00
exact CCSD			-0.637947		

Table 6.21: Comparison of the incremental correlation energies with the full CCSD results for the Hg clusters in figure 6.16.

^a relativistic large-core pseudopotential ECP78MWB and corresponding polarized double- ζ basis set of Küchle et al. [133].

At second-order we find an error less than 1.6 % or 6.1 kcal/mol, whereas we obtain excellent agreement at third-order with errors below 0.08 % or 0.4 kcal/mol. Note that the second-order result would determine the cohesive energy per atom of Hg₁₃

with an accuracy of 0.2 kcal/mol compared to estimated cohesive energies per atom of 4.0 and 6.6 kcal/mol from CCSD and pure diffusion quantum Monte Carlo (PDQMC) studies [134]. The cheap second-order level would thus allow to determine CCSD cohesive energies per atom for a given basis set with an accuracy of 5 %. This accuracy is already attractive, since larger basis sets than those feasible in standard CCSD calculations could be applied and it was demonstrated for Hg_4 that the basis set yielding a cohesive energy per atom of 4.0 kcal/mol for Hg_{13} recovers at most 62 % of the cohesive energy per atom of Hg_4 . Thus an approximate CCSD cohesive energy per atom with an estimated error of 5 % obtained with a large basis set is most likely closer to the complete basis-set limit than an exact result derived with a small basis set.

6.5 Potential Energy Surfaces

6.5.1 Octane

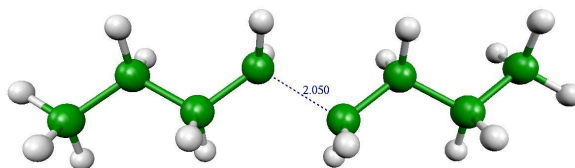


Figure 6.17: RI-BP86/SVP optimized structure of n-octane with a fixed C4-C5 distance of 2.05 Å.

We choose n-octane (figure 6.17) as a test system in order to obtain a significant result with respect to higher order terms. A too small system would yield the exact result already at low order. The geometries used for the potential curve were obtained by a relaxed scan of the C4-C5 distance using the RI-BP86/SVP method. With the chosen parameters we have 162 CCSD calculations for the full incremental calculation ($f = \infty$) at every point of the potential curve. The 6-31G** basis set of Pople and coworkers [99, 100] was used for the CCSD calculations.

First we note that we cannot recognize a difference between the incremental potential energy surface (PES) and the exact CCSD PES at the scale of figure 6.18. Comparing the approximate CCSD energy with the exact CCSD energy in figure 6.19 we get very accurate energies, if f is set to 14 or higher and the expansion is truncated at third-order. Furthermore we can see in figure 6.19 that the approximations with respect to the order and with respect to R_{\min} behave very systematically. One can obtain relatively smooth potential energy surfaces, if f is set to 16.

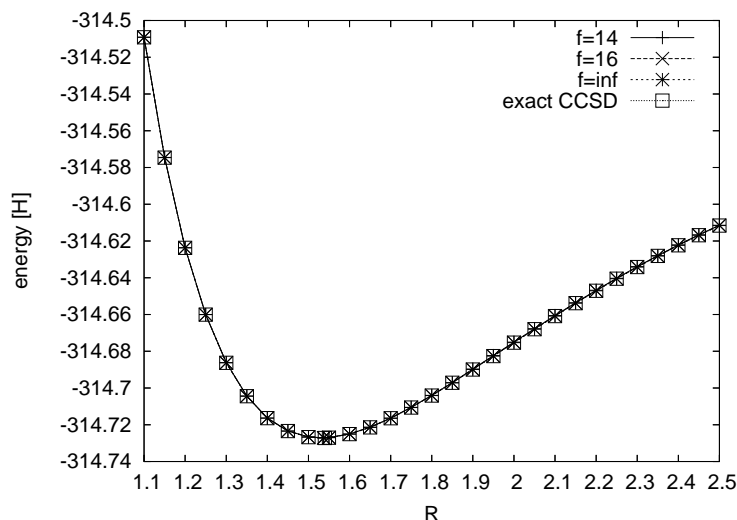


Figure 6.18: Potential energy scan along the C4-C5 distance [\AA] of n-octane (figure 6.17) for the exact CCSD energies, the full ($f = \infty$) and the distance truncated ($f = 14, 16$) fourth-order incremental expansions.

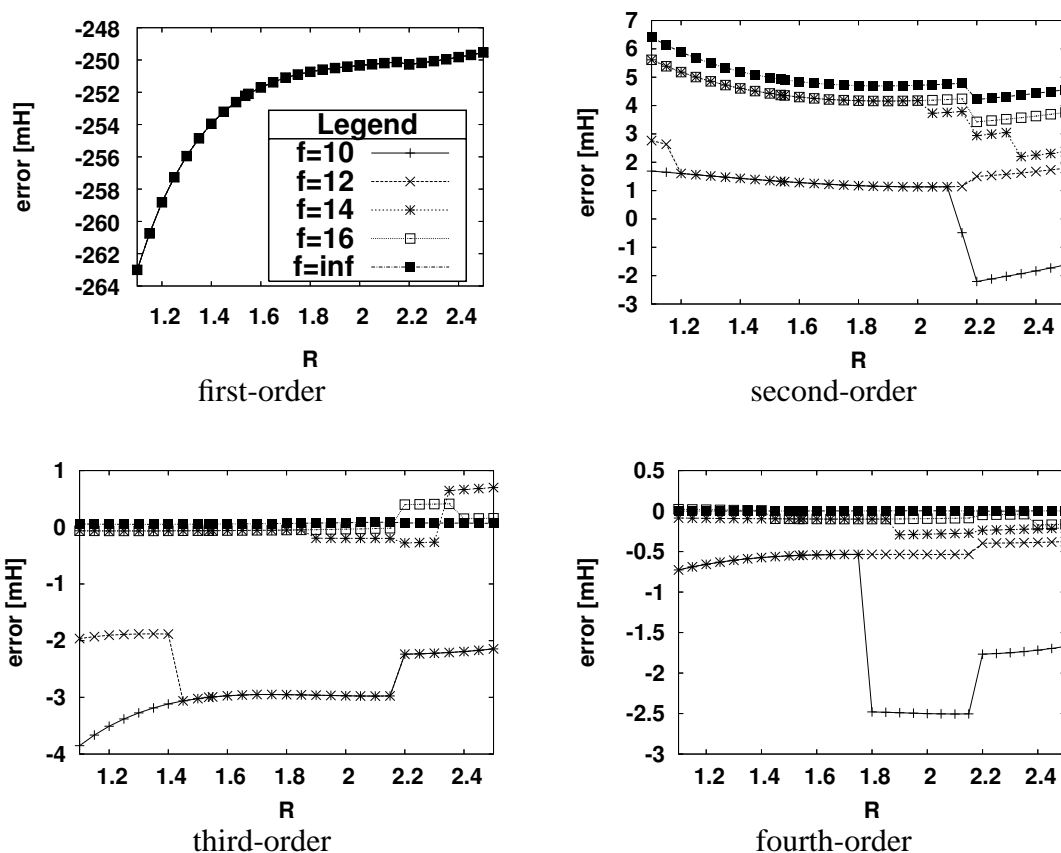


Figure 6.19: Errors of the incremental expansion of the CCSD correlation energy of n-octane (figure 6.17) for a variation of the C4-C5 distance R [\AA] with respect to the expansion order and the truncation parameter f . The legends for all four graphs are given in the plot for the first-order.

The PES is already smooth at second-order level over a quite long distance interval. The non-linearity error, however, is still quite large in this case. At third-order level the results are already very accurate. The largest errors are 0.09 mH for $f = \infty$, 0.41 mH for $f=16$ and 0.70 mH for $f=14$. Comparing these results with those of Mata and Werner [135] with a largest error of 2.8-4.2 mH, we find that with the given parameters the proposed approach is more accurate.

Figure 6.19 shows that the accuracy of the incremental expansion critically depends on the distance parameter f . If f is set to 14 or higher, we get a smooth and accurate PES. Further we find that the PES depends on the connectivity parameter t_{con} . It leads to poor domains, if the bonding orbital of the C4-C5 bond is not connected with the rest of the molecule anymore ($R > 2.5 \text{ \AA}$). We can fix this problem by increasing the connectivity threshold t_{con} to a larger value (4.0).

6.5.2 Hexayne

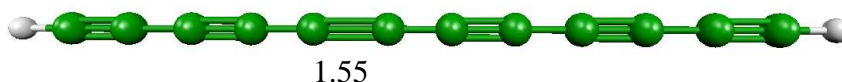


Figure 6.20: RI-BP86/SVP optimized structure of dodeca-hexayne with a fixed C5-C6 distance of 1.55 Å.

The geometries were obtained by a relaxed scan of the C5-C6 distance using the RI-BP86/SVP method. We used the 6-31G** basis set of Pople and coworkers [99, 100] for the CCSD calculations. Since the breaking of a single C-C bond is not very difficult, we decided to check the performance of the incremental scheme in breaking a C-C triple bond (figure 6.20). Although the CCSD method is not suitable for such a case, one can check for the convergence of the incremental expansion. In this case we find a small deviation of the exact PES and the approximate PES already at the scale of figure 6.21. From figure 6.22 we find again that the first-order expansion is not sufficient. The second-order is not smoothly varying and has a large error in contrast to the breaking of the C-C single bond. The expansion up to third-order yields a fairly smooth potential for all chosen distance parameters f , whereas the non-linearity error is still large for $f \leq 16$. At fourth-order we get a fairly smooth potential for $f \geq 12$.

The non-linearity error is about 1.2 mH for $12 \leq f \leq 14$. If we choose f to be 16 we find an error below 0.5 mH. The maximum errors are 0.35 mH for $f=\infty$, 0.38 mH for $f=16$ and 1.86 mH for $f=14$. If we compare our results with those of Mata and Werner [135] and Subotnik and Head-Gordon [27] we find a better total accuracy if we truncate at fourth-order and choose $f=16$. However, the efficiency with respect to the computational time of our current test implementation is certainly worse than the efficiency of the approaches of Werner et al. [135] or Head-Gordon et al. [27].

A general strategy to obtain suitable values for f is to do a series of test calculations with different f in a small basis set and use the best value for f in the calculation with the larger basis set. This can be done because the convergence behavior of the incremental series is not very sensitive with respect to the one particle basis set [86]. An alternative to this is to transfer the factor f from a similar class of compounds (see section 6.3).

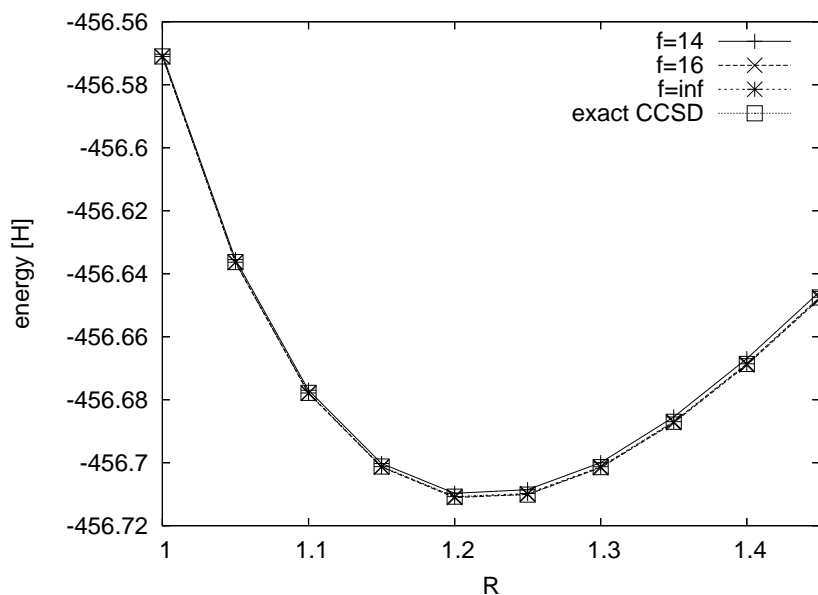


Figure 6.21: Potential energy scan along the C5-C6 distance [\AA] of dodeca-hexayne (figure 6.20) for the exact CCSD energies, the full ($f=\infty$) and the distance truncated ($f=14,16$) fourth-order incremental expansions.

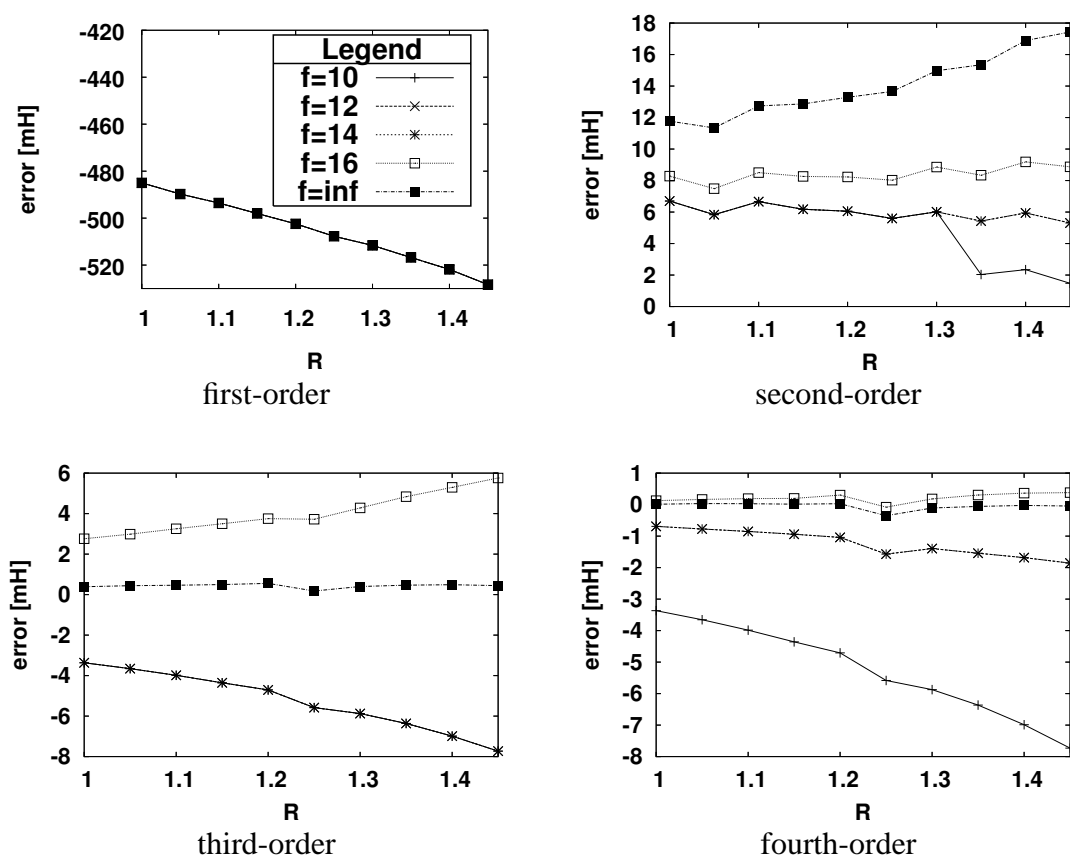


Figure 6.22: Errors of the incremental expansion of the CCSD correlation energy of hexayne (figure 6.20) for a variation of the C5-C6 distance R [Å] with respect to the expansion order and the truncation parameter f . The legends for all four graphs are given in the plot for the first-order.

6.6 Polymers

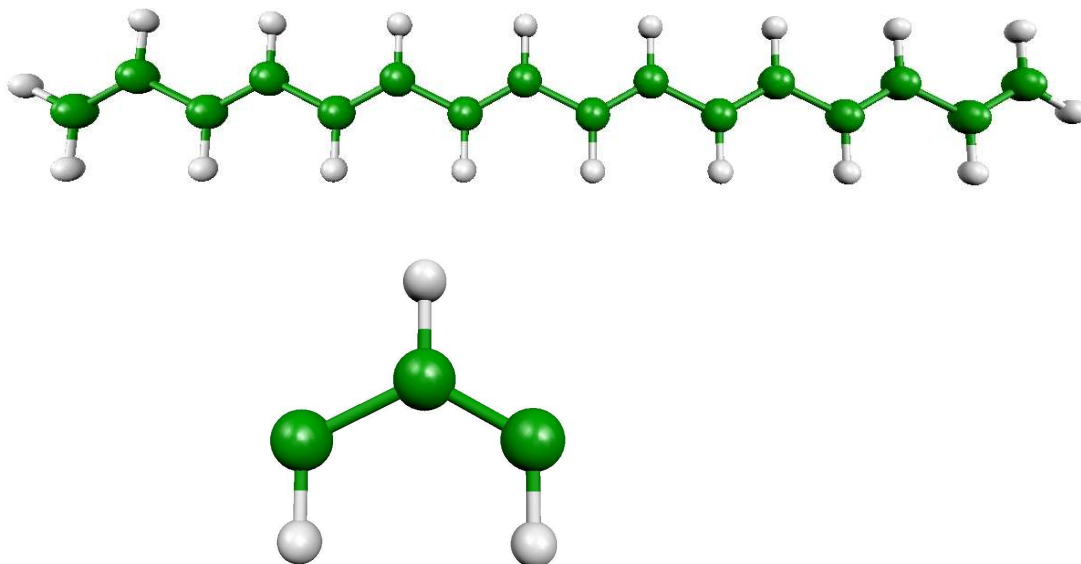


Figure 6.23: Structure of the used hydrocarbon chain and the symmetric cell of $(\text{C}_2\text{H}_2)_n$. The geometry parameters were extracted from the PhD-thesis of Kalvoda [136].

Table 6.22 shows the convergence of the CCSD/6-31G** incremental correlation energies for *all-trans* polyacetylene. The infinite summations for second and third-order increments were truncated at next neighbour cells. In this case we did not use the frozen-core approximation. A comparison to the data of Kalvoda [136] is not immediately possible, since we used a different basis set and Kalvoda used the frozen-core approximation. We note that the orders of magnitude for the contribution of a given order in the incremental expansion are equal. The CCSD energies were iterated to an accuracy of 10^{-6} Hartree.

Order	\mathcal{O}_i	\mathcal{O}_i -Correction	$E_{\text{corr}}[\text{au}]$
1		-0.166227	-0.166227
2		-0.123505	-0.289733
3		0.006539	-0.283193

Table 6.22: 6-31G** incremental CCSD correlation energies per cell for *all-trans* polyacetylene.

The benchmark results of table 6.23 were obtained by the finite-cluster approach with the unit cell and the molecule of figure 6.23 ($n=7$). According to the third column in table 6.23 we see that the benchmark energy is already converged to about 10^{-4} - 10^{-5} Hartree which is quite accurate. A comparison with the third-order incremental expansion yields an error of about 1.4 kcal/mol (100.8 %). This is somewhat less accurate than we expected from the convergence of the molecular calculations. There are three possible sources for this error: firstly the accuracy of the CCSD calculations within the domains was not strict enough, secondly the approximate symmetric orbitals introduce an error and thirdly the restriction of the infinite sums was too strict, since only increments within the next neighbour cells were taken into account.

Size of the Supercell	$E_{\text{corr}}(\mathcal{U}_n)$	$E_{\text{corr}}^{\text{cell}}(\mathcal{U}_n - \mathcal{U}_{n-1})$
n	[au]	[au]
5	-1.714325	
6	-1.995280	-0.280954
7	-2.276236	-0.280957

Table 6.23: 6-31G** finite-cluster CCSD correlation energies per cell for *all-trans* polyacetylene.

6.7 Excited States

Within the framework of the incremental scheme it is possible to calculate excited states using a local ansatz. If the CAS-wavefunction represents a reasonable approximation of the lowest states, we can account for dynamical correlation using the incremental scheme. Necessary conditions for the application of the incremental scheme are size-extensivity and size-consistency of the correlation method. Size-extensivity is important, because the size of the system is successively increased with increasing order of the incremental expansion. This can be demonstrated by a numerical study of the test system in figure 6.24. Since the domains were chosen to be H_2 -pairs, we have 6 non-interacting parts of our system. The electrons of the H_4 square in the middle are set to be active and are included in every calculation. We can expect that a first-order expansion yields the correct incremental energy for the system from a physical point of view. Due to the lack of size-consistency of MR-CISD, MR-ACPF and MR-AQCC we find rather large corrections even at fourth-order level as we can see in table 6.24. The sums of the fourth-order corrections for these methods are of comparable magnitude. Note that the MR-AQCC seems to perform quite well up to third-order, whereas it collapses at fourth-order level. If we compare the results of MR-CISD, MR-ACPF and MR-AQCC with the results obtained by MR-CEPA(0), we see that MR-CEPA(0)

is superior to the other methods from the perspective of convergence. MR-CEPA(0) yields the correct correlation energy at first-order level as expected, because it is size-consistent and size-extensive. Furthermore we can see from the MR-CEPA(0) results that we can obtain reasonable correlation energies for local excited states within the framework of the incremental scheme.

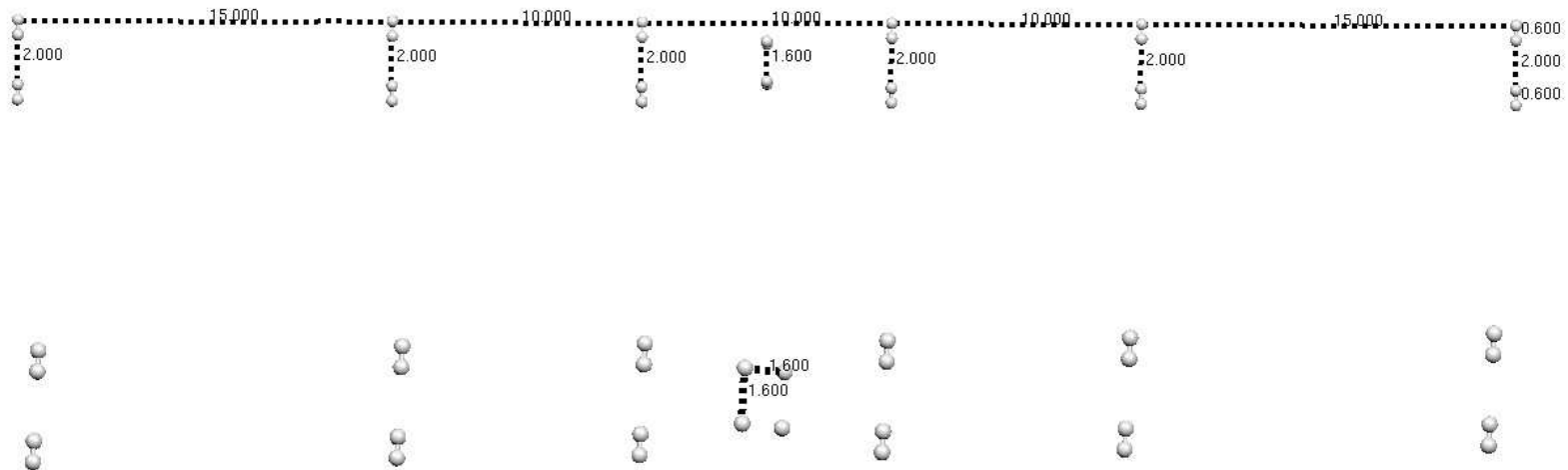


Figure 6.24: H₂₈ test system (distances in Å).

Method	MR-CISD		MR-ACPF		MR-AQCC		MR-CEPA(0)	
	Multiplicity=1							
Order \mathcal{O}_i	\mathcal{O}_i -Correction	$E_{\text{corr}}[\text{au}]$	\mathcal{O}_i -Correction	$E_{\text{corr}}[\text{au}]$	\mathcal{O}_i -Correction	$E_{\text{corr}}[\text{au}]$	\mathcal{O}_i -Correction	$E_{\text{corr}}[\text{au}]$
1	-0.391946	-0.391946	-0.399291	-0.399291	-0.396420	-0.396420	-0.401879	-0.401879
2	0.044513	-0.347433	0.004203	-0.395088	0.039031	-0.357389	0.000003	-0.401876
3	-0.008475	-0.355908	-0.004279	-0.399367	0.000316	-0.357072	-0.000002	-0.401879
4	0.001389	-0.354519	0.002581	-0.396786	-0.004384	-0.361457	0.000000	-0.401878
$E_{\text{corr}}^{\text{active}}$	-0.007517		-0.007527		-0.007534		-0.007538	
	Multiplicity=3							
Order \mathcal{O}_i	\mathcal{O}_i -Correction	$E_{\text{corr}}[\text{au}]$	\mathcal{O}_i -Correction	$E_{\text{corr}}[\text{au}]$	\mathcal{O}_i -Correction	$E_{\text{corr}}[\text{au}]$	\mathcal{O}_i -Correction	$E_{\text{corr}}[\text{au}]$
1	-0.391582	-0.391582	-0.399199	-0.399199	-0.396208	-0.396208	-0.401880	-0.401880
2	0.045728	-0.345853	0.004307	-0.394892	0.040210	-0.355998	0.000019	-0.401861
3	-0.009449	-0.355302	-0.004469	-0.399362	-0.000372	-0.356370	-0.000011	-0.401873
4	0.001587	-0.353715	0.002692	-0.396670	-0.004424	-0.360794	0.000001	-0.401872
$E_{\text{corr}}^{\text{active}}$	-0.013941		-0.013986		-0.014016		-0.014031	

Table 6.24: Convergence behavior of incremental correlation energies of the H_{28} aggregate for MR-CISD, MR-ACPF, MR-AQCC and MR-CEPA(0) correlation energies.

Chapter 7

Symmetric Localization

In the framework of the incremental scheme it is beneficial to use symmetric local orbitals, since this leads to large savings of computer time. The standard Foster-Boys procedure [3, 79] does not keep the symmetry of the whole system in all cases. In special cases it leads to good numerical results, despite the symmetry breaking due to the 2×2 rotations. But in general it may happen that the localization criterion contradicts with the symmetry of the system.

Example: CrCl_6

The total symmetry of the system is O_h but the core orbitals of the chromium will have a lower symmetry, since the Foster-Boys procedure yields local orbitals with a maximum distance of their centers of charge. This is for a sp-core shell a tetrahedron, which is not in agreement with the total symmetry of the octahedron.

We try to solve this problem by a unitary transformation of the orbitals which yields the correct symmetry in the centers of charge. This is done by minimizing the distance of the centers of charge to a set of template vectors $\tilde{\mathbf{r}}_i$.

$$D(\phi) = \min \left[\sum_i \left(\vec{R}_i - \tilde{\mathbf{r}}_i \right)^2 \right] = \min \left[\sum_i \sum_{j=1}^3 \left(\langle \phi_i | r_j | \phi_i \rangle - \tilde{r}_{ij} \right)^2 \right] \quad (7.1)$$

The template vectors are obtained as symmetrized centers of charge from a Foster-

Boys guess. If they are core orbitals we set them to zero¹, otherwise we perform a symmetrization. Let \mathbb{O} be a set of centers of charge of the occupied Foster-Boys orbitals, $\bar{\mathbb{O}}$ the set of centers of charge of occupied Foster-Boys orbitals without the centers of charge of core orbitals and \mathbb{M} the set of all symmetry operators of the point group. We define an equivalence class \mathbb{E} according to eqn. 7.2:

$$\mathbb{E}_k := \{\vec{R}_i \in \bar{\mathbb{O}} : \hat{m}_{ik} \vec{R}_k \approx \vec{R}_i\} \quad (7.2)$$

We can obtain all equivalence classes by:

1. start with an arbitrary orbital o_x from $\bar{\mathbb{O}}$ and build the equivalence class \mathbb{E}_x according to eqn. 7.2
2. build the complement of $\bar{\mathbb{O}}$ and \mathbb{E}_x : $\tilde{\mathbb{O}} = \bar{\mathbb{O}} \setminus \mathbb{E}_x$
3. end if $\tilde{\mathbb{O}} = \emptyset$ else goto 1 and use $\tilde{\mathbb{O}}$ as $\bar{\mathbb{O}}$

In the next step we apply all symmetry operators to every point of the equivalence class \mathbb{E}_x . Now we add all vectors within an arbitrary non-redundant part and divide by the number of vectors in this part. After applying all symmetry operators to the symmetrized vector we obtain the set of template vectors $\tilde{\mathbf{r}}_{x_i}$ for the equivalence class \mathbb{E}_x .

Now we apply the orthogonal 2×2 rotation of eqn. 7.3 to the occupied orbitals in eqn. 7.1.

$$\begin{aligned} u_i &= \cos(\gamma)\phi_i - \sin(\gamma)\phi_j \\ u_j &= \cos(\gamma)\phi_j + \sin(\gamma)\phi_i \end{aligned} \quad (7.3)$$

This leads to:

$$\begin{aligned} D(u) &= \sum_{\substack{i \\ i \neq a, b}} \left(\vec{R}_i - \tilde{\mathbf{r}}_i \right)^2 + \sum_{j=1}^3 \left[\langle (\cos(\gamma)\phi_a - \sin(\gamma)\phi_b) | r_j | (\cos(\gamma)\phi_a - \sin(\gamma)\phi_b) \rangle^2 \right. \\ &\quad - 2 \langle (\cos(\gamma)\phi_a - \sin(\gamma)\phi_b) | r_j | (\cos(\gamma)\phi_a - \sin(\gamma)\phi_b) \rangle \tilde{r}_{a_j} + \tilde{r}_{a_j}^2 \\ &\quad \langle (\cos(\gamma)\phi_b + \sin(\gamma)\phi_a) | r_j | (\cos(\gamma)\phi_b + \sin(\gamma)\phi_a) \rangle^2 \\ &\quad \left. - 2 \langle (\cos(\gamma)\phi_b + \sin(\gamma)\phi_a) | r_j | (\cos(\gamma)\phi_b + \sin(\gamma)\phi_a) \rangle \tilde{r}_{b_j} + \tilde{r}_{b_j}^2 \right] \end{aligned} \quad (7.4)$$

¹This is a special case. The generalization would be an analogous treatment of all points with higher symmetry to the treatment of the central point.

Expanding the terms and using the addition theorem of cosine and sine we get:

$$\begin{aligned}
 F(\gamma)_{ab} &= D(u)_{ab} - D(\phi)_{ab} = A + B \cos(2\gamma) + C \sin(2\gamma) + D \cos(4\gamma) + E \sin(4\gamma) \\
 A &= \frac{3}{4} \langle \phi_a | r_j | \phi_a \rangle^2 + \langle \phi_a | r_j | \phi_b \rangle^2 + \frac{1}{2} \langle \phi_a | r_j | \phi_a \rangle \langle \phi_b | r_j | \phi_b \rangle \\
 &\quad + \frac{3}{4} \langle \phi_b | r_j | \phi_b \rangle^2 - \langle \phi_a | r_j | \phi_a \rangle \tilde{r}_{a_j} - \langle \phi_b | r_j | \phi_b \rangle \tilde{r}_{a_j} + \tilde{r}_{a_j}^2 \\
 &\quad - \langle \phi_a | r_j | \phi_a \rangle \tilde{r}_{b_j} - \langle \phi_b | r_j | \phi_b \rangle \tilde{r}_{b_j} + \tilde{r}_{b_j}^2 \\
 B &= -\langle \phi_a | r_j | \phi_a \rangle \tilde{r}_{a_j} \\
 &\quad + \langle \phi_b | r_j | \phi_b \rangle \tilde{r}_{a_j} + \langle \phi_a | r_j | \phi_a \rangle \tilde{r}_{b_j} - \langle \phi_b | r_j | \phi_b \rangle \tilde{r}_{b_j} \\
 C &= 2 \langle \phi_a | r_j | \phi_b \rangle \tilde{r}_{a_j} - 2 \langle \phi_a | r_j | \phi_b \rangle \tilde{r}_{b_j} \\
 D &= \frac{1}{4} \langle \phi_a | r_j | \phi_a \rangle^2 - \langle \phi_a | r_j | \phi_b \rangle^2 \\
 &\quad - \frac{1}{2} \langle \phi_a | r_j | \phi_a \rangle \langle \phi_b | r_j | \phi_b \rangle + \frac{1}{4} \langle \phi_b | r_j | \phi_b \rangle^2 \\
 E &= -\langle \phi_a | r_j | \phi_a \rangle \langle \phi_a | r_j | \phi_b \rangle + \langle \phi_a | r_j | \phi_b \rangle \langle \phi_b | r_j | \phi_b \rangle
 \end{aligned} \tag{7.5}$$

The functional in eqn. 7.5 is minimal if the derivative is zero. Since the function has several maxima and minima in the interval $[0, 2\pi]$ we choose the angle γ where $F(\gamma)_{ab}$ is lowest and $\frac{dF(\gamma)_{ab}}{d\gamma} = 0$. We choose the transformation u according to the largest value of $F(\gamma)_{ab}$. After the transformation of the orbitals we build the matrix $F(\gamma)$ again and iterate until all values in $F(\gamma)$ are below some threshold (10^{-12}).

The procedure works quite well to obtain symmetry adapted centers of charge, but it fails in producing symmetric orbitals, i.e. further work along these lines is required.

Chapter 8

Summary and Outlook

8.1 Summary

In this thesis the first fully automatized implementation of the incremental scheme was presented. It was applied to molecules, open-shell systems, the 4-*exo*/5-*endo* cyclization, potential energy surfaces and a polymer. Furthermore the error propagation was analyzed theoretically and by a numerical study with uniformly distributed random numbers. A cheap and systematic screening procedure to select only the most important increments was introduced. The convergence of the proposed incremental method is fast for almost all systems considered in this work. The convergence behavior did not change significantly by increasing the basis set from double- ζ to triple- ζ basis sets as well as for the usage of a diffuse basis set (6-31++G** and aug-cc-pVDZ). The proposed method was demonstrated to be quite robust with respect to the choice of the domains. It was shown that the errors can be below one kcal/mol, if the series is truncated at a proper order (usually \mathcal{O} can be kept below 4). Furthermore it was shown that this small error was still obtained if the important increments were selected by the distance of two groups of one-site domains.

The approximate treatment of symmetry in combination with Foster-Boys orbitals was explored for two intermolecular systems as well as a polymer. For the symmetric water cluster and the symmetric cyclotriazine cluster we obtained fast convergence for the incremental expansion. We got almost the exact result at second- and third-order, respectively. The convergence of the incremental series is faster for the water cluster, because the one-site domains were fixed to a whole water molecule, whereas the cyclotriazine molecules were divided into two one-site domains (according to the chosen dsp values). The incremental calculation on polyacetylene yielded an error of 1.4 kcal/mol compared to the finite-cluster result.

The performance of CI-based correlation methods as MR-CISD, MR-ACPF, MR-

AQCC and MR-CEPA(0) was tested for a non-interacting system. We found that the size-extensivity/size-consistency errors prohibit the convergence of the incremental series for MR-CISD, MR-ACPF and MR-AQCC. According to this error the increments do not decay with respect to the distance of the localized orbitals and it is therefore not possible to truncate the incremental series with respect to the distance of the domains. This is a necessary condition for the treatment of periodic systems, since the infinite summations have to be truncated. The MR-CEPA(0) results however yield the expected result for a non-interacting system: the sum of the first-order correlation energies corresponds to the exact result.

The incremental scheme performed quite well in the calculation of potential energy surfaces. The error due to the incremental expansion was kept below one kcal/mol and we obtained smooth potential energy surfaces, if the expansion was truncated at third or fourth-order. Furthermore we still obtained smooth potential energy surfaces if we selected the important increments by an order-dependent distance threshold ($t_{\text{dist}} = \frac{f}{\sigma_i}$).

For molecules the incremental series converged to an accuracy better than one kcal/mol for all molecules in this work. We found that the convergence was slower for sphere-shaped molecules like TiCp_2Cl_2 compared to σ -bonded chains like n-decane. A screening procedure for negligibly small increments was implemented and applied to a set of test molecules. The errors introduced by the distance screening were shown to be below one kcal/mol, if the truncation parameter f is chosen properly. The energy dependence of this parameter was checked for a set of real life molecules as large as the cytosine-guanine base pair.

8.2 Outlook

In a future work it is possible to test the incremental scheme with respect to molecular properties. Since the server and client structure is already adapted for hyperpolarizabilities within the DALTON framework, it is straightforward to go in this direction. Another interesting question is to adapt a gradient to the proposed incremental method. A strategy could be the expansion of the analytical gradient in an incremental series. This has the major advantage that the gradient is well defined, even if the incremental PES is not smooth. A further advantage is that the precision of the energy is not necessarily required to be better than 10^{-6} Hartree. For test purposes one could extract the analytical CCSD gradient from DALTON [88] quantum chemistry package.

Since the CCSD(T) method is one of today's most reliable quantum chemical methods, it would be useful to include this method into the incremental program. The simplest

way to adapt the CCSD(T) method into the framework of the incremental scheme is to diagonalize the Fock matrix within the space of the fragments and perform a standard CCSD(T) calculation afterwards. This modification is necessary to converge to the exact CCSD(T) result, since the perturbative triples correction is not invariant with respect to the unitary transformations of the MO basis. If the Fock matrix is partially diagonalized for every n -site domain, the differences with respect to the choice of the one particle basis set are integrated into the incremental series. A different way to combine the CCSD(T) approach with the incremental scheme is to solve the triples correction in an iterative fashion as discussed by Schütz [21].

Another point is the implementation of a linear scaling incremental method, which can be based on the results of the current work. The first step to do this is the implementation of a fast linear scaling AO-MO-transformation with the possibility to freeze and delete orbitals. Next one has to write an efficient correlation code to get the incremental energies in a reasonable time.

In combination with an efficient implementation of the correlation modules it is possible to perform automatic incremental calculations for periodic systems. In order to get symmetry adapted MOs one could extract the required data from WANNIER [44, 137] or CRYSTAL [138–140] quantum chemistry packages.

One could explore the usage of a configuration selection based on perturbation theory for incremental CEPA(0) energies using the DIESEL quantum chemistry package [77]. This should be straightforward, since the current implementation can handle the corresponding threshold already. Furthermore it would be desirable to extend the treatment of multi-reference cases to MRCC approaches (e.g. MR-expT of Hanrath [141]). For this purpose one can use the experience of the MR-CEPA(0) calculations from this work.

Finally it is an interesting point to adapt the symmetry of the system to the localization procedure in order to use the full point group symmetry for the incremental expansion, in cases where the localization criterion does not agree with the symmetry of the system.

Bibliography

- [1] R. K. Nesbet, *Adv. Chem. Phys.* **14**, 1 (1969).
- [2] L. Bytautas, J. Ivanic, K. Ruedenberg, *J. Chem. Phys.* **119**, 8217 (2003).
- [3] J. M. Foster, S. F. Boys, *Rev. Mod. Phys.* **32**, 300 (1960).
- [4] J. Pipek, P. G. Mezey, *J. Chem. Phys.* **90**, 4916 (1989).
- [5] P. Pulay, S. Saebø, *Theor. Chem. Acc.* **69**, 357 (1986).
- [6] S. Saebø, P. Pulay, *J. Chem. Phys.* **86** (2), 914 (1987).
- [7] S. Saebø, P. Pulay, *Annu. Rev. Phys. Chem.* **44**, 213 (1993).
- [8] J. W. Boughton, P. Pulay, *J. Comput. Chem.* **14**, 736 (1993).
- [9] G. Hetzer, M. Schütz, H. Stoll, H.-J. Werner, *J. Chem. Phys.* **113**, 9443 (2000).
- [10] M. Schütz, G. Hetzer, H. J. Werner, *J. Chem. Phys.* **111** (13), 5691 (1999).
- [11] P. E. Maslen, M. Head-Gordon, *J. Chem. Phys.* **109**, 7093 (1998).
- [12] P. E. Maslen, M. Head-Gordon, *Chem. Phys. Lett.* **283**, 102 (1998).
- [13] M. S. Lee, P. E. Maslen, M. Head-Gordon, *J. Chem. Phys.* **112**, 3592 (2000).
- [14] P. E. Maslen, M. S. Lee, M. Head-Gordon, *Chem. Phys. Lett.* **319**, 205 (2000).
- [15] P. Y. Ayala, G. E. Scuseria, *J. Chem. Phys.* **110**, 3660 (1999).
- [16] D. G. Federov, K. Kitaura, *J. Chem. Phys.* **121**, 2483 (2004).
- [17] D. Walter, A. B. Szilva, K. Niedfeld, E. A. Carter, *J. Chem. Phys.* **117** (5), 1982 (2002).
- [18] D. Walter, A. Venkatnathan, E. A. Carter, *J. Chem. Phys.* **118** (18), 8127 (2003).

- [19] A. Venkatnathan, A. B. Szilva, D. Walter, *J. Chem. Phys.* **120** (4), 1693 (2004).
- [20] C. Hampel, H.-J. Werner, *J. Chem. Phys.* **104**, 6286 (1996).
- [21] M. Schütz, *J. Chem. Phys.* **113** (22), 9986 (2000).
- [22] M. Schütz, H.-J. Werner, *J. Chem. Phys.* **114** (2), 661 (2001).
- [23] M. Schütz, *Phys. Chem. Chem. Phys.* **4**, 3941 (2002).
- [24] M. Schütz, F. R. Manby, *Phys. Chem. Chem. Phys.* **5**, 3349 (2003).
- [25] N. Flocke, R. J. Bartlett, *J. Chem. Phys.* **121**, 10935 (2004).
- [26] J. E. Subotnik, M. Head-Gordon, *J. Chem. Phys.* **123**, 64108 (2005).
- [27] J. E. Subotnik, A. Sodt, M. Head-Gordon, *J. Chem. Phys.* **125**, 74116 (2006).
- [28] A. Auer, M. Nooijen, *J. Chem. Phys.* **125**, 24104 (2006).
- [29] O. Christiansen, P. Manninen, P. Jørgenson, J. Olsen, *J. Chem. Phys.* **124**, 084103 (2006).
- [30] G. Stollhoff, P. Fulde, *J. Chem. Phys.* **73**, 4548 (1980).
- [31] G. Stollhoff, *J. Chem. Phys.* **105**, 227 (1996).
- [32] C. Pisani, M. Busso, G. Capecchi, S. Casassa, R. Dovesi, L. Maschio, C. Zicovich-Wilson, M. Schütz, *J. Chem. Phys.* **122**, 94113 (2005).
- [33] C. Pisani, G. Capecchi, S. Casassa, L. Maschio, *Mol. Phys.* **103**, 2527 (2005).
- [34] D. Usvyat, M. Schütz, *Theor. Chem. Acc.* **114**, 276 (2005).
- [35] H. Stoll, *Chem. Phys. Lett.* **91**, 548 (1992).
- [36] H. Stoll, *J. Chem. Phys.* **97**, 8449 (1992).
- [37] H. Stoll, *Phys. Rev. B: Condens. Matter Mater. Phys.* **46**, 6700 (1992).
- [38] A. Abdurahman, A. Shukla, M. Dolg, *J. Chem. Phys.* **112**, 4801 (2000).
- [39] A. Abdurahman, A. Shukla, M. Dolg, *Chem. Phys. Lett.* **257**, 301 (2000).
- [40] A. Abdurahman, A. Shukla, M. Dolg, *Phys. Rev. B* (2002).
- [41] C. Buth, B. Paulus, *Phys. Rev. B* **74**, 45122 (2006).

- [42] K. Doll, M. Dolg, P. Fulde, H. Stoll, *Phys. Rev. B* **55**, 10282 (1997).
- [43] S. Kalvoda, M. Dolg, H.-J. Flad, P. Fulde, H. Stoll, *Phys. Rev. B* **57**, 2127 (1998).
- [44] A. Shukla, M. Dolg, P. Fulde, H. Stoll, *Phys. Rev. B* **60**, 5211 (1999).
- [45] M. Albrecht, B. Paulus, H. Stoll, *Phys. Rev. B: Condens. Matter Mater. Phys.* **56**, 7339 (1997).
- [46] B. Paulus, *Chem. Phys. Lett.* **371**, 7 (2003).
- [47] B. Paulus, K. Rosciszewski, H. Stoll, P. Fulde, *Phys. Chem. Chem. Phys.* **5**, 5523 (2003).
- [48] C. Willenauer, U. Birkenheuer, *J. Chem. Phys.* **120**, 11910 (2004).
- [49] A. B. Mukhopadhyay, M. Dolg, *J. Chem. Phys.* **120**, 8734 (2004).
- [50] E. Voloshina, B. Paulus, *Theor. Chem. Acc.* **114**, 259 (2005).
- [51] E. Voloshina, B. Paulus, *J. Chem. Phys.* **124**, 234711 (2006).
- [52] N. Gaston, B. Paulus, K. Rosciszewski, P. Schwerdtfeger, H. Stoll, *Phys. Rev. B* **74**, 94102 (2006).
- [53] S. Kalvoda, B. Paulus, M. Dolg, H. Stoll, H.-J. Werner, *Phys. Chem. Chem. Phys.* **3**, 514 (2001).
- [54] B. Paulus, K. Rosciszewski, P. Fulde, H. Stoll, *Phys. Rev. B* **68**, 235115 (2003).
- [55] V. Bezugly, *Wavefunction-based method for excited- state electron correlations in periodic systems*, phd thesis, Dresden (2004).
- [56] V. Bezugly, U. Birkenheuer, *Chem. Phys. Lett.* **399**, 57 (2004).
- [57] M. Mödl, M. Dolg, P. Fulde, H. Stoll, *J. Chem. Phys.* **106**, 1836 (1997).
- [58] P. Fulde, H. Stoll, *Found. Phys.* **30**, 2049 (2000).
- [59] P. Fulde, *Adv. Phys.* **51**, 909 (2002).
- [60] W. Li, S. Li, *J. Chem. Phys.* **121**, 6649 (2004).
- [61] S. Li, J. Ma, Y. Jiang, *J. Comput. Chem.* **23**, 237 (2002).
- [62] S. Li, J. Shen, W. Li, Y. Jiang, *J. Chem. Phys.* **125**, 074109 (2006).

- [63] S. Li, W. Li, T. Fang, *J. Am. Chem. Soc.* **127**, 7215 (2005).
- [64] X. He, J. Z. H. Zhang, *J. Chem. Phys.* **124**, 184703 (2006).
- [65] V. Deev, M. Collins, *J. Chem. Phys.* **122**, 154102 (2005).
- [66] W. Kutzelnigg, *Einführung in die Theoretische Chemie*, Wiley-VCH, Weinheim (2002).
- [67] A. Szabo, N. S. Ostlund, *Modern Quantum Chemistry*, Dover Publications, New York (1996).
- [68] T. Helgaker, P. Jørgensen, J. Olsen, *Molecular Electronic Structure Theory*, Wiley (2004).
- [69] K. A. Brueckner, *Phys. Rev.* **100**, 36 (1955).
- [70] A. Hugenholtz, *Physica* **23**, 481 (1957).
- [71] R. J. Bartlett, *Ann. Rev. Phys. Chem.* **32**, 359 (1981).
- [72] J. A. Pople, J. S. Binkley, R. Seeger, *Int. J. Quantum Chem. Symp.* **10**, 1 (1976).
- [73] W. Duch, G. H. F. Diercksen, *J. Chem. Phys.* **101**, 3018 (1994).
- [74] P. R. Taylor, *Lecture Notes on Quantum Chemistry* **64**, 125 (1994).
- [75] B. O. Roos, P.-O. Widmark (Eds.), *European Summerschool in Quantum Chemistry 2003 Book I*, University of Lund (2003).
- [76] B. O. Roos, P.-O. Widmark (Eds.), *European Summerschool in Quantum Chemistry 2003 Book II*, University of Lund (2003).
- [77] M. Hanrath, *Ein individuell selektierendes intern-extern-separiertes Multireferenz-Konfigurationswechselwirkungsverfahren*, phd thesis, Mathematisch Naturwissenschaftliche Fakultät der Rheinischen Friedrich-Wilhelms Universität Bonn (1999).
- [78] T. D. Crawford, H. F. Schaefer III, in: *Reviews in computational chemistry* (Eds. K. N. Lipkowitz, D. B. Boyd), Vol. 14, Chap. 2, S. 33–136, Wiley-VCH (2000).
- [79] C. Edmiston, K. Ruedenberg, *Rev. Mod. Phys.* **35**, 457 (1963).
- [80] C. Kittel, *Introduction to Solid States Physics*, John Wiley (1996).
- [81] R. Fondermann, M. Dolg, M. Raab, E. Nieke, *Chem. Phys.* **325**, 291 (2006).

- [82] *Glossary of graph theory at*
http://en.wikipedia.org/wiki/Glossary_of_graph_theory.
- [83] G. Karypis, V. Kumar, SIAM J. Sci. Comput. **20** (1), 359 (1998).
- [84] H. Stoll, B. Paulus, P. Fulde, J. Chem. Phys. **123**, 144108 (2005).
- [85] J. Friedrich, M. Hanrath, M. Dolg, J. Chem. Phys. **126**, 154110 (2007).
- [86] J. Friedrich, M. Hanrath, M. Dolg, J. Phys. Chem. A (2007), submitted.
- [87] C. Hampel, K. Peterson, H.-J. Werner, Chem. Phys. Lett. **190**, 1 (1992).
- [88] *DALTON, a molecular electronic structure program, Release 2.0 (2005), see*
<http://www.kjemi.uio.no/software/dalton/dalton.html>.
- [89] R. D. Amos, A. Bernhausen, A. Berning, P. Celani, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, C. Hampel, G. Hetzer, P. J. Knowles, T. Korona, R. Lindh, A. W. Lloyd, S. J. McNicholas, R. R. Manby, W. Meyer, M. E. Mura, A. Nicklass, P. Palmieri, R. Pitzer, G. Rauhut, M. Schütz, U. Schumann, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, H. J. Werner, *Molpro, a package of ab initio programs designed by H.-J. Werner and P. J. Knowles*, Tech. Rport, University of Birmingham (2001).
- [90] [Http://gridengine.sunsource.net/](http://gridengine.sunsource.net/).
- [91] G. Schaftenaar, J. Noordik, J. Comput.-Aided Mol. Design **14**, 123 (2000).
- [92] S. Portmann, H. P. Lüthi, Chimia **54**, 766 (2000).
- [93] P. Flükiger, H. Lüthi, S. Portmann, J. Weber, *MOLEKEL 4.2*, Tech. Rport, Swiss Center for Scientific Computing, Manno (Switzerland) (2000-2002).
- [94] T. Hahn (Ed.), *International Tables for Crystallography*, Vol. A, Kluwer Academic Publishers (2002).
- [95] *Bilbao crystallographic server* <http://www.cryst.ehu.es/index.html>.
- [96] A. D. Becke, Phys. Rev. A **38** (6), 3098 (1988).
- [97] J. P. Perdew, Phys. Rev. B **33** (12), 8822 (1986).
- [98] R. Ahlrichs, M. Bär, H.-P. Baron, R. Bauernschmitt, S. Böcker, M. Ehrig, K. Eichkorn, S. Elliott, F. Furche, F. Haase, M. Häser, H. Horn, C. Huber, U. Huniar, C. Kölmel, M. Kollwitz, C. Ochsenfeld, H. Öhm, A. Schäfer, U. Schneider,

- O. Treutler, M. von Arnim, F. Weigend, P. Weis, H. Weiss, *Turbomole 5*, Institut für Physikal. Chemie, Universität Karlsruhe, Kaiserstr. 12, D-76128 Karlsruhe (2002).
- [99] W. J. Hehre, R. Ditchfield, J. A. Pople, *J. Chem. Phys.* **56**, 2257 (1972).
- [100] P. C. Hariharan, J. A. Pople, *Theor. Chim. Acta* **28**, 213 (1973).
- [101] J. Wiebke, A. Moritz, X. Cao, M. Dolg, *Phys. Chem. Chem. Phys.* (2007).
- [102] T. H. Dunning Jr., *J. Chem. Phys.* **90**, 1007 (1989).
- [103] M. Dolg, U. Wedig, H. Stoll, H. Preuss, *J. Chem. Phys.* **86**, 866 (1987).
- [104] D. Andrae, U. Häussermann, M. Dolg, H. Stoll, H. Preuss, *Theor. Chem. Acc.* **77**, 123 (1990).
- [105] A. Moritz, M. Dolg, X. Cao, *Theor. Chem. Acc.* (2006).
- [106] S. Bulusu, S. Yoo, E. Apra, S. Xantheas, X. C. Zeng, *J. Phys. Chem. A* **110**, 11781 (2006).
- [107] S. Tsuzuki, K. Honda, T. Uchimaru, M. Mikami, K. Tanabe, *J. Am. Chem. Soc.* **124**, 104 (2002).
- [108] M. O. Sinnokrot, E. F. Valeev, C. D. Sherrill, *J. Am. Chem. Soc.* **124**, 10887 (2002).
- [109] M. O. Sinnokrot, C. D. Sherrill, *J. Phys. Chem. A* **108**, 10200 (2004).
- [110] O. Zhikol, O. V. Shishkin, K. A. Lyssenko, J. Leszczynski, *J. Chem. Phys.* **122**, 144104 (2005).
- [111] Y. C. Park, J. S. Lee, *J. Phys. Chem. A* **110**, 5091 (2006).
- [112] A. Puzder, M. Dion, D. C. Langreth, *J. Chem. Phys.* **124**, 164105 (2006).
- [113] J. G. Hill, J. A. Platts, H.-J. Werner, *Phys. Chem. Chem. Phys.* (2006).
- [114] S. Tsuzuki, K. Honda, T. Uchimaru, M. Mikami, K. Tanabe, *J. Am. Chem. Soc.* (2000).
- [115] K. Shibasaki, A. Fujii, N. Mikami, S. Tsuzuki, *J. Phys. Chem. A* **110**, 4397 (2006).

- [116] A. L. Ringer, M. S. Figgs, M. O. Sinnokrot, C. D. Sherrill, *J. Phys. Chem. A* **110**, 10822 (2006).
- [117] J. Gauss, J. F. Stanton, *J. Phys. Chem. A* **104**, 2865 (2000).
- [118] F. Mendizabal, P. Pyykkö, *Phys. Chem. Chem. Phys.* **6**, 900 (2004).
- [119] P. Schwerdtfeger, M. Dolg, W. Schwarz, G. Bowmaker, P. Boyd, *J. Chem. Phys.* **91**, 1762 (1989).
- [120] J. Friedrich, M. Dolg, A. Gansäuer, D. Geich-Gimbel, T. Lauterbach, *J. Am. Chem. Soc.* **127**, 7071 (2005).
- [121] K. Laasonen, M. Parrinello, R. Car, C. Lee, D. Vanderbilt, *Chem. Phys. Lett.* **207**, 208 (1993).
- [122] C. Kozmutza, E. S. Kryachko, E. Tfirst, *THEOCHEM* **501**, 435 (2000).
- [123] D. M. Upadhyay, M. K. Shukla, P. C. Mishra, *Int. J. Quantum Chem.* (2001).
- [124] J. Kim, D. Majumdar, H. M. Lee, K. S. Kim, *J. Chem. Phys.* **110**, 9128 (1999).
- [125] R. A. Kendall, T. H. Dunning Jr., R. J. Harrison, *J. Chem. Phys.* **96**, 6769 (1992).
- [126] M. Zhao, B. M. Gimarc, *J. Phys. Chem.* **98**, 7497 (1994).
- [127] S. Inagaki, Y. Ishitani, T. Kakefu, *J. Am. Chem. Soc.* **116**, 5954 (1994).
- [128] E. A. Salter, R. Z. Hinrichs, C. Salter, *J. Am. Chem. Soc.* **118**, 227 (1996).
- [129] D. H. Magers, E. A. Salter, R. J. Bartlett, C. S. and B. A. Hess, L. J. Schaad, *J. Am. Chem. Soc.* **110**, 3435 (1988).
- [130] H.-J. Song, H.-M. Xiao, H.-S. Dong, *J. Chem. Phys.* **125**, 074308 (2006).
- [131] G. Chalasinski, M. M. Szczesniak, S. M. Cybulski, *J. Chem. Phys.* **92**, 2481 (1990).
- [132] G. Chalasinski, M. M. Szczesniak, *Chem. Rev. (Washington, D. C.)* (1994).
- [133] W. Küchle, M. Dolg, H. Stoll, H. Preuss, *Mol. Phys.* **74**, 1245 (1991).
- [134] M. Dolg, H.-J. Flad, *Mol. Phys.* **91**, 815 (1997).
- [135] R. A. Mata, H.-J. Werner, *J. Chem. Phys.* **125**, 184110 (2006).

-
- [136] S. Kalvoda, *Lokale elektronische Korrelationen in Gadoliniumnitrid und anderen periodischen Systemen*, phd thesis, Max-Planck-Institut für Physik Komplexer Systeme, Dresden (1998).
- [137] A. Shukla, M. Dolg, P. Fulde, H. Stoll, *Phys. Rev. B* **57**, 1471 (1998).
- [138] C. Pisani, R. Dovesi, *Int. J. Quantum Chem.* **17**, 501 (1980).
- [139] V. R. Saunders, *Faraday Symp. Chem. Soc.* **19**, 79 (1984).
- [140] C. Pisani, R. Dovesi, C. Roetti, *Hartree-Fock ab-initio of crystalline systems*, Vol. 48 from *Lecture Notes in Chemistry*, Springer Verlag, Heidelberg (1988).
- [141] M. Hanrath, *J. Chem. Phys.* **123**, 084192 (2005).
- [142] *Gaussian basis set order form* <http://www.emsl.pnl.gov/forms/basisform.html>.

Appendix A

List of Abbreviations

ACPF	averaged coupled pair functional
AO	atomic orbital
AQCC	averaged quadratic coupled cluster
a.u.	atomic units
aug-cc-pVDZ	augmented correlation consistent polarized valence double- ζ basis set of Dunning
BP86	Becke-Perdew gradient corrected exchange and correlation density functional
CASSCF	complete active space self-consistent field method
cc-pVDZ	correlation-consistent polarized valence double- ζ basis set of Dunning
cc-pVTZ	correlation-consistent polarized valence triple- ζ basis set of Dunning
CC	coupled cluster
CCSD	coupled cluster with singles and doubles substitutions method
CCSD(T)	coupled cluster with singles and doubles and perturbative triples substitutions method
CEPA(0)	coupled electron-pair approximation of zeroth order
CI	configuration interaction
CISD	configuration interaction with singles and doubles substitutions method
DFT	density functional theory
dsp	domain size parameter
ECP	effective core potential
FCI	full configuration interaction method
kcal/mol	kilocalories per mol
MCSCF	multi-configuration self-consistent field method

MDF	ECP generated on the basis of all-electron Dirac-Fock calculations
MWB	ECP generated on the basis of all-electron calculations using a Wood-Boring Hamiltonian
HF	Hartree-Fock approach
LMO	localized molecular orbital
MO	molecular orbital
MP2	Møller-Plesset perturbation theory of second-order
MPn	Møller-Plesset perturbation theory of n -th order
MR	multi-reference
MR-CI	multi-reference configuration interaction method
MR-CISD	multi-reference configuration interaction with singles and doubles substitutions method
nop	number of parts parameter
occ	occupied orbitals
PAO	projected atomic orbital
PES	potential energy surface
RCCSD	restricted open-shell coupled cluster with singles and doubles substitutions method
RI	resolution of identity (density fitting)
SCF	self-consistent field method
STL	standard template library of C++
SVP	standard basis in TURBOMOLE 5.6 of polarized double- ζ quality
TS	transition state
ZPE	zero-point energy correction
6-31G**	basis of polarized double- ζ quality
6-31G*	double- ζ basis with polarization functions on all atoms of the second period or higher
6-31G	double- ζ basis set

Appendix B

The Incremental Code

B.1 Required Libraries

The incremental code is completely written in C++. It was tested on a cluster of Pentium IV PC's. In order to have a fast matrix multiplication and diagonalization it uses the Lapack and Blas Libraries. For a convenient integration in C++ the CPPLapack interface was used. Since FORTRAN numbers are not necessarily written in the C-standard format we check the read numbers by a conversion class (MOLCAS). In order to avoid an excessive case differentiation we used the regular expressions form regexx library for this purpose. For the server/client communication the socket++ library is used. The Graph partitioning is done with METIS-graph partitioning using the METIS library. All together we need the following libraries installed:

- liblapack.a
- libblas.a
- cpplapack.h
- libregexx.a
- socket++
- libmetis.a
- libf2c.a

The libraries are included in the Makefile.conf.local files in the SolidMRCC/Library/-module_name/test directory. The actual path of a given library can be adapted in this file, too.

B.2 Quantum Chemistry Packages

For all three quantum chemistry packages (MOLPRO, MOLCAS, DALTON) we modified the source code in order to obtain the required data in the desired precision. These modifications have to be adapted in a new installation or in an update of the version.

B.3 Molpro Input Example

The usage of ECP's is allowed. The most convenient way to use an ECP in our framework is to extract the ECP and basis set data from Gaussian basis set order form [142] and paste it to the input file.

B.4 MOLCAS 6.4 Dependencies

In order to use the 6.4 version of MOLCAS to calculate the MO's and the transformed one- and two-particle integrals we installed the DIESEL patch into MOLCAS 6.4. A picture of the data flow for incremental MR-CISD, ACPF, AQCC and CEPA(0) calculations within the MOLCAS/DIESEL environment is given in figure B.1.

B.5 Dalton Dependencies

Figure B.2 gives an overview of the data flow for incremental CCSD calculations within the DALTON framework. For property calculations the data structure is similar. The main difference is that the server, the clients and the wrapper classes for the network transfer are different. The data container in the server and the wrapper classes have to include the property and the client has to do different I/O with DALTON.

```
***,  
  
MEMORY,170,M;  
geomtyp=xyz  
geometry={  
  nosymm  
  16  
  B5H11  
  B   -0.1890489    0.7287746   -1.5817852  
  B   -0.1353763   -0.8908484   -0.9001497  
  B   -0.0752367   -0.9533659    0.9031186  
  H   -1.0068527   -1.4188981    0.0452508  
  B   -0.2285015    0.7612458    1.5383154  
  B    0.7575139    0.3037702    0.0050513  
  H    0.4672242    0.9182539    2.5248233  
  H   -1.0886271   -0.3515950   -1.6382591  
  H    0.1889543    1.4138737    0.3095822  
  H    0.4153058   -1.8148292   -1.4596877  
  H    0.4753712   -1.8912681    1.4370111  
  H   -1.2484971    1.4395625    1.4641646  
  H   -0.9673691   -0.4327263    1.6981598  
  H   -1.0017159    1.6177074   -1.3855590  
  H    0.4094919    0.7257819   -2.6436917  
  H    1.9693627    0.3349574    0.0284204}  
  
PUNCH test;  
basis=6-31G**;  
GTHRESH,ENERGY=1.d-10;  
hf  
orbprint,56;  
matrop;  
LOAD,ORBITAL,ORB;  
PRINT,ORBITAL;  
LOAD,OVERLAP,S;  
PRINT,OVERLAP;  
LOAD,opx,OPER,DMX;  
PRINT,opx;  
LOAD,opy,OPER,DY;  
PRINT,opy;  
LOAD,opz,OPER,DMZ;  
PRINT,opz;  
write,ORBITAL,test;
```

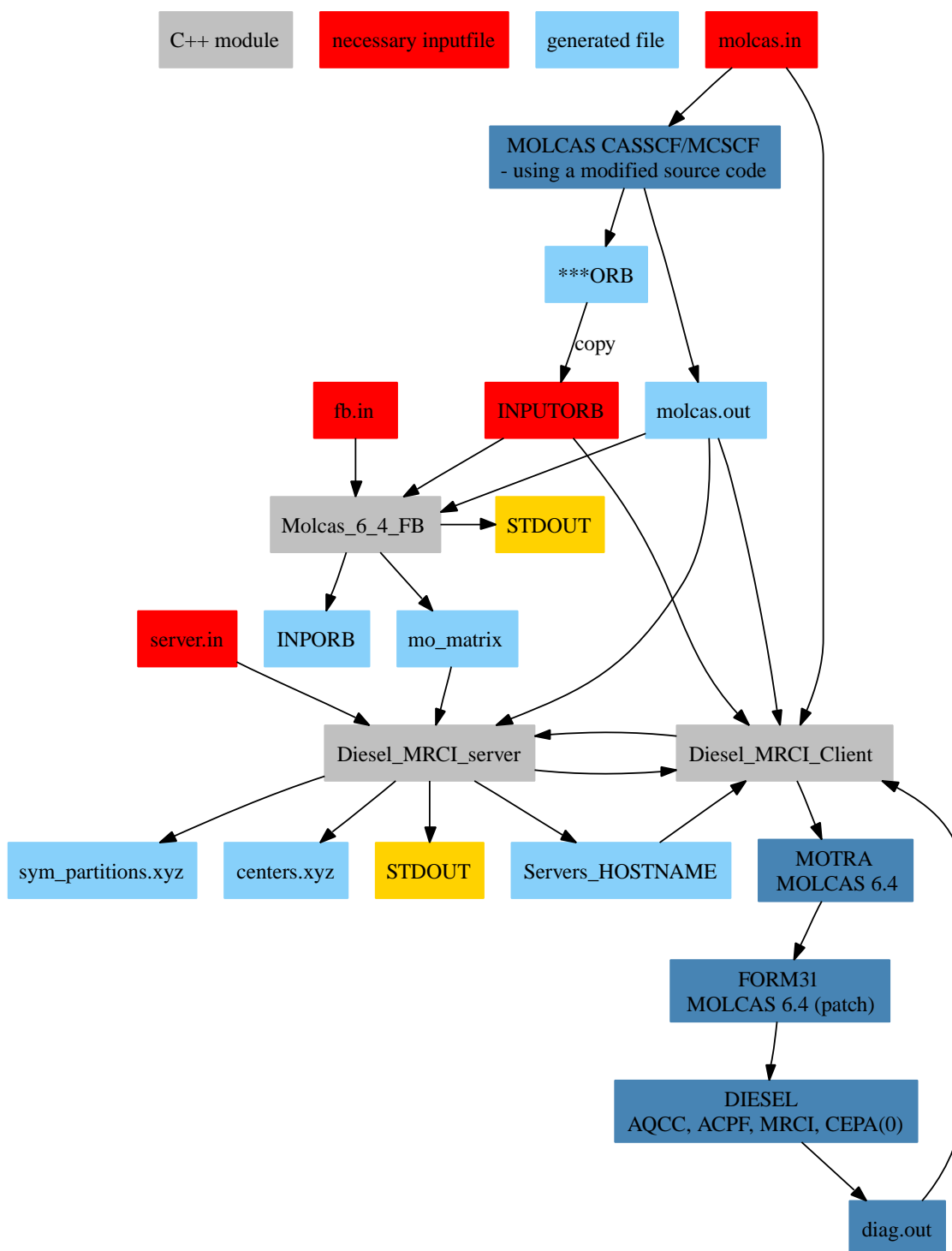


Figure B.1: File dependencies for CI-based calculations within the MOLCAS 6.4 and DIESEL framework.

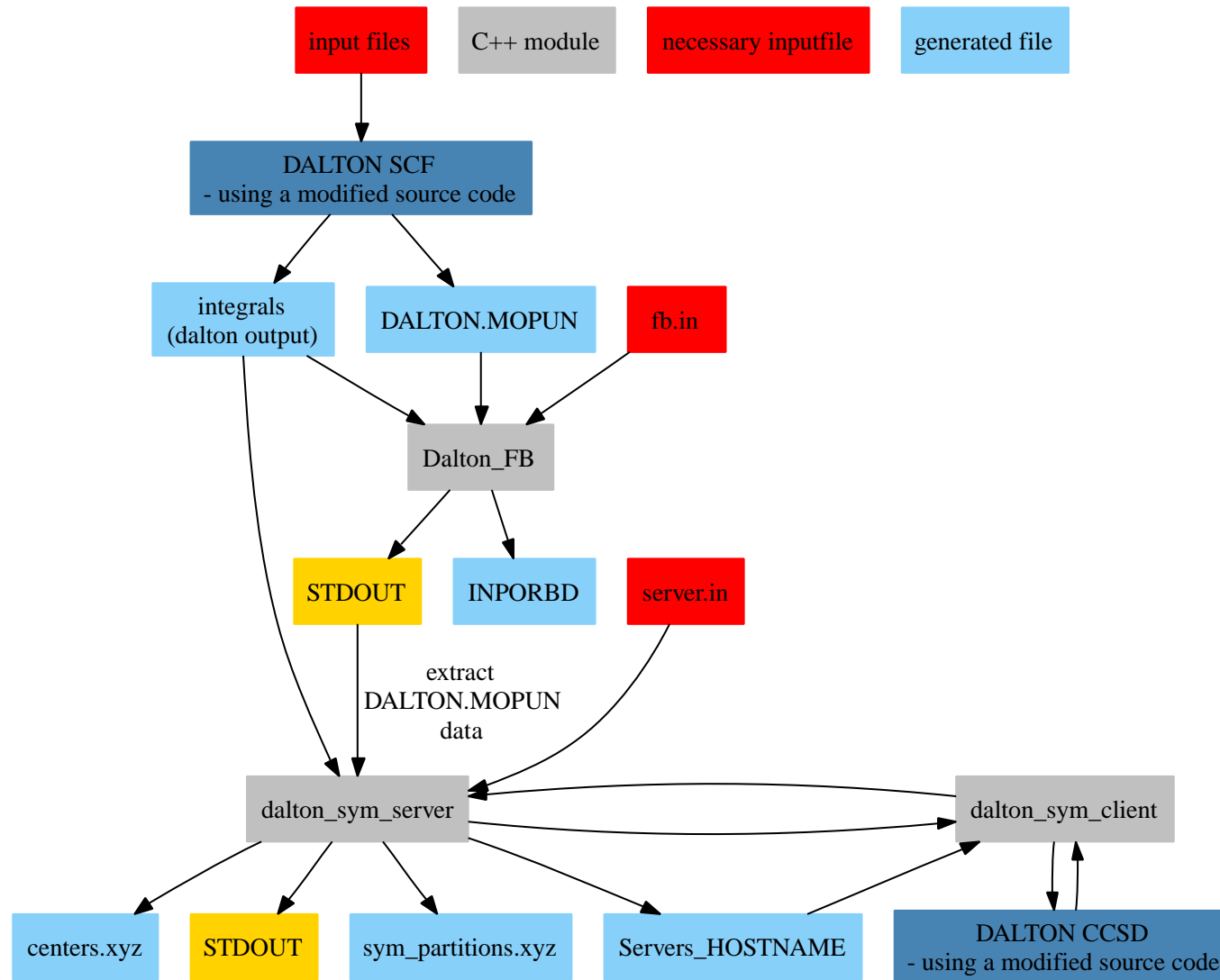


Figure B.2: File dependencies for the incremental CCSD calculations within the DALTON framework.

Appendix C

Molecular Data

C.1 Exemplary Lists of Increments

1st Order		2nd Order	
Combination	E [a.u]	Combination	E [a.u]
0	-2.023440305500e-01	0 1	-3.064693590000e-03
1	-2.023439138770e-01	0 2	-3.064542781000e-03
2	-2.023440006240e-01	0 3	-3.064692038000e-03
3	-2.023439422190e-01	0 4	-3.064541640000e-03
4	-2.023440289280e-01	0 5	-2.183536504000e-03
5	-2.023439166520e-01	1 2	-3.064697970000e-03
3rd Order		1 3	-3.064566762000e-03
Combination	E [a.u]	1 4	-2.183537037000e-03
0 1 2	3.318627070000e-04	1 5	-3.064561147000e-03
0 1 3	3.318624680000e-04	2 3	-2.183536117000e-03
0 1 4	1.786724220000e-04	2 4	-3.064549303000e-03
0 1 5	1.786743770001e-04	2 5	-3.064698248000e-03
0 2 3	1.786709700000e-04	3 4	-3.064689295000e-03
0 2 4	3.318595760000e-04	3 5	-3.064563025000e-03
0 2 5	1.786723570000e-04	4 5	-3.064693862000e-03
0 3 4	3.318617370000e-04	4th Order	
0 3 5	1.786740760000e-04	Combination	E [a.u]
0 4 5	1.786721120001e-04	0 1 2 3	-6.277379600000e-05
1 2 3	1.786757150000e-04	0 1 2 4	-6.277439499991e-05
1 2 4	1.786724040000e-04	0 1 2 5	-6.277393499998e-05
1 2 5	3.318633770000e-04	0 1 3 4	-6.277367500004e-05
1 3 4	1.786739860001e-04	0 1 3 5	-6.277498800009e-05
1 3 5	3.318622680001e-04	0 1 4 5	-1.727596000003e-05
1 4 5	1.786746860001e-04	0 2 3 4	-6.277410099986e-05
2 3 4	1.786716889999e-04	0 2 3 5	-1.727589299994e-05
2 3 5	1.786749639999e-04	0 2 4 5	-6.277433500007e-05
2 4 5	3.318619460000e-04	0 3 4 5	-6.277378200001e-05
3 4 5	3.318632139999e-04	1 2 3 4	-1.727602200011e-05
		1 2 3 5	-6.277526400003e-05
		1 2 4 5	-6.277382899997e-05
		1 3 4 5	-6.277503300012e-05
		2 3 4 5	-6.277379399985e-05

Table C.1: List of the incremental energies of MoF₆.

1st Order		2nd Order	
Combination	E [a.u.]	Combination	E [a.u.]
0	-1.639168251110e-01	0 1	-6.377502534000e-03
1	-1.639168124940e-01	0 2	-6.377028175000e-03
2	-1.639168217730e-01	0 3	-6.377502445000e-03
3	-1.639168144960e-01	0 4	-6.377033329000e-03
4	-1.639168201900e-01	0 5	-4.782358658000e-03
5	-1.639168201860e-01	1 2	-6.377505556000e-03
		1 3	-6.377038392000e-03
		1 4	-4.782359830000e-03
		1 5	-6.377035920000e-03
		2 3	-4.782360032000e-03
		2 4	-6.377035465000e-03
		2 5	-6.377502152000e-03
		3 4	-6.377506054000e-03
		3 5	-6.377030975000e-03
		4 5	-6.377502701000e-03
3rd Order		4th Order	
Combination	E [a.u.]	Combination	E [a.u.]
0 1 2	1.366408537000e-03	0 1 2 3	-4.869852500000e-04
0 1 3	1.366408990000e-03	0 1 2 4	-4.869942750000e-04
0 1 4	8.943944420000e-04	0 1 2 5	-4.869850940001e-04
0 1 5	8.943940140000e-04	0 1 3 4	-4.869850780000e-04
0 2 3	8.943933640000e-04	0 1 3 5	-4.869946160000e-04
0 2 4	1.366399406000e-03	0 1 4 5	-3.076106720001e-04
0 2 5	8.943943280000e-04	0 2 3 4	-4.869942010000e-04
0 3 4	1.366408179000e-03	0 2 3 5	-3.076105610000e-04
0 3 5	8.943948620000e-04	0 2 4 5	-4.869943040001e-04
0 4 5	8.943934290000e-04	0 3 4 5	-4.869850950000e-04
1 2 3	8.943952350001e-04	1 2 3 4	-3.076108310002e-04
1 2 4	8.943936700000e-04	1 2 3 5	-4.869944380001e-04
1 2 5	1.366408321000e-03	1 2 4 5	-4.869850150001e-04
1 3 4	8.943942740001e-04	1 3 4 5	-4.869945750000e-04
1 3 5	1.366400359000e-03	2 3 4 5	-4.869852239999e-04
1 4 5	8.943949630000e-04		
2 3 4	8.943948780000e-04		
2 3 5	8.943936880000e-04		
2 4 5	1.366408594000e-03		
3 4 5	1.366408896000e-03		

Table C.2: List of the incremental energies of MoCl₆.

C.2 Lists of Increments for CI-based Methods

Multiplicity 1		Multiplicity 3	
Combination	E [a.u.]	Combination	E [a.u.]
1st Order		1st Order	
0	-6.535316300000e-02	0	-6.535316400000e-02
1	-6.535319500000e-02	1	-6.535319900000e-02
2	-6.526644400000e-02	2	-6.508440500000e-02
3	-6.526644400000e-02	3	-6.508440500000e-02
4	-6.535319600000e-02	4	-6.535319900000e-02
5	-6.535316300000e-02	5	-6.535316300000e-02
2nd Order		2nd Order	
0 1	2.926792000000e-03	0 1	2.926792000000e-03
0 2	3.004811000000e-03	0 2	3.160259000000e-03
0 3	3.004814000000e-03	0 3	3.160262000000e-03
0 4	2.926808000000e-03	0 4	2.926808000000e-03
0 5	2.926800000000e-03	0 5	2.926800000000e-03
1 2	3.004670000000e-03	1 2	3.160118000000e-03
1 3	3.004814000000e-03	1 3	3.160261000000e-03
1 4	2.926807000000e-03	1 4	2.926817000000e-03
1 5	2.926807000000e-03	1 5	2.926807000000e-03
2 3	2.913654000000e-03	2 3	2.885662000000e-03
2 4	3.004816000000e-03	2 4	3.160262000000e-03
2 5	3.004815000000e-03	2 5	3.160261000000e-03
3 4	3.004671000000e-03	3 4	3.160118000000e-03
3 5	3.004811000000e-03	3 5	3.160258000000e-03
4 5	2.926793000000e-03	4 5	2.926792000000e-03
3rd Order		3rd Order	
0 1 2	-4.103410000000e-04	0 1 2	-4.323960000000e-04
0 1 3	-4.103470000000e-04	0 1 3	-4.324020000000e-04
0 1 4	-3.996910000000e-04	0 1 4	-3.997000000000e-04
0 1 5	-3.996880000000e-04	0 1 5	-3.996870000000e-04
0 2 3	-4.880470000000e-04	0 2 3	-6.652760000000e-04
0 2 4	-4.103380000000e-04	0 2 4	-4.324030000000e-04
0 2 5	-4.103400000000e-04	0 2 5	-4.324040000000e-04
0 3 4	-4.103410000000e-04	0 3 4	-4.324070000000e-04
0 3 5	-4.103350000000e-04	0 3 5	-4.324030000000e-04
0 4 5	-3.996910000000e-04	0 4 5	-3.996900000000e-04
1 2 3	-4.880400000000e-04	1 2 3	-6.652680000000e-04
1 2 4	-4.103420000000e-04	1 2 4	-4.324070000000e-04
1 2 5	-4.103430000000e-04	1 2 5	-4.324070000000e-04
1 3 4	-4.103410000000e-04	1 3 4	-4.324060000000e-04
1 3 5	-4.103380000000e-04	1 3 5	-4.324030000000e-04
1 4 5	-3.996860000000e-04	1 4 5	-3.996950000000e-04
2 3 4	-4.880410000000e-04	2 3 4	-6.652680000000e-04
2 3 5	-4.880470000000e-04	2 3 5	-6.652740000000e-04
2 4 5	-4.103500000000e-04	2 4 5	-4.324030000000e-04
3 4 5	-4.103410000000e-04	3 4 5	-4.323960000000e-04
4th Order		4th Order	
0 1 2 3	9.910099999999e-05	0 1 2 3	1.258210000000e-04
0 1 2 4	8.848500000003e-05	0 1 2 4	9.318100000001e-05
0 1 2 5	8.848599999998e-05	0 1 2 5	9.318899999999e-05
0 1 3 4	8.849000000002e-05	0 1 3 4	9.318700000004e-05
0 1 3 5	8.847200000002e-05	0 1 3 5	9.318000000000e-05
0 1 4 5	8.624699999997e-05	0 1 4 5	8.625300000002e-05
0 2 3 4	9.908900000000e-05	0 2 3 4	1.258290000000e-04
0 2 3 5	9.909300000001e-05	0 2 3 5	1.258330000000e-04
0 2 4 5	8.847900000002e-05	0 2 4 5	9.318300000000e-05
0 3 4 5	8.848600000001e-05	0 3 4 5	9.319599999999e-05
1 2 3 4	9.909900000005e-05	1 2 3 4	1.258190000000e-04
1 2 3 5	9.909300000001e-05	1 2 3 5	1.258310000000e-04
1 2 4 5	8.849100000004e-05	1 2 4 5	9.318500000002e-05
1 3 4 5	8.847900000002e-05	1 3 4 5	9.317600000003e-05
2 3 4 5	9.910000000000e-05	2 3 4 5	1.258190000000e-04

Table C.3: List of the MR-CISD-incremental energies of H_{28} .

Multiplicity 1		Multiplicity 3	
Combination	E [a.u.]	Combination	E [a.u.]
1st Order		1st Order	
0	-6.655497500000e-02	0	-6.655497600000e-02
1	-6.655501000000e-02	1	-6.655501400000e-02
2	-6.653556500000e-02	2	-6.648943600000e-02
3	-6.653556500000e-02	3	-6.648943600000e-02
4	-6.655501000000e-02	4	-6.655501400000e-02
5	-6.655497500000e-02	5	-6.655497600000e-02
2nd Order		2nd Order	
0 1	2.769260000000e-04	0 1	2.769260000000e-04
0 2	2.851700000000e-04	0 2	3.019970000000e-04
0 3	2.851730000000e-04	0 3	3.020000000000e-04
0 4	2.769430000000e-04	0 4	2.769440000000e-04
0 5	2.769380000000e-04	0 5	2.769390000000e-04
1 2	2.850130000000e-04	1 2	3.018390000000e-04
1 3	2.851710000000e-04	1 3	3.019960000000e-04
1 4	2.769400000000e-04	1 4	2.769500000000e-04
1 5	2.769430000000e-04	1 5	2.769430000000e-04
2 3	2.606640000000e-04	2 3	2.293250000000e-04
2 4	2.851720000000e-04	2 4	3.019970000000e-04
2 5	2.851730000000e-04	2 5	3.020000000000e-04
3 4	2.850130000000e-04	3 4	3.018390000000e-04
3 5	2.851700000000e-04	3 5	3.019970000000e-04
4 5	2.769270000000e-04	4 5	2.769270000000e-04
3rd Order		3rd Order	
0 1 2	-2.131250000000e-04	0 1 2	-2.207940000000e-04
0 1 3	-2.131240000000e-04	0 1 3	-2.207930000000e-04
0 1 4	-2.091570000000e-04	0 1 4	-2.091680000000e-04
0 1 5	-2.091550000000e-04	0 1 5	-2.091550000000e-04
0 2 3	-2.213320000000e-04	0 2 3	-2.458180000000e-04
0 2 4	-2.131150000000e-04	0 2 4	-2.207950000000e-04
0 2 5	-2.131160000000e-04	0 2 5	-2.207960000000e-04
0 3 4	-2.131240000000e-04	0 3 4	-2.208049999999e-04
0 3 5	-2.131120000000e-04	0 3 5	-2.207920000000e-04
0 4 5	-2.091580000000e-04	0 4 5	-2.091600000000e-04
1 2 3	-2.213320000000e-04	1 2 3	-2.458160000000e-04
1 2 4	-2.131250000000e-04	1 2 4	-2.208040000000e-04
1 2 5	-2.131250000000e-04	1 2 5	-2.208050000000e-04
1 3 4	-2.131240000000e-04	1 3 4	-2.208040000000e-04
1 3 5	-2.131160000000e-04	1 3 5	-2.207950000000e-04
1 4 5	-2.091520000000e-04	1 4 5	-2.091630000000e-04
2 3 4	-2.213320000000e-04	2 3 4	-2.458170000000e-04
2 3 5	-2.213310000000e-04	2 3 5	-2.458170000000e-04
2 4 5	-2.131260000000e-04	2 4 5	-2.207950000000e-04
3 4 5	-2.131250000000e-04	3 4 5	-2.207950000000e-04
4th Order		4th Order	
0 1 2 3	1.746130000000e-04	0 1 2 3	1.868800000000e-04
0 1 2 4	1.706560000000e-04	0 1 2 4	1.752450000000e-04
0 1 2 5	1.706550000000e-04	0 1 2 5	1.752540000000e-04
0 1 3 4	1.706590000000e-04	0 1 3 4	1.752509999999e-04
0 1 3 5	1.706420000000e-04	0 1 3 5	1.752409999999e-04
0 1 4 5	1.682570000000e-04	0 1 4 5	1.682700000000e-04
0 2 3 4	1.746010000000e-04	0 2 3 4	1.868909999999e-04
0 2 3 5	1.746020000000e-04	0 2 3 5	1.868910000000e-04
0 2 4 5	1.706470000000e-04	0 2 4 5	1.752479999999e-04
0 3 4 5	1.706580000000e-04	0 3 4 5	1.752599999999e-04
1 2 3 4	1.746110000000e-04	1 2 3 4	1.868809999999e-04
1 2 3 5	1.746040000000e-04	1 2 3 5	1.868930000000e-04
1 2 4 5	1.706580000000e-04	1 2 4 5	1.752489999999e-04
1 3 4 5	1.706500000000e-04	1 3 4 5	1.752409999999e-04
2 3 4 5	1.746100000000e-04	2 3 4 5	1.868800000000e-04

Table C.4: List of the MR-ACPF-incremental energies of H₂₈.

Multiplicity 1		Multiplicity 3	
Combination	E [a.u.]	Combination	E [a.u.]
1st Order		1st Order	
0	-6.608552200000e-02	0	-6.608552300000e-02
1	-6.608555600000e-02	1	-6.608556000000e-02
2	-6.603870200000e-02	2	-6.593272400000e-02
3	-6.603870200000e-02	3	-6.593272400000e-02
4	-6.608555600000e-02	4	-6.608556000000e-02
5	-6.608552200000e-02	5	-6.608552300000e-02
2nd Order		2nd Order	
0 1	2.563067000000e-03	0 1	2.563069000000e-03
0 2	2.633805000000e-03	0 2	2.776714000000e-03
0 3	2.633808000000e-03	0 3	2.776717000000e-03
0 4	2.563083000000e-03	0 4	2.563086000000e-03
0 5	2.563076000000e-03	0 5	2.563078000000e-03
1 2	2.633659000000e-03	1 2	2.776568000000e-03
1 3	2.633808000000e-03	1 3	2.776717000000e-03
1 4	2.563082000000e-03	1 4	2.563094000000e-03
1 5	2.563083000000e-03	1 5	2.563085000000e-03
2 3	2.582320000000e-03	2 3	2.617646000000e-03
2 4	2.633809000000e-03	2 4	2.776717000000e-03
2 5	2.633808000000e-03	2 5	2.776717000000e-03
3 4	2.633659000000e-03	3 4	2.776568000000e-03
3 5	2.633805000000e-03	3 5	2.776714000000e-03
4 5	2.563068000000e-03	4 5	2.563070000000e-03
3rd Order		3rd Order	
0 1 2	2.931700000000e-05	0 1 2	2.196600000001e-05
0 1 3	2.931100000000e-05	0 1 3	2.195999999999e-05
0 1 4	3.221700000001e-05	0 1 4	3.220200000000e-05
0 1 5	3.221899999999e-05	0 1 5	3.221499999999e-05
0 2 3	-4.110200000000e-05	0 2 3	-1.910660000000e-04
0 2 4	2.932000000000e-05	0 2 4	2.195900000002e-05
0 2 5	2.931800000000e-05	0 2 5	2.195799999999e-05
0 3 4	2.931700000000e-05	0 3 4	2.195500000002e-05
0 3 5	2.932200000000e-05	0 3 5	2.196099999999e-05
0 4 5	3.221599999999e-05	0 4 5	3.221099999999e-05
1 2 3	-4.109399999998e-05	1 2 3	-1.910590000000e-04
1 2 4	2.931700000003e-05	1 2 4	2.195699999999e-05
1 2 5	2.931600000000e-05	1 2 5	2.195500000002e-05
1 3 4	2.931800000003e-05	1 3 4	2.195599999999e-05
1 3 5	2.931900000000e-05	1 3 5	2.195800000002e-05
1 4 5	3.222200000001e-05	1 4 5	3.220699999999e-05
2 3 4	-4.109399999998e-05	2 3 4	-1.910580000000e-04
2 3 5	-4.110100000000e-05	2 3 5	-1.910650000000e-04
2 4 5	2.930900000001e-05	2 4 5	2.195899999999e-05
3 4 5	2.931600000000e-05	3 4 5	2.196600000001e-05
4th Order		4th Order	
0 1 2 3	-2.907690000000e-04	0 1 2 3	-2.894600000000e-04
0 1 2 4	-2.936450000000e-04	0 1 2 4	-2.996500000000e-04
0 1 2 5	-2.936440000000e-04	0 1 2 5	-2.996410000000e-04
0 1 3 4	-2.936420000000e-04	0 1 3 4	-2.996430000000e-04
0 1 3 5	-2.936590000000e-04	0 1 3 5	-2.996520000000e-04
0 1 4 5	-2.903450000000e-04	0 1 4 5	-2.903260000000e-04
0 2 3 4	-2.907810000000e-04	0 2 3 4	-2.894520000000e-04
0 2 3 5	-2.907770000000e-04	0 2 3 5	-2.894500000000e-04
0 2 4 5	-2.936540000000e-04	0 2 4 5	-2.996480000000e-04
0 3 4 5	-2.936430000000e-04	0 3 4 5	-2.996360000000e-04
1 2 3 4	-2.907730000001e-04	1 2 3 4	-2.894629999999e-04
1 2 3 5	-2.907770000000e-04	1 2 3 5	-2.894490000000e-04
1 2 4 5	-2.936410000000e-04	1 2 4 5	-2.996460000000e-04
1 3 4 5	-2.936520000000e-04	1 3 4 5	-2.996540000000e-04
2 3 4 5	-2.907700000000e-04	2 3 4 5	-2.894630000000e-04

Table C.5: List of the MR-AQCC-incremental energies of H₂₈.

Multiplicity 1		Multiplicity 3	
Combination	E [a.u.]	Combination	E [a.u.]
1st Order		1st Order	
0	-6.697835200000e-02	0	-6.697835100000e-02
1	-6.697838800000e-02	1	-6.697839100000e-02
2	-6.698298200000e-02	2	-6.698317000000e-02
3	-6.698298200000e-02	3	-6.698317000000e-02
4	-6.697838800000e-02	4	-6.697839100000e-02
5	-6.697835200000e-02	5	-6.697835100000e-02
2nd Order		2nd Order	
0 1	-1.600000000546e-08	0 1	-1.900000000388e-08
0 2	4.679999999896e-07	0 2	2.401000000013e-06
0 3	4.709999999880e-07	0 3	2.404000000011e-06
0 4	2.000000012825e-09	0 4	-9.999999855959e-10
0 5	-3.999999997895e-09	0 5	-6.999999996316e-09
1 2	3.080000000044e-07	1 2	2.240000000001e-06
1 3	4.690000000029e-07	1 3	2.400999999999e-06
1 4	-1.999999998947e-09	1 4	4.999999997368e-09
1 5	1.000000013351e-09	1 5	-1.999999985070e-09
2 3	-1.599999999990e-07	2 3	-2.200000000230e-07
2 4	4.700000000024e-07	2 4	2.400999999999e-06
2 5	4.719999999875e-07	2 5	2.404000000011e-06
3 4	3.080000000044e-07	3 4	2.240000000001e-06
3 5	4.679999999896e-07	3 5	2.400000000014e-06
4 5	-1.500000000598e-08	4 5	-1.800000000440e-08
3rd Order		3rd Order	
0 1 2	-3.299999998263e-08	0 1 2	-1.069999999853e-07
0 1 3	-3.199999998316e-08	0 1 3	-1.059999999858e-07
0 1 4	-3.000000026177e-09	0 1 4	-8.999999995263e-09
0 1 5	9.999999855959e-10	0 1 5	5.000000011246e-09
0 2 3	-4.979999999877e-07	0 2 3	-2.513000000010e-06
0 2 4	-2.299999998789e-08	0 2 4	-1.060000000414e-07
0 2 5	-2.399999994573e-08	0 2 5	-1.06999999992e-07
0 3 4	-3.299999998263e-08	0 3 4	-1.170000000356e-07
0 3 5	-1.799999997665e-08	0 3 5	-1.040000000008e-07
0 4 5	-3.000000012299e-09	0 4 5	-1.387778780781e-17
1 2 3	-4.990000000149e-07	1 2 3	-2.511999999982e-06
1 2 4	-3.300000001039e-08	1 2 4	-1.160000000083e-07
1 2 5	-3.39999999598e-08	1 2 5	-1.170000000217e-07
1 3 4	-3.200000001091e-08	1 3 4	-1.160000000083e-07
1 3 5	-2.300000000177e-08	1 3 5	-1.0600000000275e-07
1 4 5	2.999999984543e-09	1 4 5	-3.000000012299e-09
2 3 4	-4.990000000149e-07	2 3 4	-2.51099999997e-06
2 3 5	-4.97999999738e-07	2 3 5	-2.510999999983e-06
2 4 5	-3.49999999546e-08	2 4 5	-1.069999999853e-07
3 4 5	-3.29999999651e-08	3 4 5	-1.060000000136e-07
4th Order		4th Order	
0 1 2 3	4.700000001689e-08	0 1 2 3	1.239999999486e-07
0 1 2 4	1.500000000598e-08	0 1 2 4	9.999999994736e-10
0 1 2 5	1.299999996540e-08	0 1 2 5	8.999999925874e-09
0 1 3 4	1.900000000388e-08	0 1 3 4	5.99999996842e-09
0 1 3 5	-1.999999998947e-09	0 1 3 5	-3.000000067810e-09
0 1 4 5	-9.999999162069e-10	0 1 4 5	6.999999968560e-09
0 2 3 4	3.599999996717e-08	0 2 3 4	1.310000000282e-07
0 2 3 5	3.599999995330e-08	0 2 3 5	1.339999999295e-07
0 2 4 5	6.999999968560e-09	0 2 4 5	2.999999956788e-09
0 3 4 5	1.600000001933e-08	0 3 4 5	1.600000000546e-08
1 2 3 4	4.500000003183e-08	1 2 3 4	1.210000000196e-07
1 2 3 5	3.899999999335e-08	1 2 3 5	1.339999999989e-07
1 2 4 5	1.800000000440e-08	1 2 4 5	3.000000026177e-09
1 3 4 5	8.000000037423e-09	1 3 4 5	-5.999999969086e-09
2 3 4 5	4.599999998967e-08	2 3 4 5	1.199999999230e-07

Table C.6: List of the MR-CEPA(0)-incremental energies of H₂₈.

C.3 4-*exo*/5-*endo* Cyclization

system	order i	i -th order correction [au]	$E_{\text{corr}}(i)$ [au]	error [kcal/mol]	% E_{corr}
16 a					
	1	-0.908354	-0.908354	204.50	73.60
	2	-0.333906	-1.242260	-5.03	100.65
	3	0.008175	-1.234085	0.10	99.99
	4	-0.000174	-1.234259	-0.01	100.00
exact CCSD			-1.234249		
16 b					
	1	-0.907860	-0.907860	204.55	73.58
	2	-0.333982	-1.241842	-5.02	100.65
	3	0.008126	-1.233716	0.08	99.99
	4	-0.000120	-1.233836	0.00	100.00
exact CCSD			-1.233836		
16 c					
	1	-0.907649	-0.907649	204.71	73.56
	2	-0.334255	-1.241904	-5.04	100.65
	3	0.008151	-1.233753	0.07	99.99
	4	-0.000116	-1.233869	0.00	100.00
exact CCSD			-1.233869		
14 a					
	1	-0.911709	-0.911709	198.12	74.28
	2	-0.322797	-1.234506	-4.44	100.58
	3	0.007072	-1.227435	0.00	100.00
	4	-0.000002	-1.227437	0.00	100.00
exact CCSD			-1.227435		

Table C.7: Convergence behavior for the incremental RCCSD/6-31G** correlation energy of the intermediates of the cyclization of **14**. (dsp=3, core=8)

system	order i	i -th order correction [au]	$E_{\text{corr}}(i)$ [au]	error [kcal/mol]	% E_{corr}
14 b					
	1	-0.915735	-0.915735	194.87	74.68
	2	-0.318038	-1.233773	-4.70	100.61
	3	0.007503	-1.226270	0.01	100.00
	4	-0.000026	-1.226296	-0.01	100.00
exact CCSD			-1.226281		
14 c					
	1	-0.916631	-0.916631	194.25	74.75
	2	-0.316954	-1.233586	-4.64	100.60
	3	0.007401	-1.226184	0.00	100.00
	4	-0.000013	-1.226197	-0.01	100.00
exact CCSD			-1.226187		
13 a					
	1	-0.965772	-0.965772	169.30	78.16
	2	-0.277944	-1.243716	-5.11	100.66
	3	0.008119	-1.235597	-0.02	100.00
	4	0.000063	-1.235534	0.02	100.00
exact CCSD			-1.235571		
13 b					
	1	-0.965237	-0.965237	170.26	78.06
	2	-0.279755	-1.244992	-5.29	100.68
	3	0.008430	-1.236562	0.00	100.00
	4	0.000024	-1.236537	0.02	100.00
exact CCSD			-1.236569		
13 c					
	1	-0.965306	-0.965306	169.54	78.13
	2	-0.278399	-1.243705	-5.16	100.67
	3	0.008187	-1.235518	-0.02	100.00
	4	0.000069	-1.235449	0.02	100.00
exact CCSD			-1.235483		

Table C.8: Convergence behavior for the incremental RCCSD/6-31G** correlation energy of the intermediates of the cyclization of **14**. (dsp=3, core=8)

system	order i	i -th order correction [au]	$E_{\text{corr}}(i)$ [au]	error [kcal/mol]	% E_{corr}
TS 12 a					
	1	-0.924125	-0.924125	198.34	74.51
	2	-0.323896	-1.248021	-4.91	100.63
	3	0.007964	-1.240057	0.09	99.99
	4	-0.000167	-1.240224	-0.02	100.00
exact CCSD			-1.240199		
TS 12 b					
	1	-0.925685	-0.925685	198.76	74.51
	2	-0.324541	-1.250226	-4.89	100.63
	3	0.007952	-1.242274	0.10	99.99
	4	-0.000181	-1.242455	-0.01	100.00
exact CCSD			-1.242433		
TS 12 c					
	1	-0.930893	-0.930893	194.57	-75.01
	2	-0.317796	-1.248688	-4.85	-100.62
	3	0.007865	-1.240824	0.09	-99.99
	4	-0.000182	-1.241006	-0.03	-100.00
exact CCSD			-1.240961		
TS 15 a					
	1	-0.920398	-0.920398	204.39	73.86
	2	-0.335680	-1.256078	-6.25	100.80
	3	0.010135	-1.245943	0.11	99.99
	4	-0.000167	-1.246110	0.00	100.00
exact CCSD			-1.246115		
	1	-0.919738	-0.919738	204.65	73.82
	2	-0.336376	-1.256114	-6.43	100.82
	3	0.010573	-1.245540	0.21	99.97
	4	-0.000332	-1.245872	0.00	100.00
exact CCSD			-1.245870		

Table C.9: Convergence behavior for the incremental RCCSD/6-31G** correlation energy of the intermediates of the cyclization of **14**. (dsp=3, core=8)

order i	i-th order correction	$E_{\text{corr}}(i)$	order i	i-th order correction	$E_{\text{corr}}(i)$
16 a			14 a		
1	-1.092849	-1.092849	1	-1.093156	-1.093156
2	-0.405839	-1.498688	2	-0.397282	-1.490438
3	0.008991	-1.489697	3	0.008124	-1.482315
16 b			14 b		
1	-1.092181	-1.092181	1	-1.102117	-1.102117
2	-0.405937	-1.498118	2	-0.387245	-1.489362
3	0.008894	-1.489225	3	0.008373	-1.480989
16 c			14 c		
1	-1.092154	-1.092154	1	-1.103006	-1.103006
2	-0.406000	-1.498153	2	-0.385985	-1.488991
3	0.008913	-1.489241	3	0.008250	-1.480741
13 a			TS 12 a		
1	-1.163704	-1.163704	1	-1.105809	-1.105809
2	-0.336198	-1.499903	2	-0.398601	-1.504410
3	0.008745	-1.491158	3	0.009017	-1.495393
13 b			TS 12 b		
1	-1.162791	-1.162791	1	-1.107251	-1.107251
2	-0.338636	-1.501427	2	-0.399377	-1.506628
3	0.009173	-1.492255	3	0.009052	-1.497576
13 c			TS 12 c		
1	-1.163037	-1.163037	1	-1.117433	-1.117433
2	-0.336725	-1.499761	2	-0.387324	-1.504757
3	0.008807	-1.490954	3	0.008762	-1.495995
TS 15 a					
1	-1.105403	-1.105403			
2	-0.406089	-1.511493			
3	0.010671	-1.500822			
TS 15 b					
1	-1.104931	-1.104931			
2	-0.407062	-1.511993			
3	0.011275	-1.500718			

Table C.10: Incremental RCCSD/cc-pVTZ correlation energies for different conformers on the PES of the 4-*exo*/5-*endo* cyclization of **14**. (dsp=3, core=8)

Acknowledgments

I would like to thank Prof. Dr. Michael Dolg for giving me the opportunity to work in the interesting and developing field of local correlation methods, for an excellent scientific environment and for supervision.

I would like to thank Dr. Michael Hanrath for introducing me to C++, Perl, Bash, for the fruitful theoretical discussions and for being an excellent supervisor.

I would like to thank Anna Engels-Putzka, Anna Moritz, Jonas Wiebke, Jun Yang, and Mark Burkatzki for carefully reading the manuscript of this thesis and for fruitful discussions about theory. I would like to thank my colleagues Dr. Alexander Schnurpfeil, Birgitt Börsch-Pulm, Martin Böhler, Rebecca Fondermann, Sombat Kettrat, Dr. Xiaoyan Cao-Dolg for the nice time at the university of cologne.

Apart from work I want to thank my daughter Anna-Lena Boor, my girl friend Sandra Boor, my parents my brother and my sisters for spending time with me and for having patience with me.

Last but not least I gratefully acknowledge financial support of the German Science Foundation (DFG) through the priority programme 1145, the SFB 624 and the Graduiertenkolleg 549.

Abstract

A general fully automated implementation of the incremental scheme for molecules and embedded clusters in the framework of the Coupled Cluster singles and doubles theory is presented. The code can be applied to arbitrary order of the incremental expansion and is parallelized in a master/slave structure. We found that the error in the total correlation energy is lower than 1 kcal/mol with respect to the canonical CCSD calculation if the incremental series is truncated in a proper way. The potential accuracy of the incremental scheme is demonstrated explicitly for transition metal complexes, intermolecular systems, the aurophilic attraction, the 4-*exo*/5-*endo* cyclization and cluster compounds. The symmetry is exploited for a C_3 symmetric water cluster, a C_5 symmetric triazine pentamer and *all-trans* polyacetylene. For open-shell molecules the MR-CEPA(0)-method was adapted into the framework of the incremental scheme. Furthermore a systematic screening procedure for small contributions in the incremental expansion of the correlation energy is presented. The performance of the proposed scheme is checked for the calculation of intermolecular interactions in realistic test systems as large as a guanine-cytosine base pair. It was found that the computational cost for the incremental expansion can be considerably reduced without significant loss of accuracy. Typically the errors of the systems investigated here amount to less than 5 %, 1 % and 0.1 % for second, third and fourth order expansions, respectively. Additionally an analysis of the propagation of errors in the incremental expansion of the correlation energy is presented. Finally the performance of the incremental scheme in calculating potential energy surfaces is demonstrated.

Kurzzusammenfassung

In der vorliegenden Arbeit wird eine voll automatische Implementierung des Inkrementenverfahrens für CCSD und MR-CEPA(0) Korrelationsenergien vorgestellt. Mit der vorgestellten Implementation können Inkremente beliebiger Ordnung berechnet werden. Des weiteren wurde die Rechenzeit durch die Parallelisierung in einer Master/Slave-Struktur deutlich verkürzt. Mit dem vorgestellten Inkrementenverfahren kann die CCSD Energie mit einem Fehler von weniger als einer kcal/mol ermittelt werden. Dies wird durch Testrechnungen an gesättigten Kohlenwasserstoffen, an ungesättigten Kohlenwasserstoffen mit konjugierten π -Systemen, an aromatischen Verbindungen, an Übergangsmetallverbindungen, an Actinoidkomplexen, an intermolekularen Verbindungen, an den Intermediaten der 4-*exo*/5-*endo* Cyclisierung und an Clustern bestätigt. Des weiteren wird der Einfluss von näherungsweise symmetrischen Foster-Boys Orbitalen auf die Konvergenz der Inkrementenreihe behandelt. Explizite Rechnungen wurden in diesem Rahmen an einem C_3 symmetrischen Wassercluster, an dem C_5 symmetrischen Triazin-Pentamer und an Polyacetylen durchgeführt. Für offenschalige Moleküle wurden die MR-CEPA(0)-Methode, die MR-CISD-Methode, die MR-ACPF-Methode, die MR-AQCC-Methode und die RCCSD-Methode im Rahmen des Inkrementenverfahrens getestet.

Zur Beschleunigung der Inkrementen Rechnungen wurde ein Abstands basiertes Selektionskriterium implementiert. Mit diesem systematischen screening Verfahren kann die Rechenzeit bei gleicher Genauigkeit enorm verkürzt werden. Die Leistungsfähigkeit dieses screening Verfahrens wurde an realistischen Testsystemen wie z.B. dem Guanin-Cytosin-Basenpaar getestet. Die Fehler die im Rahmen der Inkrementen Rechnungen eingeführt wurden, sind typischerweise kleiner als 5 %, 1 % und 0.1 % der Korrelationsenergie für zweite, dritte und vierte Ordnung. Des weiteren wurde die Inkrementenmethode auf die Potentialhyperflächen von Octan und Dodeca-hexain angewendet. Schließlich wurde auch die Fehlerfortpflanzung in der Inkrementenmethode analysiert.

Lebenslauf

Name:	Joachim Friedrich
Geburtstag:	08.10.1977
Geburtsort:	Simmern
Familienstand:	Ledig

Ausbildung:	
Schule:	1984 - 1997 (Allgemeine Hochschulreife)
Studium:	1998 - 2004 (Diplom Chemie) Rheinische Friedrich-Wilhelms-Universität Bonn <i>Quantenchemische Untersuchungen Titanocenchlorid-induzierter 3-exo-Cyclisierungen an Epoxyolefinen</i>
Betreuer:	Prof. Dr. Michael Dolg
Dissertation	2004 - Implementation of the Incremental Scheme for highly efficient Correlation Methods
Betreuer:	Prof. Dr. Michael Dolg

Zivildienst	1997 - 1998
-------------	-------------

List of Publications

1. Joachim Friedrich, Michael Hanrath, Michael Dolg
Fully Automated Implementation of the Incremental Scheme: Application to CCSD Energies for Hydrocarbons and Transition Metal Compounds,
J. Chem. Phys. **126**, 154110, (2007)
2. Joachim Friedrich, Michael Hanrath, Michael Dolg
Energy Screening for the Incremental Scheme: Application to Intermolecular Interactions,
J. Phys. Chem. submitted
3. Joachim Friedrich, Michael Hanrath, Michael Dolg
Error Analysis of Incremental Electron Correlation Calculations and Applications to Clusters and Potential Energy Surfaces,
Chem. Phys. submitted

Erklärung

Ich versichere, daß ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; daß diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; daß sie - abgesehen von unten angegebenen Teilpublikationen - noch nicht veröffentlicht worden ist sowie, daß ich eine solche Veröffentlichung vor Abschluß des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Michael Dolg betreut worden.

Köln, 08.05.07