

Efficient Elimination of the Basis Set Superposition Error

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

vorgelegt von
Katarzyna Walczak
aus Ostrów Wielkopolski

Köln
2012

Berichterstatter: Prof. Dr. M. Dolg

Priv.-Doz. Dr. M. Hanrath

Tag der mündlichen Prüfung: 15.06.2012

Life is not easy for any of us. But what of that? We must have perseverance and above all confidence in ourselves. We must believe that we are gifted for something and that this thing must be attained.

Maria Skłodowska-Curie

*This dissertation is dedicated to the memory of my mum,
I will be eternally grateful for her love.*

Kurzzusammenfassung

Der Basissatzsuperpositionsfehler (BSSE) stellt eines der größten Hindernisse in quantenchemischen Berechnungen dar, die eine genaue Berechnung von Wechselwirkungsenergien anstreben. Die Bedeutung eines BSSE Eliminierungsverfahrens ist unter anderem darin begründet, dass der BSSE in der Größenordnung der zu berechnenden Wechselwirkungsenergie liegen kann und daher die Genauigkeit dieser signifikant beeinträchtigt.

In der vorliegenden Arbeit werden neue Ansätze vorgestellt, um den BSSE in wellenfunktionsbasierten quantenchemischen Rechnungen an großen molekularen Clustern effizient zu eliminieren. Die Anwendbarkeit der Methoden wurde ausführlich unter anderem an Wasser Clustern diskutiert, deren Größe sich von einem Wasser-Dimer bis hin zu einem $(\text{H}_2\text{O})_{20}$ Cluster erstreckte. Eine Übersicht über die in der Literatur bekannten BSSE Eliminierungsverfahren wird ebenfalls dargestellt und die hier vorgestellten Methoden werden damit verglichen.

Die neu vorgestellten Verfahren bieten mit nur geringeren Einbußen in der Genauigkeit der Rechnungen einen sehr effizienten Weg BSSE korrigierte Wechselwirkungsenergien zu erhalten, welche sogar teilweise mit den verfügbaren Standardmethoden aufgrund der Systemgröße nicht mehr zugänglich wären.

Abstract

The basis set superposition error (BSSE) is one of the major obstacle occurring in quantum chemical calculations which aim at a accurately prediction of interaction energies. The importance of a BSSE elimination procedure is among other things manifested by the fact, that the magnitude of the BSSE can be as large as the interaction energy itself, affecting the accuracy of the calculated interaction energies therefore significantly.

In this work new approaches to eliminate the BSSE efficiently from wavefunction based quantum chemical calculations on large molecular clusters are presented. The applicability of these schemes is studied in great detail among others on a water cluster series ranging in size from a water dimer up to even a $(\text{H}_2\text{O})_{20}$ water cluster. An overview of the correction schemes known from the literature is also given and the newly developed schemes are compared with the literature ones.

The presented schemes allow to account with only small loss in accuracy very efficiently for BSSE corrected interaction energies, which are partly no more feasible to calculate with standard methods due to the large system size.

Contents

Kurzzusammenfassung	vii
Abstract	ix
1 Introduction	3
2 Theory	5
2.1 Methods of Quantum Chemistry	5
2.1.1 Hartree-Fock Theory	5
2.1.2 Perturbation Theory	9
2.1.3 Coupled-Cluster Theory	11
2.1.4 Explicitly correlated methods	13
2.2 Incremental Scheme	15
2.3 Basis Set Superposition Error	22
3 Results and Discussion	25
3.1 Incremental evaluation of interaction energies	25
3.1.1 Introduction	25
3.1.2 Applications	32
3.2 BSSE correction schemes for n-body clusters	42
3.2.1 SSFC, PAFC and VMFC schemes	42
3.2.2 Approximate SSFC(R) and VMFC(2)(R)	45
3.2.3 Approximate SSFC _{inc} scheme	47
3.2.4 Approximate SSFC(S) scheme	48
3.3 Applications	50
3.3.1 Comparison of the SSFC, PAFC and VMFC(2) schemes	52
3.3.2 Approximate SSFC(R) scheme applied to water clusters	55
3.3.3 Approximate SSFC(R) scheme applied to methanol clusters	61
3.3.4 Basis set dependency on the approximate SSFC(R) scheme	62

3.3.5	SSFC(R) corrected stabilization energies of the water cluster series	65
3.3.6	Approximate VMFC(2)(R) scheme	71
3.3.7	Approximate SSFC _{inc} scheme	74
3.3.8	Approximate SSFC(S) scheme	82
3.4	Incremental evaluation of core, core-valence and valence correlation energies	111
3.4.1	Introduction	111
3.4.2	Applications	112
4	Conclusion	131
A	MP2 and CCSD incremental interaction energies	133
B	SSFC(R), VMFC(2)(R) and SSFC(S) results at the MP2 and CCSD levels of theory	143
C	Core, core-valence and valence correlation energies	155
	Literature	157
	Danksagung	165
	Erklärung gemäß §4 Abs. 9 der Promotionsordnung	167

List of Abbreviations

Å	Ångström
AO	atomic orbital
a.u.	atomic unit
avg.	average
BDE	bond dissociation energy
bf.	basis function
bfs.	basis functions
BO	Born-Oppenheimer
BSSE	basis set superposition error
ca.	approximately
CABS	complementary auxiliary basis set
cal	calorie
CBAS	auxiliary basis set
CBS	complete basis set
CC	coupled cluster (method)
CCSD(T)	coupled cluster method with singles and doubles and a perturbative estimate of triples
cf.	compare
CI	configuration interaction (method)
CP	counterpoise (correction)
CPU	central processing unit
D_e	dissociation energy from the potential minimum
DFT	density functional theory
DKH	Douglas–Kroll–Hess (Hamiltonian)
DZ	double- ζ
E	energy
e.g.	for example
ECP	effective core potential
edn.	edition
ed.; eds.	editor; editors

ϵ_i	one-electron orbital energy
eq.	equation
et al.	and others
etc.	and so forth
\hat{f}	Fock operator
g	g-factor
g_N	nuclear g-factor
GTO	Gaussian-type-orbital
\hat{H}	Hamilton operator
HF	Hartree–Fock (method)
i.e.	that is
J	Joule
\hat{J}_i	Coulomb operator
\hat{K}_i	Exchange operator
k	kilo (10^3)
LCAO	linear combination of atomic orbitals
log	logarithm to the base 10
MO	molecular orbital
m.r.e.	mean relative error
MP2	second-order Møller–Plesset perturbation theory
no.	number
PEC	potential energy curve
PES	potential energy surface
PT	perturbation theory
PP	pseudopotential
QZ	quadruple- ζ
RI	resolution of identity (approximation)
SCF	self-consistent field (method)
SD	Slater determinant
STO	Slater-type-orbital
TS	transition-state
TZ	triple- ζ
vide infra	see below
vs.	versus
ZPVE	zero-point vibrational energy
Z_I	atomic number of nucleus I

Chapter 1

Introduction

For the accurate evaluation of weak intermolecular interactions one has to apply high level quantum chemical methods such as the coupled-cluster (CC) approach or Møller-Plesset (MP) perturbation theory. The main drawback of the wavefunction-based correlation methods is the strong dependence of their demand of computational resources on the size of the one-particle basis set. For small clusters these calculations can be done routinely with commercially available quantum chemical software, however for large clusters in combination with large basis sets - which are clearly necessary for an accurate description - the application of CCSD(T) or even CCSD becomes quickly very time consuming or impossible at all. Even the calculations of error measures such as the basis set superposition error (BSSE, vide infra) becomes a difficult task for large clusters. In this work different approximations to account for the BSSE corrections for large clusters at high correlation level are introduced and compared among each other with respect to their efficiency. Furthermore the incremental scheme, which is a method from the large field of the so-called local correlation methods is applied to account for correlation energies of large water cluster and other systems.

Chapter 2

Theory

2.1 Methods of Quantum Chemistry

2.1.1 Hartree-Fock Theory

The Schrödinger equation 2.1 (here in its time-independent form):

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

is not soluble in closed form for the most systems of chemical interest. Thus it is therefore necessary to introduce approximations. The Hartree-Fock (HF) [1,2] approximation is central to computational chemistry, its outcome is exploited as the starting point for an more accurate solution to the Schrödinger equation. Within the HF theory as described below relativistic effects are neglected and the Born-Oppenheimer (BO) approximation is used, treating the nuclei as stationary sources of electrostatic fields. From the BO model follows, that the nuclear kinetic energy is neglected, correlation in the attractive electron-nuclear potential term is eliminated, and the repulsive nuclear-nuclear potential energy term \hat{h}_0 (vide infra, Eq. 2.3) becomes constant for a given geometry. Thus for the electronic Schrödinger equation:

$$\hat{H}_{el}\Psi_{el} = E\Psi_{el}, \quad (2.2)$$

the electronic Hamiltonian for a n-electron wave function can be written in terms of one-, two-, zero-electron terms as:

$$\hat{H}_{el} = \sum_i^n \hat{h}_i + \sum_{i<j}^n \hat{g}_{ij} + \hat{h}_0 \quad (2.3)$$

and reads in atomic units (a.u.) with the omission of the trivial additive constant \hat{h}_0 for fixed nuclear positions:

$$\hat{H}_{el} = - \sum_i^n \frac{1}{2} \nabla_i^2 - \sum_i^n \sum_I^M \frac{Z_I}{r_{iI}} + \sum_{i < j}^n \frac{1}{r_{ij}}. \quad (2.4)$$

In the HF theory [3] the electronic wave function, describing the ground state of an n -electron system is approximated by a single Slater determinant (SD). A SD is an antisymmetrized product of one-electron wave functions called spin orbitals φ_i , which depend upon spatial coordinates $\mathbf{r}_i = \{x_i, y_i, z_i\}$ and a spin part α or β :

$$\Psi_0 = C \hat{A} \Theta = C \hat{A} \prod_{i=1}^n \varphi_i(i). \quad (2.5)$$

The normalization constant C equals $(n!)^{\frac{1}{2}}$ when the spin orbitals are orthonormal (vide infra, Eq. 2.10), \hat{A} is the antisymmetrize operator and the one-electron wave function product Θ is called Hartree product. Eq. 2.5 satisfies two fundamental requirements of quantum mechanics. First the indistinguishability of electrons is ensured and second the wave function is antisymmetric with respect to interchange of the coordinates of any two electrons. The second statement is known as the Pauli principle, the crucial quantity in Eq. 2.5 regarding the Pauli exclusion is the antisymmetrizer:

$$\hat{A} = \sum_{k=1}^{n!} (-1)^{p_k} \hat{P}_k, \quad (2.6)$$

where the sum runs over the $n!$ possible Hartree products, \hat{P}_k is the permutation operator and $(-1)^{p_k}$ describes the parity of the k -th permutation (equals either 1 or -1, depending on whether an even or odd number of permutations will be necessary).

The variational method [4] is utilized to solve Eq. 2.2 approximately, for arbitrary functions Φ we have:

$$E[\Phi] = \frac{\langle \Phi | \hat{H}_{el} | \Phi \rangle}{\langle \Phi | \Phi \rangle} \geq \tilde{E}_0, \quad (2.7)$$

with \tilde{E}_0 being the exact ground state energy. Minimization of $E[\Phi]$ for a reasonable ansatz Φ leads to an approximate ground state energy, which is larger or in best case equal to \tilde{E}_0 . If the energy functional (i.e., the expectation value) is stationary with respect to all possible variations δ_Φ in the function Φ , then Φ is the searched solution:

$$\delta_\Phi E[\Phi] = 0 \quad (2.8)$$

Having selected a single determinant trial wave function, the variational principle can be then applied to derive the HF equations by minimizing the expectation value with

respect to the choice of spin orbitals:

$$\min : E_0[\{\varphi_i\}] \stackrel{!}{=} \langle \Psi_0 | \hat{H}_{el} | \Psi_0 \rangle \quad (2.9)$$

Choosing the spin orbitals to be orthonormal:

$$\int \varphi_i^* \varphi_j d\tau = \delta_{ij} \quad (2.10)$$

one can evaluate the matrix elements over one- and two-electron operators using the Slater-Condon rules [5]. The expectation value from Eq. 2.9 may be simplified to:

$$\begin{aligned} \langle \Psi_0 | \hat{H}_{el} | \Psi_0 \rangle &= \sum_{i=1}^n \langle \varphi_i | \hat{h} | \varphi_i \rangle + \frac{1}{2} \sum_{i,j}^n \left(\langle \varphi_i \varphi_j | \hat{g} | \varphi_i \varphi_j \rangle - \langle \varphi_i \varphi_j | \hat{g} | \varphi_j \varphi_i \rangle \right) \\ &= \sum_{i=1}^n \langle \varphi_i | \hat{h} | \varphi_i \rangle + \frac{1}{2} \sum_{i,j}^n \langle \varphi_i | \hat{J}_j - \hat{K}_j | \varphi_i \rangle \end{aligned} \quad (2.11)$$

The electron-electron interaction are represented via the Coulomb \hat{J}_j and exchange operator \hat{K}_j , both defined by their effects when operating on a spin orbital φ_i :

$$\hat{J}_j(1)\varphi_i(1) = \left[\int \frac{\varphi_j^*(2)\varphi_j(2)}{r_{12}} d\tau_2 \right] \varphi_i(1) \quad (2.12)$$

$$\hat{K}_j(1)\varphi_i(1) = \left[\int \frac{\varphi_j^*(2)\varphi_i(2)}{r_{12}} d\tau_2 \right] \varphi_j(1) \quad (2.13)$$

Note that the Coulomb interaction will always survive spin integration, whereas the exchange interaction only occurs between electrons having the same spin, the motion of electrons with parallel spins is therefore said to be correlated.¹

Applying the variational principle to the energy expression from Eq. 2.11 yields the HF equations:

$$\hat{h}(1)\varphi_i(1) + \sum_j \left[\hat{J}_j(1) - \hat{K}_j(1) \right] \varphi_i(1) = \epsilon_i \varphi_i(1), \quad (2.14)$$

where the Fock operator for electron (1):

$$\hat{f}(1) = \hat{h}(1) + v^{HF}(1) = \hat{h}(1) + \sum_j \left[\hat{J}_j(1) - \hat{K}_j(1) \right], \quad (2.15)$$

is the sum of the core-Hamiltonian operator $\hat{h}(1)$ and an effective one-electron potential operator called the HF potential v^{HF} . The HF equations 2.14 may be written as:

$$\hat{f}(1)\varphi_i(1) = \epsilon_i \varphi_i(1), \quad (2.16)$$

¹Since the one-electron operator in case of the nonrelativistic Hamiltonian does not depend on spin, spin integration also does not change the values of one-electron integrals.

which is an eigenvalue equation with the spin orbitals as eigenfunctions and the energy of the spin orbitals as eigenvalues. The complicated many-electron problem has been replaced by a one-electron problem in which the electron-electron repulsion is accounted for in an average fashion. Each electron is considered to be moving in the field of the nuclei and the average field of the other (n-1) electrons, therefore the HF method is referred to as a mean-field approximation.

Roothaan [1,2] described matrix algebra equations that permitted HF calculations to be carried out using a basis set approximation for the unknown molecular orbitals (MO). Each MO $\varphi_i(\mathbf{r})$ is expanded in terms of known basis functions $\chi_\nu(\mathbf{r})$, conventionally called atomic orbitals (MO=LCAO, Linear Combination of Atomic Orbitals):

$$\varphi_i(\mathbf{r}) = \sum_{\nu}^K c_{\nu i} \chi_{\nu}(\mathbf{r}) \quad i = 1, 2, \dots, K \quad (2.17)$$

If the set $\{\chi_\nu\}$ was complete (infinite number of basis function), this would be an exact expansion. As the basis set approaches completeness, one approaches the HF-limit (numerical HF solution), what is for practical computational reasons not reachable. Substituting the linear expansion 2.17 into the HF equation 2.16 and multiplying by χ_ν^* on the left and integrating turns the coupled integro-differential HF equations into:

$$\sum_{\nu} F_{\mu\nu} C_{\nu i} = \epsilon_i \sum_{\nu} S_{\mu\nu} C_{\nu i}, \quad (2.18)$$

which are known as Roothaan equations. The entire set of equations can be written as the single matrix equation:

$$\mathbf{FC} = \mathbf{SC}\epsilon. \quad (2.19)$$

The \mathbf{S} matrix contains the overlap of basis function, the Fock matrix \mathbf{F} is the matrix representation of the Fock operator and depends on the expansion coefficients \mathbf{C} :

$$\mathbf{F} = \mathbf{F}(\mathbf{C}). \quad (2.20)$$

The Roothaan equations are therefore nonlinear and they need to be solved iteratively. In order to find the eigenvectors \mathbf{C} and eigenvalues ϵ by diagonalizing \mathbf{F} , the generalized matrix eigenvalue Eq. 2.20 should be transformed into a conventional form. That means the overlap matrix must be unity, which is ensured when a transformed set of basis functions form an orthonormal set. One can choose a real and nonsingular transformation matrix \mathbf{X} such that $\mathbf{X}^T \mathbf{S} \mathbf{X} = \mathbf{1}$ and obtain:

$$\tilde{\mathbf{F}}\tilde{\mathbf{C}} = \tilde{\mathbf{C}}\epsilon \quad \text{with : } \tilde{\mathbf{F}} = \mathbf{X}^T \mathbf{F} \mathbf{X} \quad \text{and} \quad \tilde{\mathbf{C}} = \mathbf{X}^{-1} \mathbf{C}, \quad (2.21)$$

where the matrix \mathbf{X} defines an orthonormal basis $\{\phi_\tau\}$ expanded in the original basis $\{\chi_\nu\}$:

$$\phi_\tau(\mathbf{r}) = \sum_{\nu}^K \chi_\nu(\mathbf{r}) X_{\nu\tau} \quad (2.22)$$

Before the self-consistent field (SCF) procedure is carried out, the basis functions are orthogonalized by exemplary symmetric or canonical orthogonalization procedure [5]. During the SCF iteration one employs MO coefficients from a previous iteration as an orthonormal basis set.

2.1.2 Perturbation Theory

Another theory framework besides the variational principle used to solve the Schrödinger equation 2.1 is the so-called perturbation theory (PT) [2, 6]. The crucial characteristic within PT is that a solution of an approximate Schrödinger Eq. is known:

$$\hat{H}_0\Phi_i = E_i\Phi_i \quad i = 0, 1, 2, \dots, \infty, \quad (2.23)$$

and that this solution differs only slightly from the exact one. The exact Hamiltonian \hat{H} is divided into a reference or unperturbed Hamilton operator \hat{H}_0 and a perturbation operator \hat{H}' :

$$\hat{H} = \hat{H}_0 + \lambda\hat{H}', \quad (2.24)$$

\hat{H}_0 should closely represent the true Hamiltonian and for which the solutions form a complete set as indicated in Eq. 2.23. The perturbation operator \hat{H}' should capture only a small fraction of the true Hamiltonian, so that the perturbation becomes small, in Eq. 2.24 λ is a dimensionless parameter that, as it varies from 0 to 1, maps \hat{H}_0 into \hat{H}' . The energy and wavefunction of Eq. 2.1 Ψ are expanded in form of a Taylor series in powers of the perturbation parameter:

$$E = E^{(0)} + \lambda E^{(1)} + \lambda^2 E^{(2)} + \dots + \lambda^p E^{(p)} + \dots \quad (2.25)$$

$$\Psi = \Phi^{(0)} + \lambda\Phi^{(1)} + \lambda^2\Phi^{(2)} + \dots + \lambda^p\Phi^{(p)} + \dots, \quad (2.26)$$

where $E^{(p)}$ and $\Phi^{(p)}$ are the p-th order correction to the reference energy $E^{(0)}$ and the reference wavefunction $\Phi^{(0)}$ of the unperturbed system. Inserting Eqs. 2.24, 2.25 and 2.26 into the Schrödinger equation yields:

$$(\hat{H}_0 + \lambda\hat{H}')(\Phi^{(0)} + \lambda\Phi^{(1)} + \dots) = (E^{(0)} + \lambda E^{(1)} + \dots)(\Phi^{(0)} + \lambda\Phi^{(1)} + \dots) \quad (2.27)$$

Since all terms in 2.27 are linearly independent, we can collect terms with the same power of λ to:

$$(E^{(0)} - \hat{H}_0)\Phi^{(p)} = (\hat{H}' - E^{(1)})\Phi^{(p-1)} - \sum_{k=2}^p E^{(k)}\Phi^{(p-k)} \quad (2.28)$$

Eq. 2.28 may be further simplified if intermediate normalization (the overlap of the perturbed with unperturbed wave function is equal to unity) is supposed, which is

equivalent with:

$$\langle \Phi^{(0)} | \Phi^{(0)} \rangle = 1 \quad \langle \Phi^{(0)} | \Phi^{(p)} \rangle = 0 \quad \text{for all } p > 0 \quad (2.29)$$

and Eq. 2.28 may be simplified to:

$$E^{(p)} = \langle \Phi^{(0)} | \hat{H}' | \Phi^{(p-1)} \rangle. \quad (2.30)$$

Since the solutions 2.23 to the unperturbed system 2.28 (to the zero order) generate a complete set of functions, one can expand the unknown higher order correction to the wave function in terms of these known functions:

$$\Phi^{(1)} = \sum_i c_i \Phi_i. \quad (2.31)$$

The first-order energy correction can now be evaluated and is equal to the expectation value of the perturbation operator over the unperturbed wave function:

$$E^{(1)} = \langle \Phi^{(0)} | \hat{H}' | \Phi^{(0)} \rangle. \quad (2.32)$$

The major outcome is that the higher order correction - to the energy and wavefunction - may also be expressed in terms of matrix elements of the perturbation operator over unperturbed wave functions.

Møller and Plesset suggested using the sum over Fock operators for the unperturbed Hamiltonian:

$$\hat{H}_0 = \sum_i \hat{f}_i = \sum_i \left(\hat{h}_i + \sum_j (\hat{J}_j - \hat{K}_j) \right) = \sum_i \hat{h}_i + \sum_i \sum_j \langle \hat{g}_{ij} \rangle = \sum_i \hat{h}_i + 2\langle \hat{V}_{ee} \rangle, \quad (2.33)$$

in which the (average) electron-electron repulsion is counted twice. The perturbation operator returns the correct (electronic nonrelativistic) Hamiltonian and is therefore set to be the exact electron-electron repulsion operator \hat{V}_{ee} minus twice the $\langle \hat{V}_{ee} \rangle$ operator (which captures electron-electron repulsion as computed from summing over Fock operators):

$$\hat{H}' = \hat{H} - \hat{H}_0 = \sum_i \sum_{j>i} \hat{g}_{ij} - \sum_i \sum_j \langle \hat{g}_{ij} \rangle = \hat{V}_{ee} - 2\langle \hat{V}_{ee} \rangle. \quad (2.34)$$

Since the zeroth-order wave function is the HF determinant, the zeroth-order energy in MP theory is the sum of MO energies:

$$E_{MP}^{(0)} = \langle \Phi^{(0)} | \hat{H}_0 | \Phi^{(0)} \rangle = \langle \Phi^{(0)} | \sum_i \hat{f}_i | \Phi^{(0)} \rangle = \sum_i \epsilon_i. \quad (2.35)$$

And the first-order energy correction is the negative of the over-counted electron-electron repulsion:

$$E_{MP}^{(1)} = \langle \Phi^{(0)} | \hat{H}' | \Phi^{(0)} \rangle = \langle \hat{V}_{ee} \rangle - 2\langle \hat{V}_{ee} \rangle = -\langle \hat{V}_{ee} \rangle. \quad (2.36)$$

Thus the HF energy is the energy corrected through first-order in Møller-Plesset perturbation theory, exactly the sum of Eqs. 2.35 - 2.36. Electron correlation is therefore firstly accounted for when the MP second-order energy correction is evaluated. This involves the evaluation of all possible excited Slater determinants. But within a finite basis set approximation, the ways to distribute the electrons in the HF orbitals are also limited and hence the number of excited determinants is finite what means that the many-electron wave function is truncated. From the Condon-Slater rules follows (since our perturbation operator is a two-electron operator) that only double and single excited determinants have to be considered. Furthermore from the Brillouin's theorem follows that matrix elements between the closed-shell HF determinant and the singly excited ones are zero. Second-order MP energy correction therefore only involves a sum over doubly excited determinants:

$$E_{MP}^{(2)} = \sum_{i < j}^{occ} \sum_{a < b}^{vir} \frac{\langle \Phi^{(0)} | \hat{H}' | \Phi_{ij}^{ab} \rangle \langle \Phi_{ij}^{ab} | \hat{H}' | \Phi^{(0)} \rangle}{E_0 - E_{ij}^{ab}}, \quad (2.37)$$

and matrix elements between the HF and a doubly excited state are given by two-electron integrals over MOs:

$$E_{MP}^{(2)} = \sum_{i < j}^{occ} \sum_{a < b}^{vir} \frac{(\langle \phi_i \phi_j | \phi_a \phi_b \rangle - \langle \phi_i \phi_j | \phi_b \phi_a \rangle)}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b}, \quad (2.38)$$

where in the denominator the energy difference between two Slater determinants occur, these quantities correspond to differences in MO energies.

2.1.3 Coupled-Cluster Theory

In Coupled-Cluster theory [7] the many-electron wave function is constructed through an exponential ansatz of the cluster operator:

$$\Psi = e^{\hat{T}} \Psi_{HF} \quad (2.39)$$

The cluster operator is defined as:

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \dots + \hat{T}_n \quad (2.40)$$

where n is the number of electrons and the \hat{T}_i operators construct all possible determinants having i excitations from the reference one, as exemplary shown for \hat{T}_2 :

$$\hat{T}_2 \Psi_{HF} = \sum_{i < j}^{occ} \sum_{a < b}^{vir} t_{ij}^{ab} \Psi_{ij}^{ab}, \quad (2.41)$$

where i, j are the occupied MOs in the HF reference wave function and a, b are virtual MOs in Ψ_{HF} . The excited determinants are obtained by exciting an electron from occupied orbital(s) indicated by subscripts into the virtual orbitals indicated by superscripts. The amplitudes t are determined by the constraint that Eq. 2.41 be satisfied. One of the most appealing features of the CC theory (in contrast to a truncated Configuration Interaction expansion) is that it is size-consistent. Size consistency means that the energy of an A-B system, where A and B are at infinite separation is equal to the sum of the energies of A and B calculated individually. To illustrate this we consider a truncated CC expansion with the usage of only the double excitation operator, as indicated by the superscript CCD:

$$\Psi_{CCD} = e^{\hat{T}_2} \Psi_{HF} = \left(1 + \hat{T}_2 + \frac{\hat{T}_2^2}{2!} + \frac{\hat{T}_2^3}{3!} + \dots \right) \Psi_{HF}, \quad (2.42)$$

where the exponential function is expanded as a Taylor series. In Eq. 2.42 the \hat{T}_2 generates double excited determinants, the square of \hat{T}_2 quadruple excitations, the cube \hat{T}_2^3 sextuple substitution and so on. The Taylor expansion is finite in practice due to the finite number of occupied MOs and therefore a limited number of excitations. The inclusion of these higher order excitations ensures the method to be size-consistent. The CC Schrödinger equation reads:

$$\hat{H} e^{\hat{T}} |\Psi_{HF}\rangle = E e^{\hat{T}} |\Psi_{HF}\rangle \quad (2.43)$$

and it is solved by projecting the Schrödinger Eq. 2.43, which means left-multiplying by a trial wave functions expressed as determinants of the HF orbitals. This generates a set of coupled, nonlinear equations in the amplitudes which have to be solved. With these solution the CC energy may be calculated according to:

$$\langle \Psi_{HF} | \hat{H} | e^{\hat{T}} \Psi_{HF} \rangle = E_{CC} \quad (2.44)$$

Dependent on which contributions from the cluster operator are included in the exponential ansatz the CC energy is referred to CCS (only single excited determinants are considered), CCSD (additionally the double excited are included) and so on.

2.1.4 Explicitly correlated methods

Electron correlation treatment based on the finite one-particle basis function expansion of the molecular wave function suffers from the frustratingly slow convergence with respect to the size of the latter basis towards the complete basis set (CBS) limit. It has been recognized that the slow convergence is due to a poor description of the so-called correlation cusp (vide infra) when standard quantum chemical methods like CC or MP theory are applied. One can bypass the slow convergence considering a wave function that explicitly depend on the interelectronic distance, methods which incorporate these dependence are therefore called explicitly correlated ones.

The discussion on the origin of these methods is given here for the ground state of the helium atom with one electron fixed at a separation of $0.5 a_0$ from the nucleus [3]. Within the HF theory the mean-field approximation causes that the motion of one electron is unaffected by the instantaneous position of the second, meaning that the wave function amplitude for one electron depends only on its distance to the nucleus but not on the distance to the other electron. For the exact wave function the probability amplitude for one electron is affected by the second fixed electron creating the so-called Coulomb hole around this electron. The Coulomb hole is a classically forbidden region and a good description is necessarily for an accurate treatment of the so-called short range dynamical correlation. The non-relativistic electronic helium Hamiltonian with the origin at the nucleus (where \mathbf{r}_1 and \mathbf{r}_2 are the coordinates of the two electrons and Z is the nuclear charge):

$$\hat{H} = -\frac{1}{2} \sum_{i=1}^2 \nabla_i^2 - \frac{Z}{|\mathbf{r}_1|} - \frac{Z}{|\mathbf{r}_2|} + \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|}, \quad (2.45)$$

has singularities for $\mathbf{r}_1=0$, $\mathbf{r}_2=0$ and $\mathbf{r}_1=\mathbf{r}_2$. At the electron-electron and electron-nucleus coalescence points, the exact solutions to the Schrödinger equation:

$$\hat{H}\Psi(\mathbf{r}_1, \mathbf{r}_2) = E\Psi(\mathbf{r}_1, \mathbf{r}_2) \quad (2.46)$$

must provide contributions to the product $\hat{H}\Psi$ that balance the singularities in 2.45 such that the local energy:

$$\epsilon(\mathbf{r}_1, \mathbf{r}_2) = \frac{\hat{H}\Psi(\mathbf{r}_1, \mathbf{r}_2)}{\Psi(\mathbf{r}_1, \mathbf{r}_2)} \quad (2.47)$$

remains constant and equal to the eigenvalue E . In an exact wave function these singularities must be exactly canceled by the kinetic energy operator. It is convenient to employ the symmetry of the wave function and to express the helium Hamiltonian 2.45 in terms of three radial coordinates r_1 , r_2 and r_{12} (with r_{12} being the interelectronic

distance and r_1, r_2 the distances of the electrons to the nucleus):

$$\begin{aligned} \hat{H} = & -\frac{1}{2} \sum_{i=1}^2 \left(\frac{\partial^2}{\partial r_i^2} + \frac{2}{r_i} \frac{\partial}{\partial r_i} + \frac{2Z}{r_i} \right) - \left(\frac{\partial^2}{\partial r_{12}^2} + \frac{2}{r_{12}} \frac{\partial}{\partial r_{12}} - \frac{1}{r_{12}} \right) \\ & - \left(\frac{\mathbf{r}_1 \cdot \mathbf{r}_{12}}{r_1 r_{12}} \frac{\partial}{\partial r_1} + \frac{\mathbf{r}_2 \cdot \mathbf{r}_{21}}{r_2 r_{21}} \frac{\partial}{\partial r_2} \right) \frac{\partial}{\partial r_{12}}. \end{aligned} \quad (2.48)$$

The singularities at the nucleus are balanced by the kinetic energy terms proportional to $1/r_i$ since:

$$\left(\frac{2}{r_i} \frac{\partial}{\partial r_i} + \frac{2Z}{r_i} \right) \Psi \Big|_{r_i=0} = 0 \quad \Rightarrow \quad \frac{\partial \Psi}{\partial r_i} \Big|_{r_i=0} = -Z\Psi \quad (r_i = 0) \quad (2.49)$$

where $(\partial\Psi/\partial r_i) = -Z\Psi$ at $r_i = 0$ is known as the nuclear cusp condition which can be easily satisfied with the usage of Slater-type-orbitals (STOs). Likewise the terms that multiply $1/r_{12}$ at $r_{12} = 0$ must vanish in $\hat{H}\Psi$:

$$\left(\frac{2}{r_{12}} \frac{\partial}{\partial r_{12}} + \frac{1}{r_{12}} \right) \Psi \Big|_{r_{12}=0} = 0 \quad (2.50)$$

imposing the additional condition:

$$\frac{\partial \Psi}{\partial r_{12}} \Big|_{r_{12}=0} = \frac{1}{2} \Psi \quad (r_{12} = 0) \quad (2.51)$$

which is known as the Coulomb cusp condition. The Coulomb cusp condition is not satisfied for the HF wave function:

$$\frac{\partial \Psi^{HF}}{\partial r_{12}} \Big|_{r_{12}=0} = 0, \quad (2.52)$$

since the Slater determinant does not depend on the interelectronic distance as electrons approach one another. But within the single-determinant level the so-called Fermi correlation is included which occur as a consequence of the Pauli antisymmetry. In the Configuration Interaction (CI) approach the helium ground full CI wave function constructed from STOs becomes:

$$\Psi^{CI} = \exp[-\zeta(r_1 + r_2)] \sum_{ijk} C_{ijk} (r_1^i r_2^j + r_1^j r_2^i) r_{12}^{2k} \quad (2.53)$$

And since only even powers of r_{12} are included the cusp condition cannot be satisfied:

$$\frac{\partial \Psi^{CI}}{\partial r_{12}} \Big|_{r_{12}=0} = 0. \quad (2.54)$$

However when a linear term in r_{12} is included:

$$\Psi_{r_{12}}^{CI} = \left(1 + \frac{1}{2}r_{12}\right)\Psi^{CI} \quad (2.55)$$

the cusp condition is exactly satisfied since:

$$\left.\frac{\partial\Psi_{r_{12}}^{CI}}{\partial r_{12}}\right|_{r_{12}=0} = \frac{1}{2}\Psi^{CI}(r_{12}=0) = \frac{1}{2}\Psi_{r_{12}}^{CI}(r_{12}=0) \quad (2.56)$$

The cusp condition can therefore always be satisfied by multiplication with a correlating function γ :

$$\gamma = 1 + \frac{1}{2}\sum_{i>j} r_{ij}, \quad (2.57)$$

which leads to the correct nondifferentiable cusp in the product function $\gamma\Psi$. Methods that employ correlating functions or otherwise make use of the interelectronic distances r_{ij} are called explicitly correlated methods. A distinction is drawn between R12 method which includes r_{ij} linearly and the F12 method which includes a more general (exponential) dependence on r_{ij} .

2.2 Incremental Scheme

The wide branch of the local correlation methods [8–24] and also the fragment-based methods [25–38] aim at a substantial reduction of computational requirements for medium-sized and large systems while maintaining the high accuracy of wavefunction-based *ab initio* approaches. Among the fragment-based local correlation methods, the incremental scheme devised by Stoll for finite-cluster calculations modeling 3D crystals [39–41], related to earlier ideas of Nesbet for atoms [42], is quite unique due to its wide range of applicability. It allows both wavefunction-based correlated electronic structure calculations of periodic systems using Wannier-type orbitals [43, 44] and of medium-sized and large molecules using localized molecular orbitals [45]. The adsorption of molecules on crystalline surfaces can also be studied [46, 47]. In its simplest form the approach can be combined with any size-extensive correlation treatment provided by standard quantum chemical program packages without changing the correlation modules and thus extends their range of applicability beyond the one of the standard wavefunction-based correlation methods.

The incremental procedure starts with the localization of the canonical HF orbitals. The set of localized Hartree-Fock orbitals is then grouped into disjoint subsets, the so-called one-site domains, which form the set \mathbb{D} . The set of n-site domains is obtained by adding all pairs, triples etc. to the one-site domains. The resulting outcome is the

power set of the set of one-site domains $\mathcal{P}(\mathbb{D})$. Correlation energy calculations are carried out for $\mathcal{P}(\mathbb{D})$. The incremental correlation energy is expand as:

$$E_{\text{corr}} = \sum_{\mathbb{X} \in \mathcal{P}(\mathbb{D}) \wedge |\mathbb{X}| \leq \mathcal{O}} \Delta \varepsilon_{\mathbb{X}}. \quad (2.58)$$

The summation in Eq. 2.58 includes all terms of $\mathcal{P}(\mathbb{D})$ with a cardinality less or equal to the order of the expansion, denoted as \mathcal{O} . The general correlation energy increment $\Delta \varepsilon_{\mathbb{X}}$ is defined as:

$$\Delta \varepsilon_{\mathbb{X}} = \varepsilon_{\mathbb{X}} - \sum_{\mathbb{Y} \in \mathcal{P}(\mathbb{X}) \wedge |\mathbb{Y}| < |\mathbb{X}|} \Delta \varepsilon_{\mathbb{Y}}, \quad (2.59)$$

where \mathbb{X} and \mathbb{Y} are the summation indices defined in Eq. 2.58 and Eq. 2.59 respectively. Here $\varepsilon_{\mathbb{X}}$ stands for the correlation energy obtained upon correlating all electrons in \mathbb{X} . The expansion described via Eq. 2.58 is exact when carried out to the highest expansion order but advantages with respect to computational requirements arise when it is truncated at preferable low order. For example, a third-order expansion which appear to be accurate enough (chemical accuracy with errors below 1 kcal/mol) for a large variety of systems [48] reads:

$$E_{\text{corr}} = \sum_i \Delta \varepsilon_i + \sum_{i < j} \Delta \varepsilon_{ij} + \sum_{i < j < k} \Delta \varepsilon_{ijk}, \quad (2.60)$$

with the one-, two- and three-body increments:

$$\begin{aligned} \Delta \varepsilon_i &= \varepsilon_i \\ \Delta \varepsilon_{ij} &= \varepsilon_{ij} - \Delta \varepsilon_i - \Delta \varepsilon_j \\ \Delta \varepsilon_{ijk} &= \varepsilon_{ijk} - \Delta \varepsilon_{ij} - \Delta \varepsilon_{ik} - \Delta \varepsilon_{jk} - \varepsilon_i - \varepsilon_j - \varepsilon_k. \end{aligned} \quad (2.61)$$

Provided that Eq. 2.58 may be truncated at a low expansion order it further becomes more computationally attractive when small incremental contribution (small with respect to the energy they contribute to the overall expansion sum) are identified and neglected *a priori* and especially when the virtual space of the domains is reduced. Low-order truncation of the incremental expansion and an efficient screening method requires domains whose orbitals are spatially close to each other, but remote to the orbitals of other domains. Since for larger systems it is quite tedious and also error prone to set up the domains by inspection, a fully automated domain generation is crucial. Friedrich et al. [49, 50] proposed a procedure to set up the domains automatically and generated a computer code which also provides all needed input data to perform all incremental correlation energy calculations to the highest order with external quantum chemical codes.

2.2.0.1 Construction of the valence domains

The centers of charge for the set of occupied valence orbitals \mathbb{O} are obtained from the diagonal elements of the dipole integrals in MO basis:

$$\phi_a \mapsto \vec{R}_a := \begin{pmatrix} \langle \phi_a | x | \phi_a \rangle \\ \langle \phi_a | y | \phi_a \rangle \\ \langle \phi_a | z | \phi_a \rangle \end{pmatrix} = \begin{pmatrix} x_a \\ y_a \\ z_a \end{pmatrix}, \quad (2.62)$$

therefore the set of \mathbb{O} is mapped to a set of vectors in \mathbb{R}^3 :

$$\mathbb{O} \rightarrow \mathbb{R}^3. \quad (2.63)$$

The distances $D_{ab} = |\vec{R}_a - \vec{R}_b|$ between the centers of charge are used to construct an edge-weighted graph. In order to arrive at disjoint sets of orbitals forming the domains [50] a graph partitioning problem has to be solved. For this purpose the Metis graph partitioning library [51] is used. The distances $D_{ab} = |\vec{R}_a - \vec{R}_b|$ of all pairs of centers of charge define the distance matrix D , from which a connectivity matrix C is constructed:

$$C_{ab} = \begin{cases} 10^8, & \text{if } D_{ab} \leq t_{con} \wedge \frac{q}{D_{ab}} \geq 10^8 \\ \frac{q}{D_{ab}}, & \text{if } D_{ab} \leq t_{con} \wedge \frac{q}{D_{ab}} < 10^8 \\ 0, & \text{if } D_{ab} > t_{con} \end{cases} \quad (2.64)$$

Here t_{con} is a distance threshold and q is a constant stretching factor set to 10^4 , the factor of 10^8 enters as an approximation of infinity in the representation of integers on a computer. The connectivity matrix conditions enforce the construction of an edge-weighted graph, where the orbital pairs (represented via center of charge) with short distances get a large weight, and those with a large distance get a small or a zero weight. Metis partitions the graph under the side condition that the sum of weights of the cut edges is a minimum. Hereby the number of resulting valence domains can be controlled by specifying the so-called domain-size parameter (dsp) according to:

$$\text{no. valence domains} = \frac{\text{no. valence orbitals}}{\text{dsp}}. \quad (2.65)$$

It should be noted that the choice of the domain size also influences the order of the incremental expansion required for a desired accuracy as well as the computational effort. Formally any choice between single-orbital domains and the treatment of the whole system as a single domain, corresponding to the conventional calculation, is possible. For a given target accuracy smaller domain sizes require a truncation at higher order than larger domain sizes, i.e. they lead to a higher number of correlation calculations. On the other hand, choosing too large domains may deteriorate the efficiency despite a possible truncation at low order, since e.g. contributions for groups

of orbitals are implicitly included and have to be evaluated, although from a numerical point of view they could be neglected. Thus for n occupied valence orbitals the optimum domain size is $1 < \text{dsp} \ll n$ and often can be chosen according to the physical situation, e.g. $\text{dsp}=4$ is a suitable choice when treating water clusters, since a water molecule has 4 valence orbitals.

2.2.0.2 Screening procedure

A sufficient screening of small incremental contributions exploits the property that the incremental values decay with increasing order and for a given order with increasing distances between the underlying domains. An order-dependent distance truncation procedure according to:

$$t_{dist} = \frac{f}{(i-1)^2} \quad \text{with:} \quad i \geq 2, \quad (2.66)$$

where f is a variable parameter, excludes those increments for which all distances between the centers of charge of two groups of underlying domains is larger than t_{dist} .

2.2.0.3 Domain-specific basis set

The usage of a domain-specific basis set [52] within an incremental calculation provides a significant speed up in calculation time since the virtual space is reduced as well as the number of the required integrals. Within this procedure the centers of charge of the localized orbitals in a given domain are utilized to detect all atom coordinates which will be treated with the original large AO basis. The selection is controlled with a variable distance parameter t_{main} and all atom centers not covered by this threshold form the environment of the domains and are treated with the second smaller basis. Once the domain-specific basis set is constructed a second HF calculation followed by a localization is performed. With a remapping of the center of charge from the first to those from the second localization the occupied orbitals from a given domain are identified and the correlation calculation in the reduced basis set can be performed. Since the remapping may not always be unique a template localization has to be used [53,54].

2.2.0.4 Incremental MP2 and CCSD(T) correlation energy contributions

In order to converge to the canonical MP2 or CCSD(T) energies one has to take into account that neither the canonical treatment of MP2 nor the canonical treatment of the perturbative triples correction in CCSD(T) is invariant with respect to a unitary transformation of the occupied orbitals. This problem can be circumvented with a pseudo-canonical MO basis. For this purpose the Fock matrix in the local MO basis

is diagonalized in the subspace of the domain and the corresponding transformation is used to transform the local MOs to the pseudo-canonical MOs [55].

2.2.0.5 Evaluation of core and core-valence correlation energy contributions

Core correlation contributions are often neglected in studies of larger systems with many cores. The reason is mostly the computational effort, not the insignificance of the neglected effects. In order to treat the core correlation effects efficient, i.e. core-core and core-valence correlation, within the incremental framework disjoint sets of localized core and valence orbitals are required [56]. Therefore the core and valence orbitals are localized separately. The set \mathbb{D} of one-site domains is split into the set of one-site core domains \mathbb{D}_c and the set of one-site valence domains \mathbb{D}_v . With this classification Eq. 2.58 may be rewritten in terms of three sums, which account for the energy contributions arising from core-core $\Delta\varepsilon_{\mathbb{X}}$, core-valence $\Delta\varepsilon_{\mathbb{Y}}$ and valence-valence $\Delta\varepsilon_{\mathbb{Z}}$ correlation effects separately.

$$\begin{aligned}
 E_{\text{corr}} &= \sum_{\mathbb{X}} \Delta\varepsilon_{\mathbb{X}} + \sum_{\mathbb{Y}} \Delta\varepsilon_{\mathbb{Y}} + \sum_{\mathbb{Z}} \Delta\varepsilon_{\mathbb{Z}} \\
 &\mathbb{X} \in \mathcal{P}(\mathbb{D}_c) \wedge |\mathbb{X}| \leq \mathcal{O} \\
 &\mathbb{Y} \in \mathcal{P}(\mathbb{D}) \setminus [\mathcal{P}(\mathbb{D}_c) \cup \mathcal{P}(\mathbb{D}_v)] \wedge |\mathbb{Y}| \leq \mathcal{O} \\
 &\mathbb{Z} \in \mathcal{P}(\mathbb{D}_v) \wedge |\mathbb{Z}| \leq \mathcal{O}
 \end{aligned}
 \tag{2.67}$$

If the expansions in Eq. 2.67 are carried out to the highest possible order and no further approximations with respect to the local orbital character are made the incremental correlation energy E_{corr} corresponds exactly to the total correlation energy obtained with standard methods. Treating the core correlation on equal footing with valence correlation may become quite expensive especially when several larger cores are present. However, from the physical point of view such a uniform treatment is not really necessary, since core shells are usually quite compact and rather tightly bound. Previous incremental CCSD investigations showed that for not too diffuse cores neglecting inter-core correlation contributions is a significant simplification leading only to negligible errors [57]. Thus, each core can be treated individually by evaluating its intra-core correlation energy directly and applying an incremental expansion for its core-valence correlation energy contributions. Since after localization the core orbitals are more compact than the valence orbitals the order of the incremental core-valence correlation energy expansion has to be at most the order used for evaluating the valence correlation energy.

According to these considerations Eq. 2.67 can be rewritten as:

$$\begin{aligned}
 E_{\text{corr}} &= \sum_{\mathbb{X}} \Delta\varepsilon_{\mathbb{X}} + \sum_{\mathbb{Y}} \Delta\varepsilon_{\mathbb{Y}} + \sum_{\mathbb{Z}} \Delta\varepsilon_{\mathbb{Z}} \\
 \mathbb{X} &\in \mathcal{P}(\mathbb{D}_c) \wedge |\mathbb{X}| = 1 \\
 \mathbb{Y} &\in \{Y \in \mathcal{P}(\mathbb{D}) \mid |Y \cap \mathbb{D}_c| = 1 \wedge 0 < |Y \cap \mathbb{D}_v| \leq (\mathcal{O} - 1)\} \\
 \mathbb{Z} &\in \mathcal{P}(\mathbb{D}_v) \wedge |\mathbb{Z}| \leq \mathcal{O}
 \end{aligned} \tag{2.68}$$

The partitioning of the valence orbitals has been already described above. The core domains are constructed by mapping centers of charge of the localized core orbitals to the closest atom coordinates. Therefore as many core domains occur as many atoms with core orbitals are considered in a given calculation. Within this simple procedure comparatively small local domains are obtained due to the local character of the core orbitals. This procedure would not work sufficiently for the valence domains since a center of charge may be located in the middle of two atoms and therefore a unique mapping would not be possible.

2.2.0.6 Scaling behavior of the incremental scheme

The formal scaling behavior of the incremental expansion Eq. 2.67 depends on the number of individual calculations and the time needed to perform them. The total amount of individual calculations is equal to:

$$\mathcal{N}_{\text{calc}}^{\text{total}} = \sum_{i=1}^{\mathcal{O}} \binom{|\mathbb{D}|}{i}, \tag{2.69}$$

where the expansion order \mathcal{O} is equal to the number of the domains \mathbb{D} in the limiting case when Eq. 2.58 is carried out to highest possible expansion order. Eq. 2.69 is also applicable to the core-valence treatment without simplifications as described in Eq. 2.67 when \mathbb{D} is the unified set of core \mathbb{D}_c and valence \mathbb{D}_v domains. In the approximate treatment of the core correlation, incremental energy calculations are excluded *a priori* as describe via Eq. 2.68 and hence a reduced total number of individual calculations is considered according to:

$$\mathcal{N}_{\text{calc}}^{\text{core-val}} = \sum_{i=1}^{\mathcal{O}} \binom{|\mathbb{D}_v|}{i} + |\mathbb{D}_c| + \sum_{i=2}^{\mathcal{O}} \binom{|\mathbb{D}_v|}{(i-1)} |\mathbb{D}_c|. \tag{2.70}$$

Incremental energy calculation carried out in this work make use of of an implementation of the incremental scheme which contains an interface to the MOLPRO quantum chemistry package [58]. The localization is performed using the Boys [59] functional

with the algorithm of Edmiston and Ruedenberg [60] separately for the core orbitals and for the valence orbitals. The thresholds needed to specify the input data for the incremental calculations are listed at the bottom of all tables presenting the results. In order to reduce the error propagation arising from the recursive nature of the incremental expansion an energy convergence threshold E_{thres} was applied for the correlation calculations entering at highest order, whereas for the lower orders the thresholds are tightened dynamically [61].

2.3 Basis Set Superposition Error

The basis set superposition error (BSSE, *vide infra*) was reported for the first time by Kestner [62] in 1968, while trying to explain the too deep minimum on the potential energy curve for the helium dimer. One year later Jansen and Ros [63] detected the BSSE error when investigating the protonation of carbon monoxide. However the error was for the first time termed by Liu and McLean [64], who also investigated the helium-helium interaction.

The BSSE occurs in every molecular electronic structure calculation whenever orbitals are approximated by an expansion in terms of analytic basis functions (most commonly ones used are Gaussians) [65]. That is, at the HF level of theory as well as at the correlated level of theory when wavefunction based methods like CC or perturbation theory are employed. But BSSE also exist for approximate Hamiltonians such as semiempirical forms or density functional methods [66], it is also not negligible for Slater-type functions [66]. We may clearly identify the appearance of BSSE as a consequence of the usage of a truncated basis set as there is no doubt in the theoretical chemistry literature that the BSSE is completely eliminated in the limit of a complete basis set.

BSSE is primarily related to the calculation of interaction energies within the supermolecule approach, which is widely used since it only requires as a prerequisite the applied method to be size extensive, but unfortunately it suffers from the BSSE effect. Within the supermolecule approach the interaction energy for exemplary a dimer is evaluated by subtracting the energies of the monomers from the one of the dimer. However, in this prescription the BSSE arises from the more significant incompleteness of the basis sets used for the monomers than for the one of the dimer. For a given incomplete basis set the wavefunction of a supermolecule is more flexible in the sense that the energy of the individual monomers within this complex is artificially lowered due to the partial use of the basis functions centered on the other monomer. This causes an energy lowering and therefore too deep minima at too short distances on the potential energy surface (PES).

The magnitude of the BSSE is influenced by three fundamental issues. The first one is the investigated type of system. Considering the interaction energies the BSSE fraction with respect to the interaction energy depends upon the strength of the molecular interaction. The weaker the interaction energy is all the more the calculated interaction energies are affected by the BSSE effect, which may even be of the order of the interaction energy itself. Therefore it is mandatory to consider the BSSE whenever the nature of interaction is due to dispersion or electrostatic forces or when investigations are carried out on hydrogen-bonded systems. BSSE effects become less important for chemically interacting systems which represent the strongest molecular interaction. The level of theory is another factor affecting the magnitude of BSSE. The BSSE

impact is known to be larger for post-Hartree-Fock methods than at the Hartree-Fock level. However, by far the greatest effect on the size of the BSSE error has the quality of the applied basis set, what is reasonable as the usage of a finite basis set is the reason why the BSSE occurs. Since in theory the BSSE vanishes at the CBS limit, the closer the energy converges to the later, the smaller the BSSE will be. Especially the so-called correlation consistent basis sets introduced by Dunning and co-workers [67–69] provide a well defined path to the CBS limit and hence these families of basis sets provide an extremely powerful approximation in any quantum chemical application which utilizes them.

Quite recently it has been recognized that the BSSE not only arises when computational chemistry describes the interaction of two or more species. The importance of the so-called intramolecular BSSE has been understood recently [70–74]. Exemplary it was reported that intramolecular BSSE is responsible for masking the expected minima in a conformation equilibrium analysis for a dipeptide on the PES [75] or that due to intramolecular BSSE *ab initio* calculations predict wrongly benzene and arene molecules to exhibit nonplanar minima [76,77].

The strategy of how to deal with the BSSE can be roughly divided into two categories. One methodology aims at the omission of BSSE from the theory model [78–88] as exemplary realized within the symmetry-adapted perturbation theory (SAPT) [89–92], where the interaction energies are evaluated directly as a sum of physically distinct contributions. Within the other methodology one corrects the BSSE in an *a posteriori* fashion, by far the most widely used *a posteriori* prescription is the counterpoise (CP) correction introduced by Boys and Bernardi [93]. Within this correction method the energy calculations for the individual monomers are performed using the whole supermolecular basis sets instead of only the appropriate monomer basis sets. A literature survey indicates [94] that from the year of publication in 1970 the Boys and Bernardi paper was 9416 times quoted, which is an enormously amount of citations and clearly evidences the popularity of the CP scheme.

Chapter 3

Results and Discussion

3.1 Incremental evaluation of interaction energies

3.1.1 Introduction

Recently Wang and Paulus [95] proposed to calculate the binding energy of an intermolecular system, i.e. a H₂S-benzene complex, without referring to monomer calculations in the full dimer basis set. Instead they proposed to estimate the binding energy by performing three correlated calculations on the dimer in a local orbital basis: calculations correlating only the occupied orbitals of one of the monomers in the dimer, hereby treating the occupied orbitals localized on the other monomer as frozen core, and a calculation correlating simultaneously the corresponding occupied orbitals of both monomers in the dimer. The energy difference between the dimer correlation energy and the two monomer-in-dimer correlation energies was taken as an estimate for the binding energy. Using this simple prescription Wang and Paulus obtained 99% of the CCSD(T) binding energy of the H₂S-benzene dimer, when only the valence orbitals on H₂S and the π -orbitals on benzene were included in the correlation treatment. The authors emphasized that the evaluated binding energy is BSSE-free. Furthermore, since the correlation energy difference used as an estimate for the binding energy corresponds to a so-called two-body increment in the incremental expansion of the correlation energy as proposed by Stoll [39], they use the name method of increments for their approach.

The prescription of Wang and Paulus is applied here to calculate the interaction energy of the helium, hydrogen sulfide, methane and water dimer and the accuracy of that approach is compared to results obtained with standard methods.

3.1.1.1 Methodology and computational details

The study is based on the assumption of weak interactions, for which the geometries of the interacting monomers are essentially identical to the ones of the free monomers. In this case the interaction energy ΔE between two monomers is given by the difference of the energy of the dimer E_{ij}^{ij} and the ones of the monomers in the monomer basis E_i^i , E_j^j using their geometries in the dimer:

$$\Delta E = E_{ij}^{ij} - E_i^i - E_j^j \quad (3.1)$$

In Eq. 3.1 and onward the following notation is used: the basis sets are noted as superscript, whereas the considered system is noted as subscript. In the case of a dimer aggregate the most widely used prescription to correct for the BSSE is the full counterpoise (CP) correction of Boys and Bernardi [93]. Within the CP scheme all calculations, i.e. those for the dimer ij as well as those for the monomers i and j are carried out within the dimer basis set:

$$\Delta E^{CP} = E_{ij}^{ij} - E_i^{ij} - E_j^{ij} \quad (3.2)$$

where the E_i^{ij} and E_j^{ij} energies are again evaluated at the geometries taken from the dimer. Therefore for both Eq. 3.1 and Eq. 3.2 the relaxation energy contribution is neglected.

In order to obtain values for the basis set limit of the total energy as a benchmark, the two-point X^{-3} extrapolation of Halkier et al. [96] to the complete basis set (CBS) limit for the correlation energy has been used. To distinguish more easily between total energies and correlation energies the symbols E and ε are used for them, respectively.

$$E_{XY} = E_{HF,Y}^{CP} + \frac{X^3\varepsilon_X - Y^3\varepsilon_Y}{X^3 - Y^3} \quad \text{with} \quad X < Y \quad (3.3)$$

Here X, Y denote the cardinal numbers of the applied basis sets and $E_{HF,Y}^{CP}$ stands for the CP-corrected Hartree-Fock (HF) energy obtained in the larger basis set with cardinal number Y .

According to the introduced convention the notation for the correlation energy contribution $\Delta\varepsilon$ to the CP-uncorrected interaction energy of Eq. 3.1 is:

$$\Delta\varepsilon = \varepsilon_{ij}^{ij} - \varepsilon_i^i - \varepsilon_j^j \quad (3.4)$$

and for the corresponding CP-corrected interaction energy of Eq. 3.2:

$$\Delta\varepsilon^{CP} = \varepsilon_{ij}^{ij} - \varepsilon_i^{ij} - \varepsilon_j^{ij} \quad (3.5)$$

If one assumes a pure dispersion interaction one can calculate the interaction energy in a local orbital basis as a difference between correlation energies for the dimer and the monomers in the dimer, i.e.

$$\Delta\varepsilon_{MI} = \varepsilon_{ij}^{ij} - \varepsilon_i^{\bar{i}j} - \varepsilon_j^{\bar{i}j} \quad (3.6)$$

The approximation $\Delta E = \Delta\varepsilon_{MI}$ has been advocated by Wang and Paulus for the direct evaluation of the binding energy in the H₂S-benzene complex at the CCSD(T) level. Direct means that no calculations for separated monomers are needed and thus also no BSSE-corrections have to be taken into account. Since Eq. 3.6 corresponds to the definition of a two-body contribution in the incremental expansion of the correlation energy the authors used the term method of increments (MI) for their procedure. Note that Eq. 3.6 equals the second line of Eq. 2.61, however the notation here is slightly different as for Eq. 2.61 in order to be able distinguish clearly between Eq. 3.4, Eq. 3.5 and Eq. 3.6.

In general the HF contribution to the interaction energy, which contains e.g. the Pauli repulsion between the monomers, is not negligible and has to be added to the incremental correlation contribution $\Delta\varepsilon_{MI}$ in order to obtain reliable estimates. In this investigation the HF interaction energy ΔE_{HF} or the corresponding CP-corrected quantity ΔE_{HF}^{CP} , evaluated with Eq. 3.2, has been used and therefore the approximate interaction energy for the method of increments is evaluated according to:

$$\Delta E_{MI} = \Delta E_{HF} + \Delta\varepsilon_{MI}, \quad (3.7)$$

and in the CP-corrected case as:

$$\Delta E_{MI}^{CP} = \Delta E_{HF}^{CP} + \Delta\varepsilon_{MI}. \quad (3.8)$$

The calculation of the interaction energy for the H₂S-benzene complex at the CCSD(T)/aug-cc-pVDZ level of theory (compare Figs. 3.2 and 3.1) indicates, as for the other systems treated here, that $\Delta\varepsilon_{MI}$ alone does not exhibit a minimum near the equilibrium geometry.

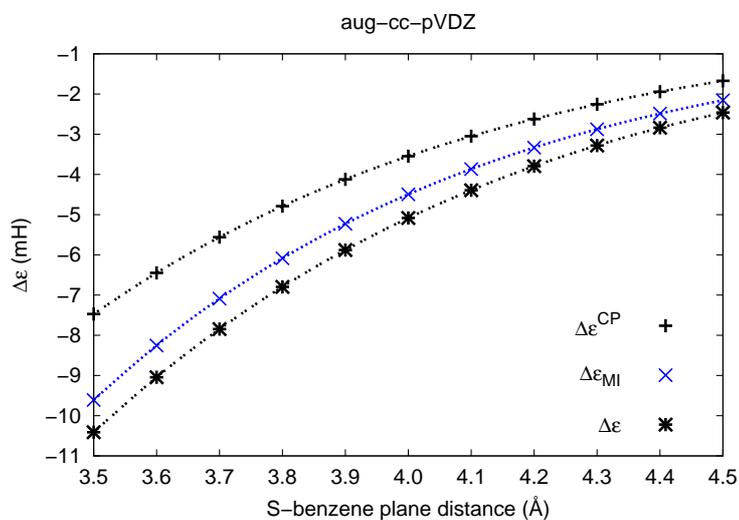


Figure 3.1: Correlation energy contribution to the interaction energy of the $\text{C}_6\text{H}_6\text{-H}_2\text{S}$ complex calculated with aug-cc-pVDZ basis set at the CCSD(T) level of theory as a function of the intermolecular sulfur atom - benzene plane distance.

Thus the potential energy curve on the H_2S -benzene system of Ref. [95] cannot be solely based on Eq. 3.6 - it must contain HF contributions to the interaction energy.

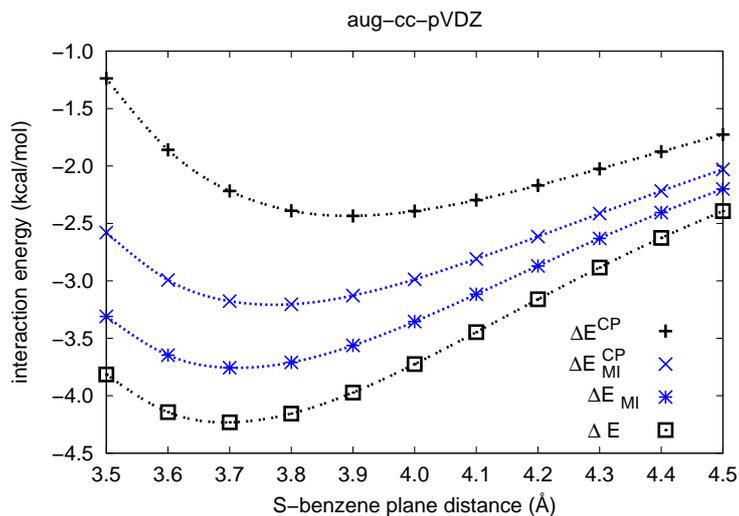


Figure 3.2: Potential energy curves of the interaction energy for the $\text{C}_6\text{H}_6\text{-H}_2\text{S}$ complex, calculated with aug-cc-pVDZ basis set at the CCSD(T) level of theory as a function of the intermolecular sulfur atom - benzene plane distance.

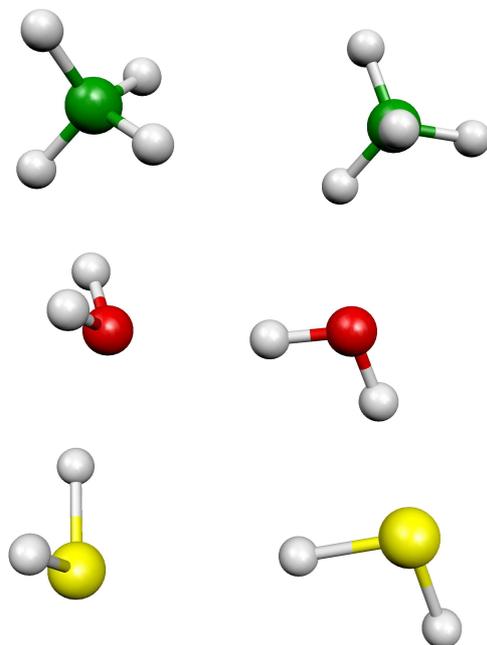


Figure 3.3: RI-MP2/aug-cc-pVDZ equilibrium structures of methane, water and hydrogen sulfide dimers.

For the methane, water and hydrogen sulfide dimers geometry optimizations were carried out at the RI-MP2/aug-cc-pVDZ [97, 98] level of theory using the TURBOMOLE 5.10 program package [99] for fixed C-C, O-O and S-S distances, respectively. No symmetry constraints were imposed. Pictures of the optimized equilibrium structures are shown in Fig. 3.3. The interaction energies were calculated at RI-MP2 geometries at the CCSD(T), CCSD and MP2 [100, 101] levels of theory using the MOLPRO program package [102] and basis sets of augmented correlation-consistent double-through quadruple- ζ quality (aug-cc-pVXZ; X=D,T,Q) [67, 103]. Additional single point calculations on the helium dimer were performed with double and triple augmented correlation-consistent basis sets of double- through sextuple- ζ quality (y-aug-cc-pVXZ; y=s,d,t; X=D,T,Q,5,6) [69]. The frozen-core approximation was applied in all calculations except for He₂.

In order to evaluate the correlation contribution to the interaction energies according to Eq. 3.8 the implementation of the incremental scheme [104] was used to get the necessary two-body increments. In order to obtain the proper fragments the domain size parameter (dsp, refer to Eq. 2.65) was set to the number of correlated orbitals on one fragment.

The convergence of the HF interaction energies was analyzed with basis sets of aug-cc-pVDZ through aug-cc-pV6Z quality at the equilibrium structures for the methane, hy-

drogen sulfide and water dimers, see Fig. 3.4. The largest change in the CP-corrected interaction energy of the quadruple- ζ basis set with respect to the sextuple- ζ basis set is only -0.01 kcal/mol. Since the differences due to the correlation contributions discussed in this work are much larger, the CP-corrected HF energies of the quadruple- ζ basis sets have been used as an approximation for the HF contributions to the extrapolated energies according to Eq. 3.3.

The discussion is limited to the CCSD(T) results in the following. The findings for the CCSD and MP2 methods show a similar behavior with respect to the accuracy and applicability of the procedure discussed. Therefore the conclusions presented below are also valid at CCSD and MP2 level of theory. The CCSD and MP2 outcomes are given in the Appendix A.

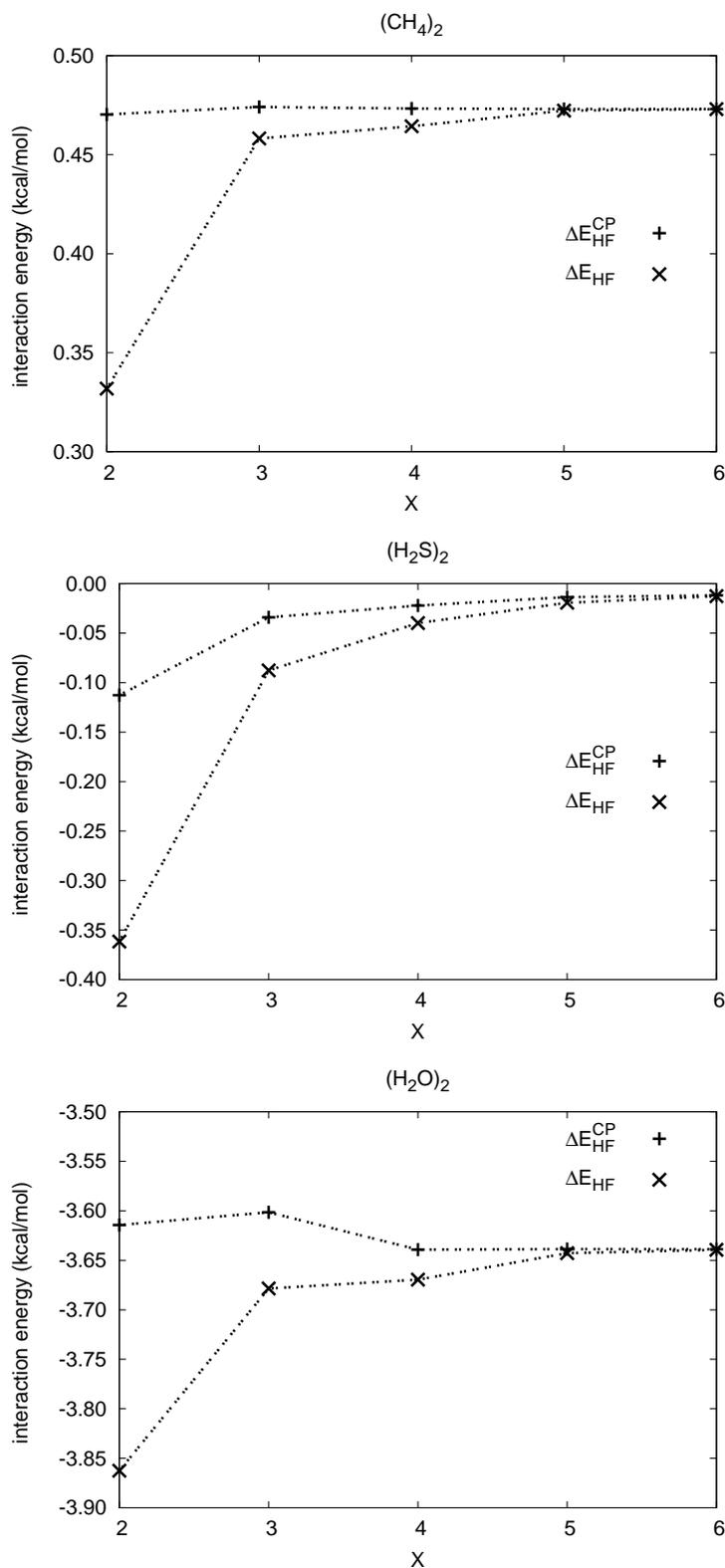


Figure 3.4: Counterpoise corrected and uncorrected HF interaction energy of the methane, hydrogen sulfide and water dimer at equilibrium distance with respect to the aug-cc-pVXZ ($X=D,T,Q,5,6$) basis set series.

3.1.2 Applications

The helium dimer was considered as a simple model system at the equilibrium separation of 5.61 Bohr [105]. The convergence behavior of Eq. 3.8 and Eq. 3.7 was checked with respect to the y-aug-cc-pVXZ (y=s,d,t and X=D,T,Q,5,6) basis set series and the results are presented in the chart of Fig. 3.5.

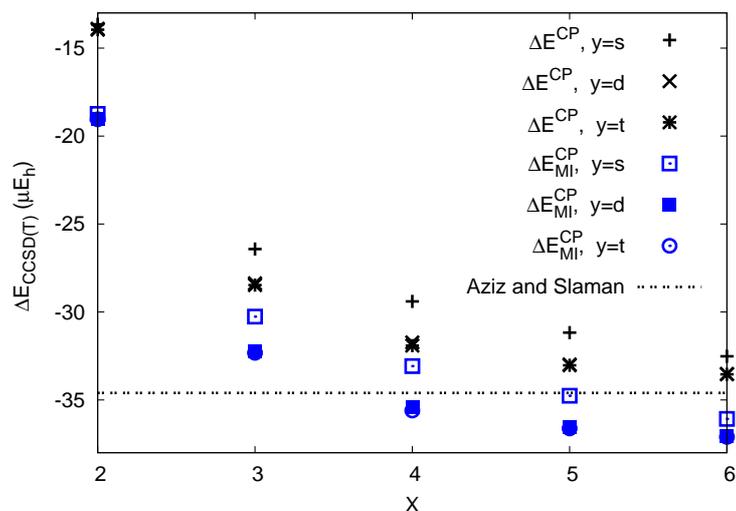


Figure 3.5: Convergence behavior of the CCSD(T) interaction energy (in μE_h) of He_2 at the equilibrium distance for the y-aug-cc-pVXZ (y=s,d,t and X=D,T,Q,5,6) basis sets. The horizontal dashed line is the well depth of He_2 taken from [105].

From the convergence of the y-aug-cc-pVXZ basis sets at different augmentation levels in Fig. 3.5 it is clear that a double augmentation is more accurate than the single augmentation for $X \geq 3$. The triple augmentation does not lead to a further improvement of the interaction energy and the d-aug-cc-pVXZ and the t-aug-cc-pVXZ energies are on top of each other. Therefore the double augmented basis sets have been used to study the convergence of the correlation contribution to the interaction energy for the approximations for $\Delta\varepsilon$, $\Delta\varepsilon^{CP}$, $\Delta\varepsilon_{MI}$ as well as the CBS extrapolated ones at the CCSD(T) level of theory. The results are presented in Fig. 3.6. The extrapolation technique was applied both to the CP-corrected and uncorrected correlation energy contributions to the interaction energies, labeled as $\Delta\varepsilon_{XY}^{CP}$ and $\Delta\varepsilon_{XY}$ respectively (see legend of the chart in Fig. 3.6).

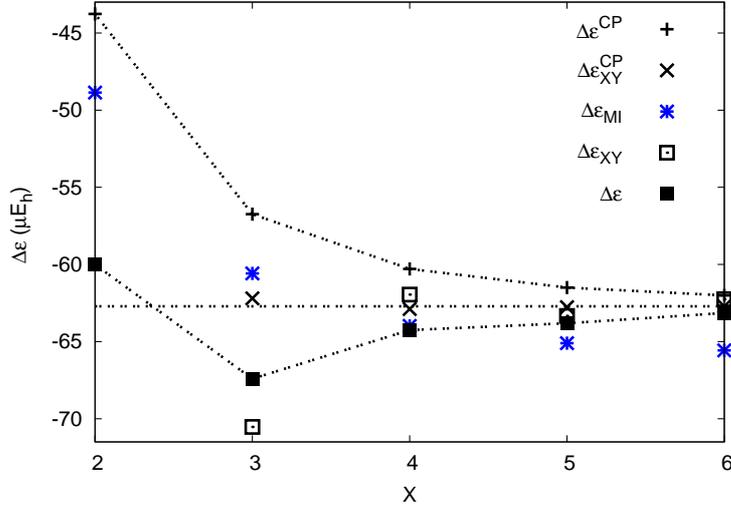


Figure 3.6: CCSD(T) correlation energy contribution (in μE_h) to the interaction energy of He_2 as a function of the cardinal number X in the d-aug-cc-pVXZ basis set series. The horizontal dashed line $\Delta\varepsilon_{56}^{CP}$ marks the estimated basis set limit.

Table 3.1: Comparison of the CCSD(T) correlation energy errors (in μE_h) for He_2 at a bond distance of 5.61 Bohr of $\Delta\varepsilon_{MI}$ and $\Delta\varepsilon^{CP}$ using the d-aug-cc-pVXZ basis set series with respect to the 5-6 extrapolated CP-corrected interaction energy $\Delta\varepsilon_{56}^{CP}$. The percentages are calculated with respect to $\Delta\varepsilon_{56}^{CP}$. The BSSE, referred to as CP correction in the sixth column of this Table, is estimated as the difference between $\Delta\varepsilon$ and $\Delta\varepsilon^{CP}$.

X	error		error		CP
	$\Delta\varepsilon_{MI}$	%	$\Delta\varepsilon^{CP}$	%	correction
2	13.85	77.91	18.94	69.80	-16.24
3	2.13	96.60	5.97	90.48	-10.67
4	-1.27	102.03	2.42	96.14	-3.96
5	-2.39	103.81	1.21	98.07	-2.31
6	-2.86	104.56	0.7	98.88	-1.14

Considering the counterpoise corrected two-point extrapolated interaction energy ($X=5$, $Y=6$ in Eq. 3.3) as estimate for the basis set limit, one finds that the CP-corrected correlation energy contribution to the interaction energy converges monotonously from above to the CBS limit. The uncorrected interaction energy $\Delta\varepsilon$ converges monotonously from below for $X=3$ to 6. The two-point X^{-3} extrapolated interaction energies based on the uncorrected energies $\Delta\varepsilon_{XY}$ are between the $\Delta\varepsilon$ and $\Delta\varepsilon^{CP}$

values except for the 2-3 extrapolated one. If the CP correction is applied in addition to the CBS extrapolation the interaction energies $\Delta\varepsilon_{XY}^{CP}$ are very close to the CBS limit already for the 2-3 extrapolation. The errors of $\Delta\varepsilon_{MI}$ and $\Delta\varepsilon^{CP}$ with respect to the 5-6 extrapolated CP-corrected energy are presented in Tab. 3.1. The $\Delta\varepsilon_{MI}$ errors are significantly smaller in comparison to the corresponding $\Delta\varepsilon^{CP}$ errors for the double-, triple- and quadruple- ζ basis sets. Using the quintuple- and sextuple- ζ basis sets, the errors of the CP-corrected interaction energy are smaller than the errors for $\Delta\varepsilon_{MI}$. The $\Delta\varepsilon_{MI}$ results overestimate the CBS limit for the quadruple to sextuple- ζ basis sets by 2, 4 and 5%, respectively, whereas the $\Delta\varepsilon^{CP}$ results converge smoothly to the CBS limit.

On the basis of these results one may conclude so far that Eq. 3.6 yields reasonable interaction energies for smaller basis sets, but it is not systematically improvable. The potential energy curves (PEC) of the interaction energy for the methane dimer at the CCSD(T)/aug-cc-pVXZ (X=D,T and Q) levels of theory are given in Fig. 3.7. The 3-4 extrapolated PECs based on the CP-corrected as well as the uncorrected energies are shown in all charts, in order to compare the results of the given basis set with the CBS limit. The difference between the ΔE_{34}^{CP} and ΔE_{34} curves is very small indicating that both ΔE^{CP} and ΔE are suitable for the 3-4 extrapolation. The ΔE^{CP} curve converges smoothly from above to the CBS limit with increasing basis set size, whereas the interaction energy curve without BSSE correction ΔE approaches the CBS limit from below when the basis set quality is improved. The BSSE is large for the double- and triple- ζ basis sets and a BSSE correction scheme is even necessary when the quadruple- ζ basis set is applied. The difference between the ΔE_{MI}^{CP} and ΔE_{MI} curves corresponds directly to the BSSE at the SCF level. The BSSE is reduced drastically for the triple- ζ basis set and almost vanishes when the quadruple- ζ basis set is applied. The ΔE_{MI}^{CP} curve is almost identical with the extrapolated curves for the double- ζ basis set, whereas the ΔE^{CP} curve underestimates the interaction energy with respect to the CBS limit. At 3.8, near the minimum of the extrapolated curve, the underestimation of ΔE^{CP} with respect to ΔE_{34}^{CP} equals 27%. The situation changes when the triple- ζ and quadruple- ζ basis sets are considered. For the triple- ζ basis set the ΔE_{MI}^{CP} curve is still between the ΔE and ΔE^{CP} curves in a region near the PEC minimum. Comparing ΔE_{MI}^{CP} to the ΔE^{CP} curve, we observe that the latter is somewhat closer to the CBS limit. For the minimum structure ΔE^{CP} underestimates ΔE_{34}^{CP} by 7%, whereas ΔE_{MI}^{CP} overestimates ΔE_{34}^{CP} by 11%. For the quadruple- ζ basis set the ΔE_{MI}^{CP} curve is below the uncorrected ΔE curve, exhibiting unacceptable large errors with respect to the CBS limit, e.g. at 3.8 ΔE_{MI}^{CP} overestimates ΔE_{34}^{CP} by 15%.

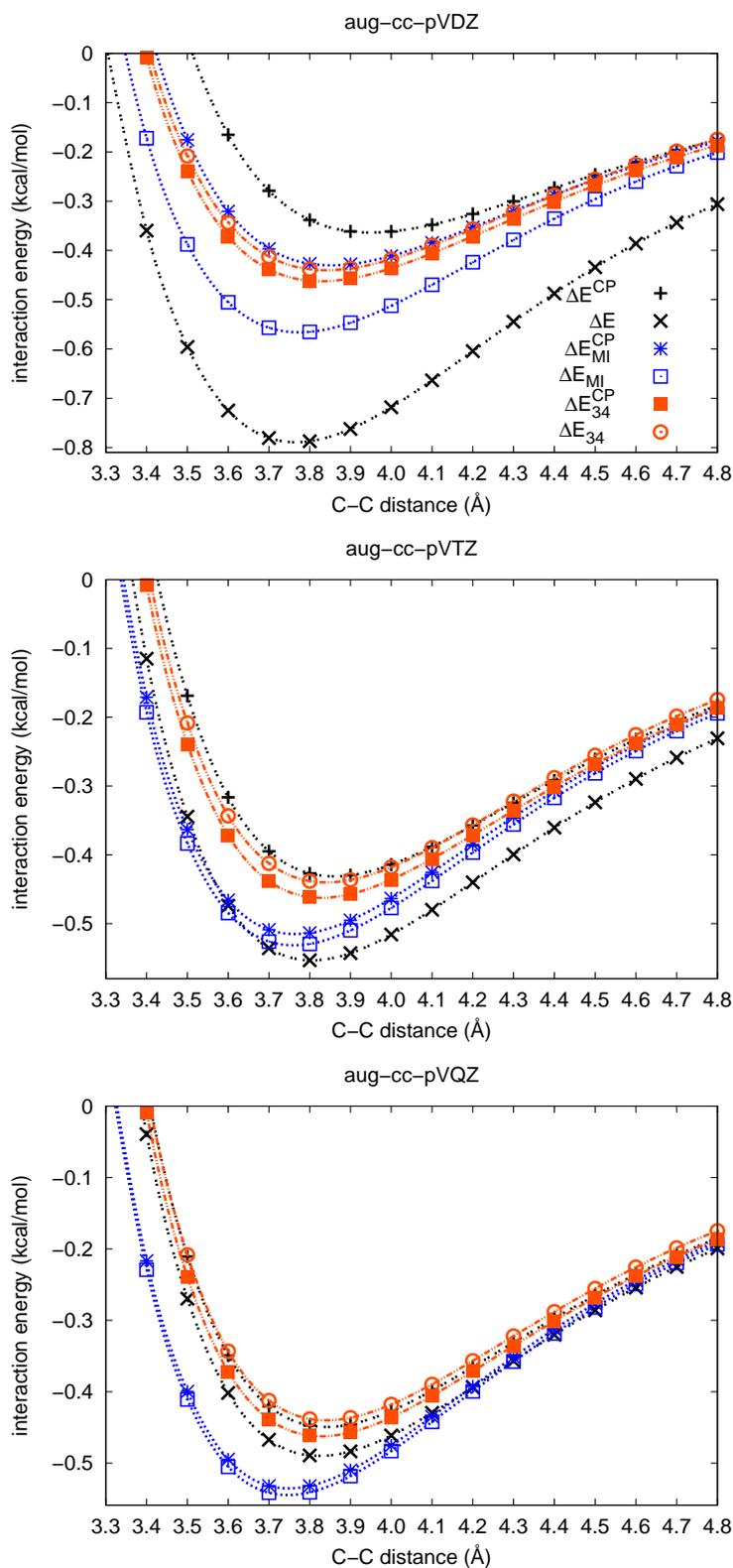


Figure 3.7: Potential energy curves of the interaction energy for the CH_4 dimer, calculated with aug-cc-pVXZ (X=D,T and Q) basis sets at the CCSD(T) level of theory as a function of the intermolecular C-C distance. The 3-4 extrapolated curves ΔE_{34}^{CP} and ΔE_{34} are based on CP-corrected and uncorrected correlation energies respectively.

For the hydrogen sulfide dimer the same analysis as for the methane dimer has been performed. Fig. 3.8 shows the comparison of the corresponding PEC for $(\text{H}_2\text{S})_2$. For all applied basis sets the interaction energies ΔE and ΔE^{CP} converge to the basis set limit from below and above, respectively. Considering the performance of Eq. 3.8 a good agreement with the CBS limit is found if the BSSE is eliminated at the HF level, i.e. for ΔE_{MI}^{CP} . Increasing the basis set to triple- or quadruple- ζ quality deteriorates the results and the PEC fall even below the lower limit of ΔE . The difference of ΔE_{MI}^{CP} and ΔE_{MI} to the 3-4 extrapolated PEC for the triple and quadruple- ζ basis sets strongly depends on the reaction coordinate of the interacting molecules. The stronger the interaction the bigger the errors and vice versa. This is due to different HF reference wavefunctions which are employed for the correlation treatment of ε_i^{12} ($i = 1, 2$) and $\varepsilon_i^{\overline{12}}$ ($i = 1, 2$). Within the incremental calculations we start with a HF solution obtained for the whole complex, whereas in the CP scheme we consider HF solutions for monomer units using the AO-basis of the dimer. This error source causes larger difference of ΔE_{MI}^{CP} with respect to the 3-4 extrapolated PEC when the interaction force is stronger.

In general one can conclude that if no calculations on the isolated monomers in whatever form are performed, the results cannot yield accurate interaction energies according to the supermolecular approach. Therefore the most accurate results presented here so far are obtained for the system with the weakest interaction in this study, i.e. the helium dimer.

As the last test case the water dimer is considered, where a dipole-dipole-interaction is the leading contribution in the total interaction energy. The PECs for the various methods are given in Fig. 3.9. Again one observe the convergence of the ΔE to the CBS-34 limit from below, whereas the CP-corrected PEC ΔE^{CP} converges from above. In contrast to the other systems Eq. 3.8 breaks down completely in this case. For all applied basis sets the PECs are below the lower bound defined by ΔE . Therefore Eq. 3.8 is not systematically improvable and also not generally applicable.

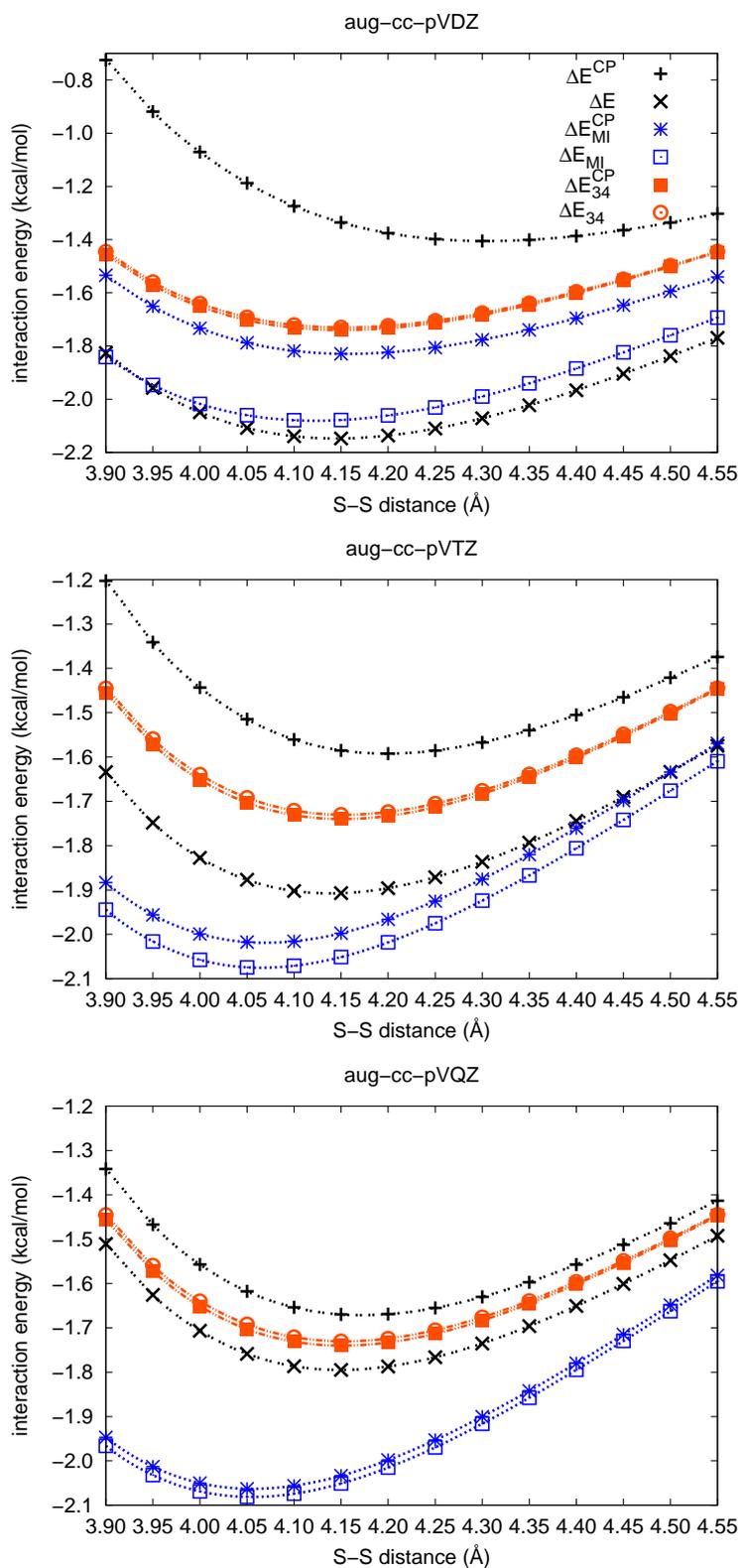


Figure 3.8: Potential energy curves of the H_2S dimer, for the aug-cc-pVXZ (X=D,T and Q) basis sets at the CCSD(T) level of theory.

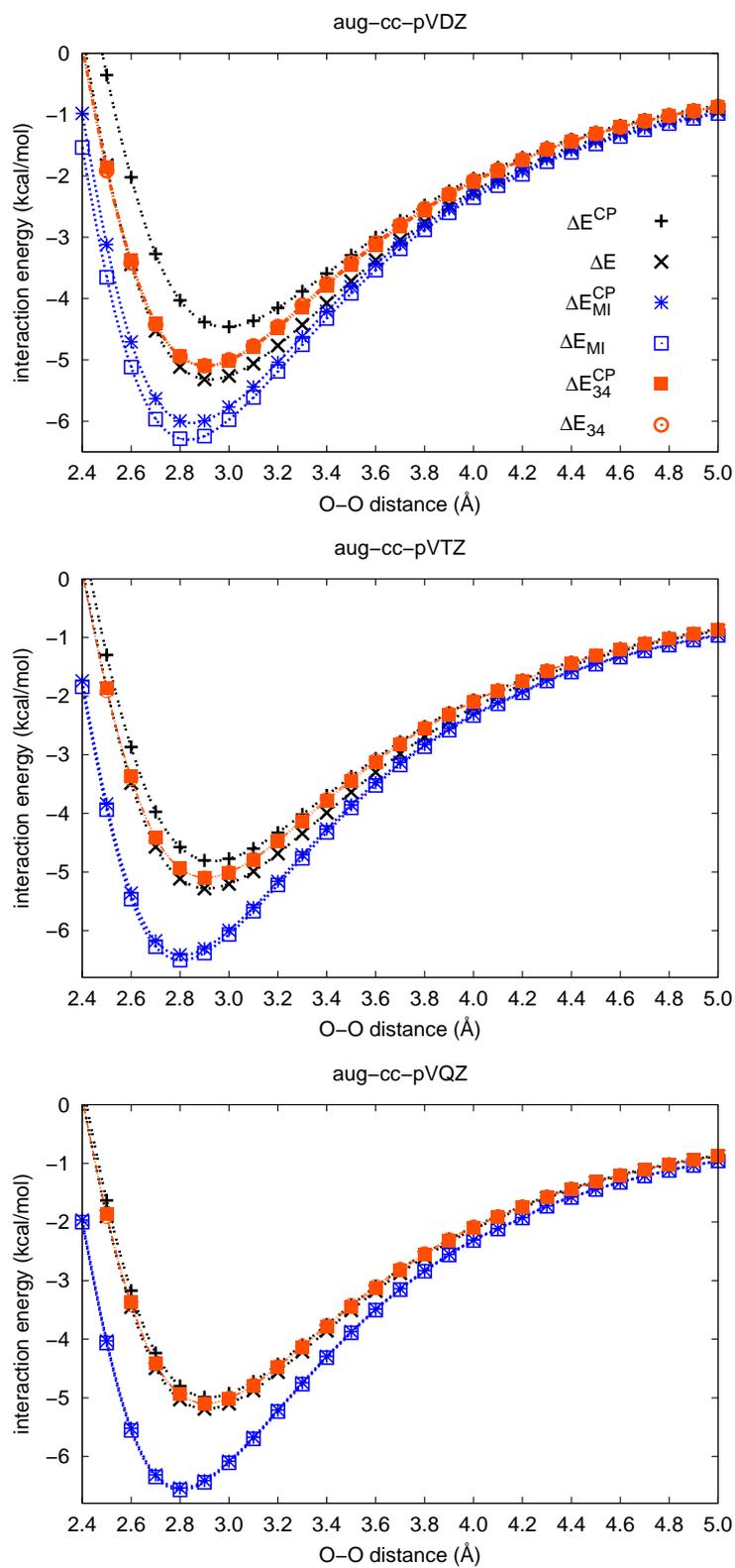


Figure 3.9: Potential energy curves of the H₂O dimer, for the aug-cc-pVXZ (X=D,T and Q) basis sets at the CCSD(T) level of theory.

A comparative summary of the ΔE_{MI}^{CP} and ΔE^{CP} errors for the equilibrium geometries of the methane, water and hydrogen sulfide dimers with respect to ΔE_{34}^{CP} is presented in Tab. 3.2.

Table 3.2: CCSD(T) errors (kcal/mol) of ΔE_{MI}^{CP} and ΔE^{CP} with respect to ΔE_{34}^{CP} at the equilibrium geometries (as predicted by ΔE_{34}^{CP}) for the basis set series aug-cc-pVXZ (X=D,T and Q).

X	error	error
	ΔE_{MI}^{CP}	ΔE^{CP}
CH ₄ -CH ₄		
2	0.03 (93%)	0.12 (73%)
3	-0.05 (111%)	0.03 (93%)
4	-0.07 (115%)	0.01 (97%)
H ₂ S-H ₂ S		
2	-0.09 (105%)	0.4 (77%)
3	-0.26 (115%)	0.15 (91%)
4	-0.29 (117%)	0.07 (97%)
H ₂ O-H ₂ O		
2	-0.9 (118%)	0.71 (86%)
3	-1.21 (124%)	0.29 (94%)
4	-1.32 (126%)	0.11 (98%)

With increasing cardinal number X the errors of the CP-corrected interaction energies ΔE^{CP} are systematically reduced and the values approach the CBS limit. This is not the case for ΔE_{MI}^{CP} , which tends to overestimate the CBS limit when larger basis sets are applied.

Beside the already mentioned different reference wavefunctions a second error source should be considered. The virtual space in the calculations of $\varepsilon_i^{\overline{12}}$ ($i = 1, 2$) is reduced in comparison to the phantom orbital calculations of ε_i^{12} ($i = 1, 2$). The decrease of the external orbitals is equal to the number of occupied orbitals frozen in the second monomer unit, which are excluded from the correlation treatment. The reduced number of external orbitals yields less negative $\varepsilon_i^{\overline{12}}$ ($i = 1, 2$) energies with respect to the CP-corrected outcomes and thus in turn causes an overestimation of the interaction energy. As evidence by the following ordering: $\Delta\varepsilon < \Delta\varepsilon_{MI} < \Delta\varepsilon^{CP}$ of the correlation energy contribution to the interaction energy of the H₂S-benzene complex as presented in Fig. 3.1. The good agreement of ΔE_{MI}^{CP} with respect to the CBS limit for the double- ζ basis sets (see Table 3.2) seems therefore to be rather a consequence of a favorable error cancellation than of an accurate description of the interaction energy

with the approach of Eq. 3.8. Note that the proposal of Wang and Paulus is indeed in line with the so-called virtual counterpoise scheme (VCP) [106]. In this scheme the monomer calculations are performed in a reduced dimer basis where the 'to be occupied orbitals' of the second monomer are projected out. The reason why Daudey et al. [106] suggested the VCP scheme is strongly related to one of the most often discussed issues regarding the counterpoise approach, that is the question whether a CP method overcorrects the interaction energy or not. The physical explanation for the validity of the VCP scheme is based on the Pauli exclusion principle. In a correlated CP calculations on a monomer, the excitations to orbitals corresponding to the occupied MOs of the second monomer are allowed, however in the dimer they are not [107]. This approach was discussed in detail in a review article by van Duijneveldt et al. [108]. It has been concluded on the basis of several studies [109–111] by comparing the results obtained with VCP and CP that only the latter scheme provides the proper correction to the interaction energies [112, 113]. The argument that in the limiting case of a complete dimer basis the monomer calculations should be performed in this complete basis too (CP scheme) rather than in a reduced basis (VCP scheme) also applies to the proposed scheme, which is thus bound to overestimate the interaction energies for large basis sets. These findings are in line with the discussion carried out here on the errors caused by the reduced number of external orbitals.

It should be noted that the procedure defined by Eq. 3.8 is an extremely simplified variant of the method of increments. The usage of the name incremental method for this approach is even misleading. Eq. 3.8 simply approximates the correlation part of the interaction energy of a dimer by a single two-body increment, i.e. the correlation contribution Eq. 3.6 is formally obtained as the difference between a second-order expansion for the dimer and first-order expansions for the monomers under the assumption, that the first-order contributions of these expansions completely cancel when taking the difference. Thus the highest order increment in the dimer expansion has to be evaluated, which corresponds to the full calculation, whereas the cheaper evaluation of lower-order increments is avoided. In contrast to this the incremental scheme [39, 42] aims at evaluating the total correlation energy of a system by considering only increments of relatively low order and thus avoids the expensive evaluation of the high-order contributions [43, 104, 114, 115]. Thus, for the proposal of Paulus et al. no advantage with respect to the computational effort is obtained. The computational cost of their procedure as well as of the CP approach is dominated by the most expensive calculation for the dimer, which is completely equivalent for both schemes. For the monomers one saves in the method suggested by Paulus et al. the HF calculations in the dimer basis. However, since a CP correction is necessary at the HF level of theory, one does neither save CPU-time nor hand work.

From the presented results it is clearly evident that the recent attempt to identify the

intermolecular second-order correlation energy increments as an approximation to the counterpoise corrected interaction energies of weakly interacting monomers can not be recommended as an accurate method in general. It is shown that the results obtained via Eq. 3.8 are not systematically improvable and that the range of applicability is limited. The incremental correlation energies $\Delta\varepsilon_{MI}$ overestimate the interaction energy in a constant manner with respect to the counterpoise corrected energies. A comparison of $\Delta\varepsilon_{MI}$ and the complete basis set limit reveals that reasonable results for the dimers of helium, hydrogen sulfide and methane are only obtained with basis sets of moderate size, and the corresponding good results appear to be due to a beneficial error cancellation. For the water dimer with strong hydrogen bonding it was found that Eq. 3.8 yields unreasonable large errors. For the hydrogen sulfide and methane dimer in quadruple- ζ basis set Eq. 3.8 also fails to estimate accurate interaction energies with respect to the complete basis set limit as well as with respect to the counterpoise approach.

The overall conclusion is that the incremental method cannot provide a direct procedure to determine BSSE-free interaction energies. One has to use one of the various *a posteriori* correction methods. Concerning weak intermolecular interactions the incremental scheme in fact can provide accurate total energies for the interacting system and therefore also accurate interaction energies at reduced computational effort. However an efficient evaluation of the BSSE effect, especially when large molecular clusters are considered, has not yet been introduced together with the incremental approach. This is actually the goal of this work.

3.2 BSSE correction schemes for n-body clusters

3.2.1 SSFC, PAFC and VMFC schemes

Within the supermolecule approach one can calculate the interaction energy (D) of n -body clusters from the energy difference of the cluster energy $E_{ij\dots n}^{ij\dots n}$ and its individual monomer fragments \bar{E}_i^i according to:

$$D = E_{ij\dots n}^{ij\dots n} - \sum_i^n \bar{E}_i^i, \quad (3.9)$$

where the bar in \bar{E}_i^i indicates that the energy is calculated at the relaxed geometry. Consistently with the notation of Eq. 3.9 the omission of the bar symbol at E_i^{ij} and E_j^{ij} in Eq. 3.2 means that these monomer energies are calculated at the monomer geometry taken from the dimer.

In the limit of the complete basis set the dimer interaction energy calculated via Eq. 3.9 (for $n=2$) and Eq. 3.2 will not be affected by the BSSE. However the obtained results will (except for diatomic systems) not converge to the same value since the monomer energies are evaluated at different geometries (isolated monomer vs. monomer in the complex) [116, 117]. In order to circumvent this discrepancy one has to account for the geometry relaxation contribution terms ΔE_{rel}^i and ΔE_{rel}^j . So the counterpoise corrected interaction energy ΔE_{rel}^{CP} which takes into account the relaxation energy in case of a dimer is:

$$\Delta E_{rel}^{CP} = \Delta E^{CP} + \Delta E_{rel}^i + \Delta E_{rel}^j, \quad (3.10)$$

where we also assume that the CP-correction and the relaxation contributions behave additiv. The fragment relaxation energies describe the energy penalty for distorting the monomers from their isolated geometries to the ones in the complex:

$$\begin{aligned} \Delta E_{rel}^i &= E_i^i - \bar{E}_i^i, \\ \Delta E_{rel}^j &= E_j^j - \bar{E}_j^j. \end{aligned} \quad (3.11)$$

Similar consideration may be carried out also for n -body interaction energies. If we aim to account for BSSE effects in n -body clusters, we may invoke the counterpoise procedure and expand Eq. 3.2 according to:

$$D(fCP) = E_{ij\dots n}^{ij\dots n} - \sum_i^n E_i^{ij\dots n}, \quad (3.12)$$

where all individual monomers calculated in the basis of the whole supermolecule $E_i^{ij\dots n}$ are subtracted from the cluster energy $E_{ij\dots n}^{ij\dots n}$. Again the omission of the bar symbol at $E_i^{ij\dots n}$ indicates that the geometries are those from the cluster and hence

Eq. 3.12 is also referred to as full counterpoise corrected interaction energy (fCP). The geometry relaxation contributions may also be generalized for the n-body case according to:

$$\sum_i^n \Delta E_{rel}^i = \sum_i^n \left(E_i^i - \bar{E}_i^i \right). \quad (3.13)$$

Together with Eq. 3.12 and Eq. 3.13 we may finally formulate the BSSE corrected interaction energy for n-body clusters which also take the fragment relaxation effects into account as:

$$D(SSFC) = D(fCP) + \sum_i^n \Delta E_{rel}^i, \quad (3.14)$$

recasting Eq. 3.14 leads to:

$$D(SSFC) = E_{ij\dots n}^{ij\dots n} - \sum_i^n \bar{E}_i^i + \sum_i^n \left(E_i^i - E_i^{ij\dots n} \right) \quad (3.15)$$

The last term in Eq. 3.15 is well-known from the literature [118] as a BSSE correction for n-body interactions, called site-site function counterpoise (SSFC) correction:

$$SSFC = \sum_i^n \left(E_i^i - E_i^{ij\dots n} \right). \quad (3.16)$$

Note that at the complete basis set limit Eq. 3.16 approaches zero and hence $D(SSFC)$ converges to D .

In the framework of the counterpoise method at least two more extensions exist which correct for BSSE in aggregates with various subunits. Wells and Wilson [118] proposed a pairwise additive function counterpoise (PAFC) scheme which describes the BSSE correction of the pairwise interaction of each monomer with every other monomer in the cluster:

$$PAFC = \sum_{i \neq j}^n \left(E_i^i - E_i^{ij} \right). \quad (3.17)$$

Valiron and Mayer [119], based on earlier work of White and Davidson [120], introduced a hierarchical scheme, the so-called Valiron-Mayer function counterpoise (VMFC) correction.

$$\begin{aligned} VMFC = & \sum_i^n \left(E_i^i - E_i^{ij\dots n} \right) + \sum_{i < j}^n \left(\Delta E_{ij}^{ij} - \Delta E_{ij}^{ij\dots n} \right) \\ & + \sum_{i < j < k}^n \left(\Delta E_{ijk}^{ijk} - \Delta E_{ijk}^{ijk\dots n} \right) + \dots \\ & + \sum_{i < j < \dots < (n-1)}^n \left(\Delta E_{ij\dots(n-1)}^{ij\dots(n-1)} - \Delta E_{ij\dots(n-1)}^{ij\dots(n-1)} \right) \end{aligned} \quad (3.18)$$

The first-order term (VMFC(1)) from Eq. 3.18 equals to the SSFC scheme. Higher-order BSSE energy contributions treat the basis set extension effects of all dimers, trimers up to n-mers present in a cluster and are calculated in a recursive manner. Exemplary the second-order of the VMFC scheme is given by Eq. 3.19:

$$VMFC(2) = \sum_{i < j}^n \left(\Delta E_{ij}^{ij} - \Delta E_{ij}^{ij\dots n} \right) = \sum_{i < j}^n \left((E_{ij}^{ij} - E_i^{ij} - E_j^{ij}) - (E_{ij}^{ij\dots n} - E_i^{ij\dots n} - E_j^{ij\dots n}) \right), \quad (3.19)$$

whereas the third-order term reads as:

$$VMFC(3) = \sum_{i < j < k}^n \left(\Delta E_{ijk}^{ijk} - \Delta E_{ijk}^{ijk\dots n} \right) = \sum_{i < j < k}^n \left((E_{ijk}^{ijk} - \Delta E_{ij}^{ijk} - \Delta E_{ik}^{ijk} - \Delta E_{jk}^{ijk}) - (E_{ijk}^{ijk\dots n} - \Delta E_{ij}^{ijk\dots n} - \Delta E_{ik}^{ijk\dots n} - \Delta E_{jk}^{ijk\dots n}) \right). \quad (3.20)$$

Note that within the notation used in this work the sum of first and second-order VMFC corrections is indicated by VMFC(1,2), whereas the VMFC(2) term describes the energy contributions arising from second-order corrections only.

One important characteristic of these schemes is the computational cost they cause. The number of individual calculations [121] which have to be performed to calculate the VMFC correction is equal to:

$$n_{calc.} = \sum_{i=1}^n 2^{n-i} \binom{n}{i}. \quad (3.21)$$

Obviously the amount of individual calculations increases very rapidly with increasing cluster size n . Therefore the VMFC scheme is applicable only to small cluster sizes. Exemplary the full VMFC BSSE correction calculation of nonsymmetric clusters with $n=6,8,10$ and 12 subunits would require $665, 6305, 58025$ and 527345 individual calculations, respectively [121]. The scaling behavior of the PAFC is n^2 and for the SSFC $2n$, which is quite low in comparison to Eq. 3.21. However, the size of the one-particle basis set employed in the individual calculations is also a limiting factor determining the computational cost of a BSSE calculation. From this perspective the PAFC method is computationally the most attractive, since the most expensive calculations needed are just in a dimer basis set, independent from the cluster size. The VMFC scheme is definitely the most expensive one, not only due to the huge number of individual calculations, but also because the most expensive ones are almost as large as the cluster

calculation itself, namely the calculation of $E_{ij\dots(n-1)}^{ij\dots n}$. Even though the SSFC scheme requires the smallest number of individual calculations it may become computational extremely demanding when BSSE corrections for larger cluster sizes are considered, since any monomer has to be calculated in the presence of the whole n -mer basis set.

3.2.2 Approximate SSFC(R) and VMFC(2)(R)

As pointed out in contrast to the computationally less demanding PAFC scheme the SSFC and mainly the VMFC methods cause an enormous computational effort. It is therefore desirable to develop approximations to these schemes. The ones introduced here aim at a reduction of the size of the n -mer basis set. The basic assumption employed here is that standard Gaussian basis sets are utilized, e.g. basis sets optimized for the corresponding atom and not aiming to describe functions of other centers. Since atom-centered Gaussian basis functions are used we may expect that far distant functions contribute only little to the additional flexibility of the wavefunction at a specific monomer.

Due to the steep scaling of Eq. 3.21 the implementation of the VMFC scheme is restricted to the second-order correction only. In order to carry out SSFC as well as VMFC(2) calculations automatically one needs to partition the n -body cluster into its monomer fragments. The fragmentation procedure is performed automatically with the usage of a graph partitioning routine [51]. The partitioning into monomer fragments is analogous to the construction of the one-site domains within the incremental scheme. But the set of vectors from which the distance matrix is constructed is no longer a set of centers of charge as in case of the incremental scheme but simply the set of atom coordinates of the cluster under consideration. Therefore one distinction has to be drawn, the first condition in the connectivity matrix Eq. 2.64 is not needed anymore as we may exclude the case that the distance between two atom coordinates is smaller than 10^{-4} Å.

Having the sets of coordinates which form the monomers constructed we turn our attention to the superscripts which occur in the second sum of Eq. 3.16 and Eq. 3.19. Firstly note that the individual calculations from the second sum in Eq. 3.16 and Eq. 3.19 are performed using so-called ghost atoms. The term ghost is used because these atoms contribute only their basis functions but neither their nuclei nor electrons to a quantum chemical calculation within the LCAO approach. The overall basis set is the sum of basis functions which are contributed by the atom coordinates i (the lower index in Eq. 3.16 and Eq. 3.19) and all the remaining ghost atoms of a given cluster. Therefore if we want to exclude in a given monomer ghost calculation basis functions contributed by far away ghost atoms *a priori*, we need a procedure to identify them. We employ the information about the distances of atom coordinates from a

given monomer fragment to all possible ghost atom coordinates of the n-body cluster in order to decide whether a basis function will significantly contribute to the BSSE calculation or not. The reduced basis set for the BSSE calculation is constructed as follows: use a distance threshold R to select all ghost atom coordinates from the whole molecule, which have a smaller distance than R to at least one atom of the fragment i . Since these selected ghost coordinates are associated with atoms of the system, they are associated with AO-basis functions, too. These AO-basis functions together with those arising from the lower index in Eq. 3.16 are exactly the basis set $\mathbb{B}_i(R)$ of the BSSE calculation. Having the set of the truncated ghost orbital space constructed, we may rewrite Eq. 3.16 into the approximate one:

$$SSFC(R) = \sum_i^n \left(E_i^i - E_i^{\mathbb{B}_i(R)} \right) \quad (3.22)$$

Note that in the limit when the distance threshold reaches infinity, the approximate SSFC(R) scheme is equal to SSFC:

$$SSFC(R \rightarrow \infty) = SSFC. \quad (3.23)$$

The distance-dependent screening procedure is also applied to the second-order VMFC scheme and hence Eq. 3.19 is approximated via:

$$VMFC(2)(R) = \sum_{i < j}^n \left(\left(E_{ij}^{ij} - E_i^{ij} - E_j^{ij} \right) - \left(E_{ij}^{\mathbb{B}_{ij}(R)} - E_i^{\mathbb{B}_{ij}(R)} - E_j^{\mathbb{B}_{ij}(R)} \right) \right), \quad (3.24)$$

where the truncated ghost orbital space is formed through the unification of the proper ghost coordinate sets for a given monomer pair:

$$\mathbb{B}_{ij}(R) = \mathbb{B}_i(R) \cup \mathbb{B}_j(R). \quad (3.25)$$

As indicated in Eq. 3.24 the evaluation of VMFC(2)(R) includes the calculation of monomer and dimer energies with the dimer basis set as well as the truncated basis set. Analogously to Eq. 3.23 in the limit when $R \rightarrow \infty$ we arrive at the VMFC(2) scheme: $VMFC(2)(\infty) = VMFC(2)$.

The proposed approximate scheme removes far distant functions in a systematic fashion so that one can control the level of desired accuracy of the BSSE correction on one hand and the gained savings in computational resources on the other hand.

3.2.3 Approximate SSFC_{inc} scheme

The approximate SSFC_{inc} scheme is termed incremental approach because the underlying procedure used to account for SSFC_{inc} closely reflects the expansion of the incremental scheme as described in Eq. 2.58. The SSFC_{inc} BSSE correction is evaluated as a sum of fragmental BSSE contributions according to:

$$SSFC_{inc} = \sum_i^n \left(\sum_{\substack{j \\ i \neq j}}^n \Delta\epsilon_i^{ij} + \sum_{\substack{j < k \\ i \neq j, k}}^n \Delta\epsilon_i^{ijk} + \dots + \sum_{\substack{j < k < \dots < m \\ i \neq j, k, \dots, m}}^n \Delta\epsilon_i^{ijk\dots m} \right), \quad (3.26)$$

where $\Delta\epsilon_i^{ij}$ refers to the fragmental first-order correction, $\Delta\epsilon_i^{ijk}$ to the fragmental second-order correction and finally $\Delta\epsilon_i^{ijk\dots m}$ marks the highest possible correction order which depends on the number of monomers present in the n-body cluster.

The fragmental first-order BSSE contribution is evaluated as the difference between the monomer energy in its own basis E_i^i and a second monomer E_i^{ij} :

$$\Delta\epsilon_i^{ij} = E_i^i - E_i^{ij} \quad (3.27)$$

The basis set enlargement built from basis function which correspond to ghost atoms of the remaining monomers is systematically increased with increasing order. Thus the second-order BSSE increments read:

$$\Delta\epsilon_i^{ijk} = E_i^i - E_i^{ijk} - \Delta\epsilon_i^{ij} - \Delta\epsilon_i^{ik}, \quad (3.28)$$

where E_i^{ijk} as well as the first-order correction $\Delta\epsilon_i^{ij}$ and $\Delta\epsilon_i^{ik}$ are subtracted from the monomer in its own basis E_i^i . The third-order BSSE correction is formulated analogously:

$$\Delta\epsilon_i^{ijkl} = E_i^i - E_i^{ijkl} - \Delta\epsilon_i^{ijk} - \Delta\epsilon_i^{ijl} - \Delta\epsilon_i^{ikl} - \Delta\epsilon_i^{ij} - \Delta\epsilon_i^{ik} - \Delta\epsilon_i^{il}. \quad (3.29)$$

So the general formula for the fragmental BSSE contribution reads:

$$\begin{aligned} \Delta\epsilon_i^{ij\dots m} = & E_i^i - E_i^{ij\dots m} - \sum_{i < j} \Delta\epsilon_i^{ij\dots pom} - \sum_{i < j} \Delta\epsilon_i^{ij\dots po} \\ & - \sum_{i < j} \Delta\epsilon_i^{ij\dots p} - \dots - \sum_{i < j} \Delta\epsilon_i^{ij}. \end{aligned} \quad (3.30)$$

Regarding the expansion order of Eq. 3.26 we can distinguish two limiting cases, once the truncation of the series after first-order and on the other hand an expansion up to the highest order. In the later case the SSFC_{inc} scheme turns into the SSFC approach,

what is exemplary proven for the case of a trimer cluster:

$$\begin{aligned}
SSFC_{inc} &= \sum_i^3 \left(\sum_{\substack{j \\ i \neq j}}^3 \Delta \epsilon_i^{ij} + \sum_{\substack{j < k \\ i \neq j, k}}^3 \Delta \epsilon_i^{ijk} \right) \\
&= \Delta \epsilon_1^{12} + \Delta \epsilon_1^{13} + \underline{\Delta \epsilon_1^{123}} + \Delta \epsilon_2^{21} + \Delta \epsilon_2^{23} + \underline{\Delta \epsilon_2^{213}} + \Delta \epsilon_3^{31} + \Delta \epsilon_3^{32} + \underline{\Delta \epsilon_3^{312}} \\
&= \Delta \epsilon_1^{12} + \Delta \epsilon_1^{13} + \underline{E_1^1 - E_1^{123} - \Delta \epsilon_1^{12} - \Delta \epsilon_1^{13}} + \\
&\quad \Delta \epsilon_2^{21} + \Delta \epsilon_2^{23} + \underline{E_2^2 - E_2^{123} - \Delta \epsilon_2^{21} - \Delta \epsilon_2^{23}} + \\
&\quad \Delta \epsilon_3^{31} + \Delta \epsilon_3^{32} + \underline{E_3^3 - E_3^{123} - \Delta \epsilon_3^{31} - \Delta \epsilon_3^{32}} \\
&= E_1^1 - E_1^{123} + E_2^2 - E_2^{123} + E_3^3 - E_3^{123} \\
&= \sum_i^3 (E_i^i - E_i^{ijk}) = SSFC
\end{aligned} \tag{3.31}$$

Whereas the $SSFC_{inc}$ first-order correction is the pairwise additive function counterpoise scheme:

$$SSFC_{inc}(first) = \sum_i^n \sum_{\substack{j \\ i \neq j}}^n \Delta \epsilon_i^{ij} = \sum_{i \neq j}^n (E_i^i - E_i^{ij}) = PAFC \tag{3.32}$$

The number of individual calculations needed for the whole series of Eq. 3.26 is:

$$n_{calc.} = n + \sum_{i=1}^{n-1} \binom{n-1}{i} n. \tag{3.33}$$

Exemplary for cluster sizes with $n=6,8,10$ and 12 the evaluation of Eq. 3.33 yields 192, 1024, 5120 and 24576 individual calculations, respectively. Of course advantages over the SSFC scheme will only occur when the expansion is truncated at a low order. The restriction to second-order fragmental contributions, leads to 96, 232, 460 and 804 individual calculations for cluster sizes $n=6,8,10$ and 12 respectively. Even though the number of individual calculations is still considerable high for a second-order expansion, one should remember that the most expensive calculations are those of E_i^{ijk} in the basis of a trimer, independent from the size of the cluster.

3.2.4 Approximate SSFC(S) scheme

Finally we propose an approximation to the SSFC scheme based on the findings of Kalvoda et al. [122]. The authors calculated the cohesive energy of a GdN cluster within the incremental method. As an estimate for the BSSE at the HF level of theory

they proposed to calculate the atomic energies of N and Gd with merely the diffuse and polarization functions, respectively placed on the ghost centers surrounding either N or Gd.

A procedure which aims at the elimination of the most dense basis functions and a survival of the diffuse ones at the ghost centers may be realized with the inspection of the overlap integrals from the overlap matrix \mathbf{S} . We call this scheme SSFC(S) method and introduce a reduced set of basis functions for every monomer i $\mathbb{B}_i(S)$:

$$SSFC(S) = \sum_i^n \left(E_i^i - E_i^{\mathbb{B}_i(S)} \right). \quad (3.34)$$

Partitioning a cluster according to the method outlined in 3.2.2, yields the information which atoms correspond to either the set of ghost centers or the set of atoms forming the active molecule, described by the subscript in 3.34. As a result we gain a set of active molecules, where each one in turn correspond to a set of atoms and every atom map to a set of basis functions. Similar consideration apply to the sets of ghost centers. In order to utilize the value of an overlap integral as a measure whether to incorporate a given basis function on a ghost center or not we need to consider all values of the overlap integrals from the set of basis functions of the active part to the set of basis functions corresponding to the ghost centers. When only one overlap between a basis function of the active part to a regarded basis function from the ghost set is greater than a chosen S threshold then the corresponding basis function survives the screening procedure.

This method allows to establish a systematically reduced set of basis functions $\mathbb{B}_i(S)$ for each monomer i . The resulting reduced basis set will depend upon the exponents of basis functions delivered within a chosen basis set, as well as the distance of the atoms where the two basis functions are placed at.

3.3 Applications

The performance and applicability of the approximate schemes SSFC(R), SSFC(S), SSFC_{inc} and the VMFC(2)(R) have been mainly tested for a water cluster series $(\text{H}_2\text{O})_n$ with sizes n ranging from 6 to 20. The optimized structures of the investigated clusters are shown in Fig. 3.35 and have been taken from [123].

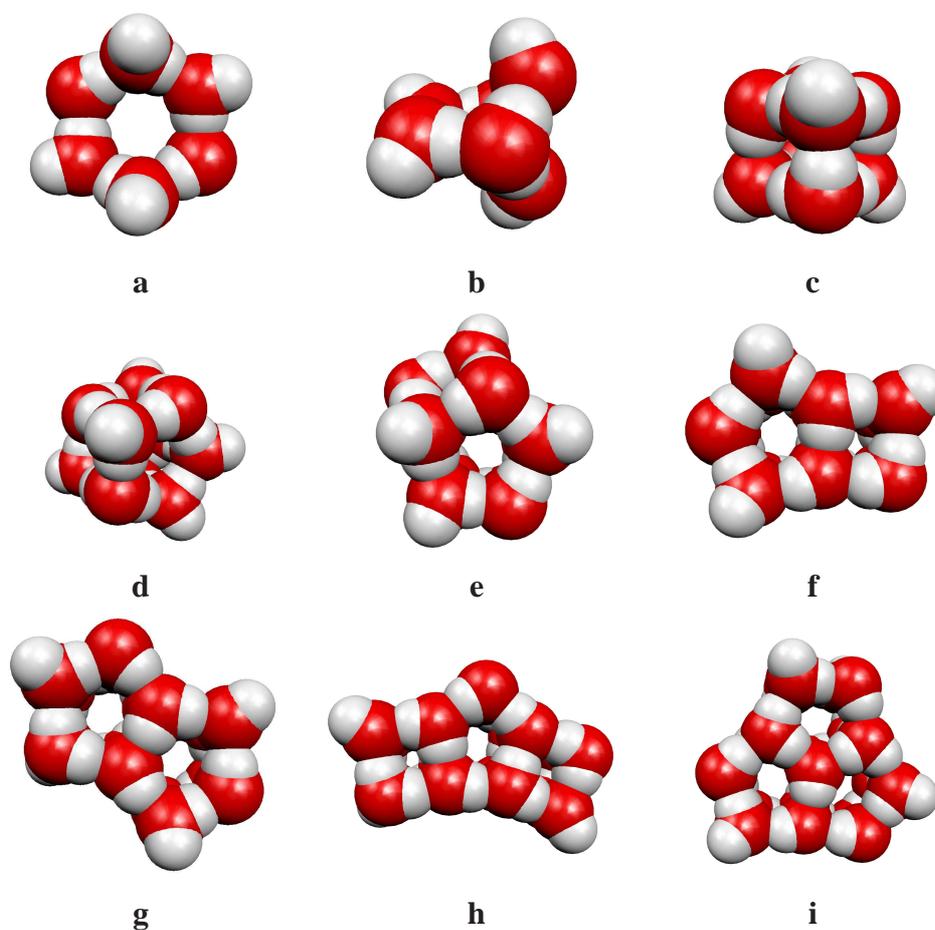


Figure 3.10: Optimized structures of the water cluster series $(\text{H}_2\text{O})_n$; $n=6$ boat structure (a), $n=6$ bag structure (b), $n=8,10,\dots,20$ (c-i) [123].

Furthermore the SSFC(R) scheme was also applied to calculate the BSSE of a methanol cluster series, the optimized geometries has been taken from [124] and are presented in Fig. 3.11.

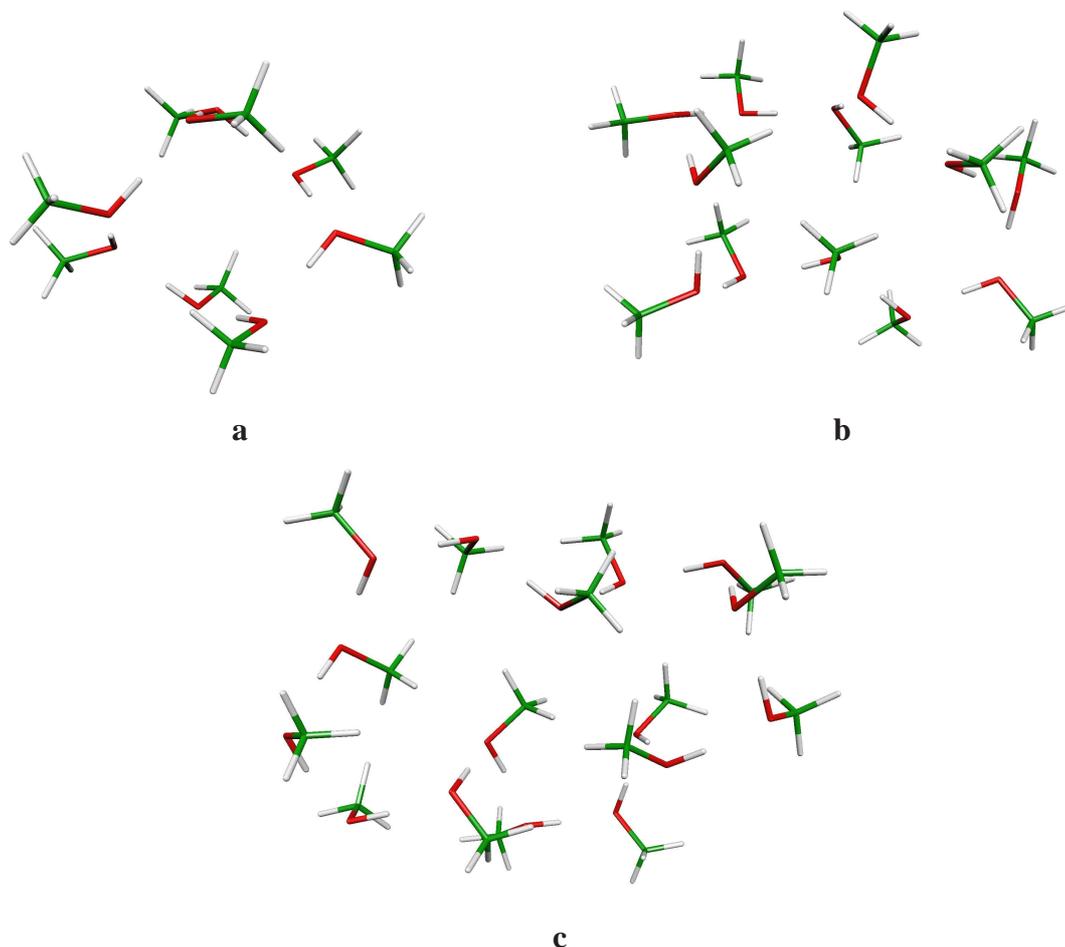


Figure 3.11: Optimized structures of the methanol cluster series $(\text{CH}_3\text{OH})_n$; $n=8$ (a), $n=12$ (b) and $n=16$ (c) [124].

In order to arrive at a proper partitioning of the methanol and water clusters the connectivity parameter t_{con} for the Metis routine was set to 1.3 (Å). The BSSE SSFC(R), SSFC_{inc} and VMFC(2)(R) calculations were carried out within the MOLPRO quantum chemistry package [58, 100, 101]. Whereas the SSFC(S) calculations were performed within the TURBOMOLE 6.3 program package [125].

Beside the BSSE correction itself we calculated also the stabilization energies of the water clusters according to Eq. 3.9. The calculation of the total energy of the cluster $E_{ij\dots n}^{ij\dots n}$ is carried out with the incremental scheme using the domain-specific basis set approach [52]. For the description of the environment of the domains we employed Pople's basis sets, namely the STO-3G for hydrogen and the 6-31G basis set for oxygen [126, 127]. The size of this second basis is controlled with the t_{main} parameter which was set to 3 Å.

3.3.1 Comparison of the SSFC, PAFC and VMFC(2) schemes

Before starting to examine the applicability of the approximate BSSE correction schemes the performance of SSFC, PAFC and VMFC(2) schemes is checked in general for the water cluster series $(\text{H}_2\text{O})_n$ (with n up to 20). Table 3.3 summarizes the BSSE corrections according to Eq. 3.16, 3.17 and 3.19 at the CCSD(T), CCSD and MP2 level of theory using aug-cc-pVDZ basis sets (and for the $(\text{H}_2\text{O})_6$ cluster also for the aug-cc-pVTZ basis sets).

The SSFC and PAFC BSSE values grow fast with the cluster size n at all the presented theory levels. The SSFC and PAFC methods estimate the BSSE correction to be even bigger than 40 kcal/mol for the biggest cluster sizes with $n=18$ and 20 at all correlation levels. For moderate cluster sizes with $n=12,14$ and 16 a SSFC and PAFC BSSE correction around 20 kcal/mol is observed. But even for the smallest clusters of the water series the SSFC and PAFC BSSE values are still considerably high at least for the double- ζ basis set and have definitely to be considered if one aims at an accurate description of interaction energies.

The aug-cc-pVDZ/CCSD(T) SSFC BSSE correction per monomer equals 0.97, 1.45, 1.48, 1.53, 1.63, 1.66 and 2.17 kcal/mol for $n=6,8,\dots,18$, respectively, growing slow but continuously with the cluster size n . Analog considerations for the PAFC outcomes reveal the following sequence of BSSE corrections: 1.04, 1.63, 1.65, 1.68, 1.83, 1.86, 2.51 and 1.95 kcal/mol also for $n=6,8,\dots,20$, respectively.

In contrast to these results the magnitude of the second-order BSSE correction VMFC(2) is very small with values around one tenths of a kcal/mol for $n=6$ and at most 1.3 kcal/mol for $n=12$ for the CCSD(T) theory and 1.8 kcal/mol for $n=16$ at MP2 theory level. The VMFC(2) is therefore only a slight improvement to the SSFC scheme which is according to Eq. 3.18 referred as the first-order of the VMFC scheme. The percentage of the VMFC(2) fraction with respect to VMFC(1,2) increases rapidly with the cluster size increasing from hexamer to octamer and stays rather constant for clusters larger or equal than the water decamer. This behavior may be explained when we refer to the geometry pattern of the investigated clusters. From the geometrical arrangement of the water clusters presented in Fig. 3.35 we note, that every monomer in the regarded boat structure of the water hexamer is coordinated to just two another water monomers. Whereas the number of hydrogen bonds felt by each monomer in the larger clusters is at least and also mostly three. Altogether we may therefore say that the VMFC(2) correction becomes more important when bigger cluster sizes are considered. Note that the second-order BSSE correction contributes even around 7% to the VMFC(1,2) value for cluster sizes between $n=8$ to $n=12$ for the CCSD(T) theory level.

Regarding the BSSE magnitude with respect to different correlation treatment we find

the highest and lowest BSSE estimates at CCSD(T) and CCSD level of theory respectively for all correction schemes considered and the aug-cc-pVDZ basis set. The MP2 theory yields BSSE values in between CCSD and CCSD(T) estimates also irrespective to the applied correction scheme for the aug-cc-pVDZ basis set. Note that this ordering is not found for the BSSE calculations on the water hexamer in the aug-cc-pVTZ basis set indicating that the described dependence is not generally valid.

The comparison of the SSFC vs. PAFC scheme leads to the following observations:

- the PAFC results tend to overestimate the BSSE with respect to the SSFC results,
- the overestimation grows when the cluster size increases, but not smoothly, and
- the overestimation ranges from 7% for the water hexamer to even 16% for octadecamer at the CCSD(T) level of theory.

A similar trend is also observed at the CCSD and MP2 level of theory, however both methods predict the overestimation to be a little smaller.

As one expects from theory the increase of the basis set quality from double to triple- ζ reduces the BSSE for the water hexamer. The BSSE correction calculated with the aug-cc-pVTZ basis set is about 50% of the BSSE correction with the aug-cc-pVDZ basis set for all correlation methods and BSSE correction schemes. The basis set increase also decreases the ratio of PAFC/SSFC (%) for the water hexamer at all theory levels. This can be rationalized by the fact that the BSSE disappears in the limit of the complete basis set.

As a final remark note that the blank entries in Table 3.3 occur because the calculation of particularly the VMFC(2) scheme becomes very quickly impossible to carry out especially for bigger clusters and more demanding correlation treatments. But also the calculation of the BSSE within the triple- ζ basis set (except for the smallest cluster or the PAFC scheme) was hardly possible to carry out due to computational limitations.

Table 3.3: Comparison of different BSSE correction schemes for $(\text{H}_2\text{O})_n$ $n=6,8,\dots,20$ ($n=6$ boat structure) clusters with aug-cc-pVDZ (and aug-cc-pVTZ for $n=6$) basis sets at the CCSD and MP2 level. BSSE in kcal/mol.

n	6	6 (X=T)	8	10	12	14	16	18	20
CCSD(T)/aug-cc-pVXZ									
SSFC	5.79	2.80	11.56	14.80	18.30	22.75	26.54	39.11	
PAFC	6.22	2.95	13.01	16.45	20.19	25.66	29.68	45.18	39.05
PAFC/SSFC (%)	107.4	105.4	112.5	111.1	110.3	112.8	111.8	115.5	
VMFC(1,2)	5.94	2.87	12.43	15.91	19.64				
VMFC(2)	0.14	0.07	0.88	1.10	1.34				
VMFC(2)/VMFC(1,2)(%)	2.36	2.44	7.08	6.91	6.82				
CCSD/aug-cc-pVXZ									
SSFC	5.20	2.64	10.45	13.39	16.55	20.63	24.05	35.50	31.55
PAFC	5.56	2.77	11.59	14.66	18.00	22.88	26.47	40.35	34.84
PAFC/SSFC (%)	106.9	104.9	110.8	109.5	108.8	110.9	110.1	113.7	110.4
VMFC(1,2)	5.27	2.68	11.17	14.29	17.65				
VMFC(2)	0.06	0.04	0.72	0.91	1.10				
VMFC(2)/VMFC(1,2)(%)	1.2	1.4	6.4	6.3	6.2				
MP2/aug-cc-pVXZ									
SSFC	5.40	2.85	10.84	13.89	17.18	21.39	24.95	36.96	32.72
PAFC	5.78	2.98	12.02	15.22	18.71	23.75	27.50	42.01	36.20
PAFC/SSFC (%)	106.9	104.8	110.9	109.6	108.9	111.1	110.2	113.7	110.6
VMFC(1,2)	5.51	2.91	11.48	14.70	18.16	22.96	26.69		
VMFC(2)	0.10	0.06	0.64	0.81	0.99	1.57	1.75		
VMFC(2)/VMFC(1,2)(%)	1.9	2.1	5.6	5.5	5.4	6.9	6.5		

3.3.2 Approximate SSFC(R) scheme applied to water clusters

The performance of the proposed approximate SSFC(R) scheme according to Eq. 3.22 has been investigated for the water cluster series with $n=6,8,\dots,18$ at the CCSD(T) with the aug-cc-pVXZ ($X=D,T$) basis set applied. As a benchmark for the accuracy of the SSFC(R) scheme the full SSFC correction according to Eq. 3.16 was calculated for the aug-cc-pVDZ basis set.

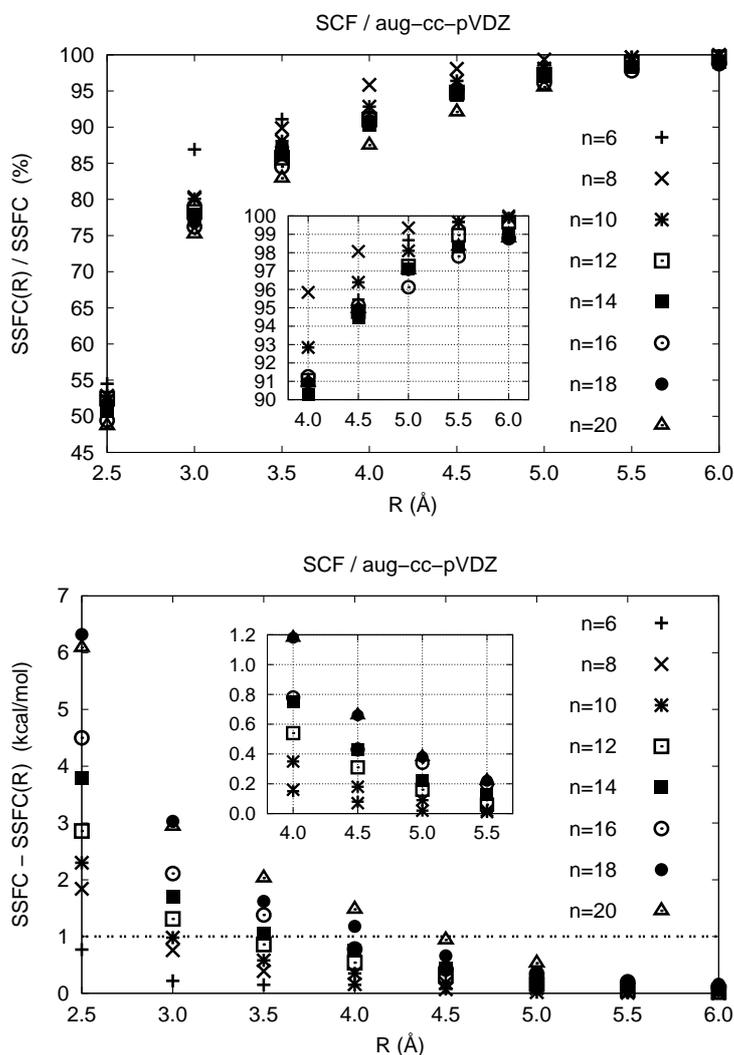


Figure 3.12: BSSE correction (in percentage) depending on the distance threshold R (Å) with respect to the SSFC method and the corresponding absolute aberration from SSFC calculated for the water cluster series $(\text{H}_2\text{O})_n$ ($n=6$ boat structure) at HF level of theory.

The plots in Figures 3.12 and 3.13 illustrate the percentage coverage of the CCSD(T) SSFC(R) BSSE correction with respect to full SSFC scheme as well as the absolute deviation from the full SSFC treatment depending on the distance threshold R . The corresponding CCSD/aug-cc-pVDZ and MP2/cc-pVDZ results for clusters with even 20 monomers are given in the the Appendix B in Fig. B.1-B.2. Since the convergence behavior of the MP2 and CCSD outcomes does not differ from the CCSD(T) results the analysis of the data may be limited to the CCSD(T) findings only. Whenever it will be necessary or worthwhile to consider also the MP2 and CCSD results we will incorporate them into the discussion. Otherwise we will list the data in the Appendix.

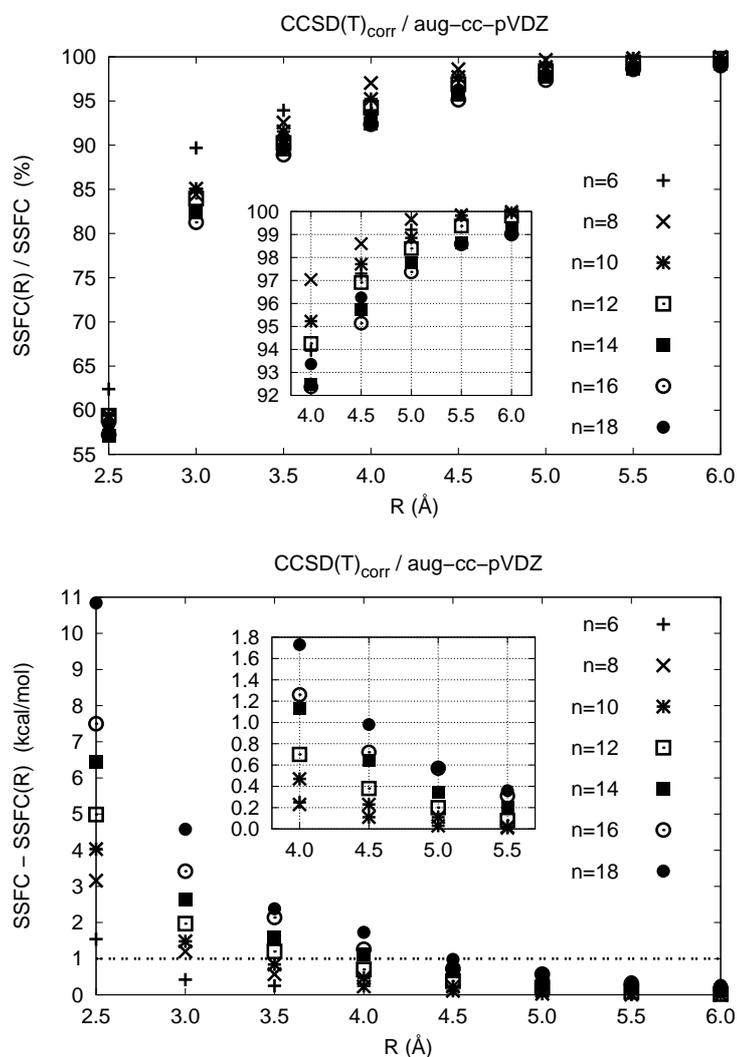


Figure 3.13: As Fig. 3.12 but for the correlation energy contribution at CCSD(T) level of theory.

The cutoff parameter R was chosen in a range between 2.5 Å and 6 Å. Correlation energies, labeled as $\text{CCSD(T)}_{\text{corr}}$, are presented separately from the HF contributions therein. The corresponding MP2 and CCSD plots are presented in the Appendix B in Figures B.1 and B.2 respectively.

The percentage coverage of the $\text{SSFC}(R)$ results with respect to SSFC grows exponentially when R increases at the HF and at the presented correlated levels of theory. Already at the middle range of the cutoff parameter with $R=3.5, 4$ and 4.5 Å very accurate results are observed. For example at $R=4$ and 4.5 Å the $\text{SSFC}(R)/\text{SSFC}$ ratio for both HF and $\text{CCSD(T)}_{\text{corr}}$ level is in between 90% to 97% and 95% to 99% respectively. A further extension of the cutoff threshold to 5.5 Å and 6 Å yields virtually exact results. For $R < 3.5$ Å however the percentage coverage becomes small and a significant loss in accuracy is observed. Exemplary the $\text{SSFC}(R=2.5)$ outcomes cover just 50% to 60% of the BSSE correction at the presented theory levels.

Regarding the absolute error of $\text{SSFC}(R)$ with respect to SSFC (refer to the second charts of Fig. 3.12 and Fig. 3.13) for the distance threshold of $R=4$ and $R=4.5$ Å deviations close to 1 kcal/mol or even smaller are gained. Negligible small errors of a few tenths of a kcal/mol are obtained for R values equal to 5 Å and 6 Å. For the smallest R value considered here with 2.5 Å the deviation is on the order of several kcal/mol. In the following the deviation of $\text{SSFC}(R)$ with respect to SSFC will be discussed together with the percentage BSSE correction with respect to the calculated interaction energies. This then allows to draw the most reliable conclusions regarding the accuracy of certain thresholds.

For a more detailed analysis of the results refer to the zoom-charts in the plots of Figures 3.12 and 3.13. They display a cutout covering a range of R in between 4 to 6 Å and 4 to 5.5 Å. Analyzing the results carefully we find that the accuracy of $\text{SSFC}(R)$ for a given R value depends on the cluster size following the trend that for smaller cluster sizes the results become more accurate as for the bigger ones. This is a consequence of the $\text{SSFC}(R)$ procedure itself since for small cluster sizes with a chosen R value a relatively large fraction of ghost atoms are included with respect to all available ones. This is not the case for larger clusters. In turn this causes a more complete basis set for the n single BSSE calculations with respect to the whole n -mer basis for smaller cluster sizes and is the reason why the percentage coverage of $\text{SSFC}(R)$ with respect to SSFC is bigger for small clusters.

In order to investigate this subject systematically the percentage coverage of the average number of basis functions (avg. no bf.) used to evaluate the second sum of Eq. 3.22 with respect to amount of basis functions used for SSFC calculations is plotted against the distance threshold R for the different cluster sizes n in Fig. 3.14.

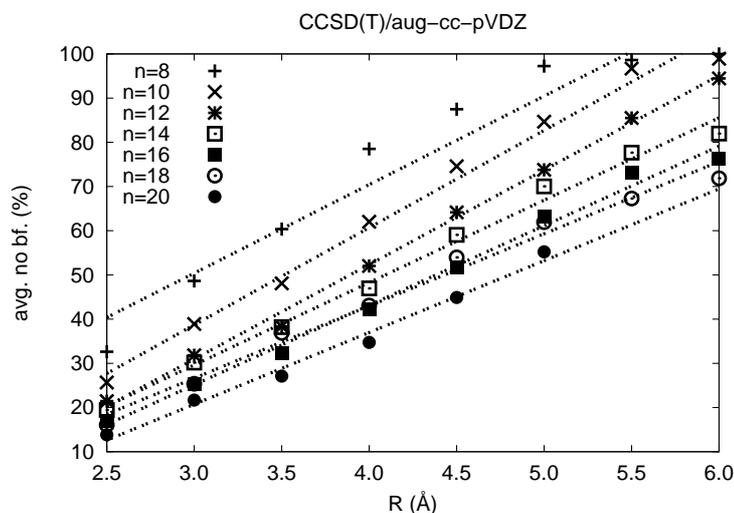


Figure 3.14: Average number of basis functions used in the SSFC(R) monomer calculations with respect to the SSFC method for the water cluster series $(\text{H}_2\text{O})_n$ $n=6,8,\dots,20$ ($n=6$ boat structure).

The avg. no bf. grows linearly with increasing R what clearly demonstrates that the approximate SSFC(R) procedure is carried out in a systematic fashion. Furthermore we note that for a given R the percentage coverage of the avg. no bf. decreases with increasing n . The findings in Fig. 3.14 may be used as a measure of the gained computational savings of the SSFC(R) scheme with respect to full SSFC method. Exemplary choosing SSFC($R=4$) as a sufficiently accurate approximation for a BSSE calculation we omit almost 50% of the basis functions in average, if we consider the middle range of n . The scaling behavior of CCSD(T), CCSD, MP2 and HF methods with respect to the size of the one-particle basis set is N^7 , N^6 , N^5 and N^4 , respectively. So if we halve the size of the basis set, what we do in average for SSFC($R=4$) method, we gain in theory a computational speed up by a factor of 128, 64, 32 and 16 for the CCSD(T), CCSD, MP2 and HF methods, respectively.

Another interesting observation is that the spread of the percentage coverage among the cluster sizes for fixed R values becomes smaller when R increases. We observe at $R=4.5$ Å a data spreading ranging from 94 to 99% whereas already at $R=5.5$ Å the data range varies only between 98 and 100%. This is a valuable hint indicating that if a certain amount of the nearest ghost atoms are employed in a single BSSE calculation $E_i^{\text{bb}_i}(\mathbf{R})$ one can neglect the basis functions which correspond to the omitted ghost atoms with almost no loss in accuracy. In order to support this conclusion Table 3.4 summarizes the CCSD(T)/aug-cc-pVDZ SSFC(R) BSSE correction for the $(\text{H}_2\text{O})_{n=14,16,18}$ water clusters with R values also above 6 Å. Therein also the average number of basis function is given.

Table 3.4: CCSD(T)/aug-cc-pVDZ SSFC(R) BSSE correction (in kcal/mol) with respect to R for three different water clusters $(\text{H}_2\text{O})_n$ and the deviation (in kcal/mol and %) from the full SSFC method. Avg. no bf. = average number of basis functions with respect to the full SSFC scheme.

R (Å)	5	5.5	6	7	8	∞
n=14						
SSFC(R)	22.19	22.42	22.56	22.70		22.75
error	0.56	0.33	0.19	0.05		0
error (%)	2.46	1.45	0.84	0.22		0
avg. no bf.	402	446	471	529		574
avg. no bf. (%)	70	78	82	92		100
n=16						
SSFC(R)	25.76	26.12	26.25	26.42		26.54
error	0.78	0.41	0.28	0.12		0
error (%)	2.94	1.56	1.07	0.45		0
avg. no bf.	415	479	501	554		656
avg. no bf. (%)	63	73	76	84		100
n=18						
SSFC(R)	38.14	38.54	38.70	38.95	39.05	39.11
error	0.97	0.58	0.42	0.16	0.06	0
error (%)	2.48	1.47	1.07	0.42	0.16	0
avg. no bf.	457	497	530	626	674	738
avg. no bf. (%)	62	67	72	85	91	100

The deviation of the SSFC($R \geq 5$) corrections for $n=14,16$ and 18 is very small with at most 3% error and definitely negligible because just in the order of few tenths of a kcal/mol for BSSE corrections obtained with $R > 6$ Å. From the data of Table 3.4 we may also conclude that a highly accurate BSSE correction obtained with $R \geq 6$ Å leads to a computational speed up. For distance threshold of 6 Å we neglect 18%, 24% and 26% basis functions in average for $n=14,16$ and 18 respectively. Whereas the usage of $R=7$ Å still leads to the omission 8%, 16% and 15% basis functions in average for $n=14,16$ and 18 respectively. And even for $R=8$ Å we neglect almost 10% of the basis functions on average.

The reason why we continuously consider the average number of basis function is due to the fact that the amount of basis functions exploited for a single BSSE calculation differs among the n monomer units. Recall that the number of ghost atoms covered by a certain R threshold depends upon the distances of all the coordinates from the considered monomer to all the remaining coordinates in a cluster. The crucial factor

influencing the amount of covered ghost atoms is therefore the geometrical arrangement of the neighbor monomers surrounding the considered monomer fragment. From these considerations follows that major differences with respect to the number of covered ghost atoms and therefore the size of applied basis set occur between monomer fragments from the central region in a cluster compared to those arranged at the periphery.

To confirm this considerations we study the variation of the average number of basis functions employed in the SSFC(R) calculation on the $(\text{H}_2\text{O})_{18}$ cluster. In Fig. 3.15 the standard deviation from the average number of basis functions is plotted against R.

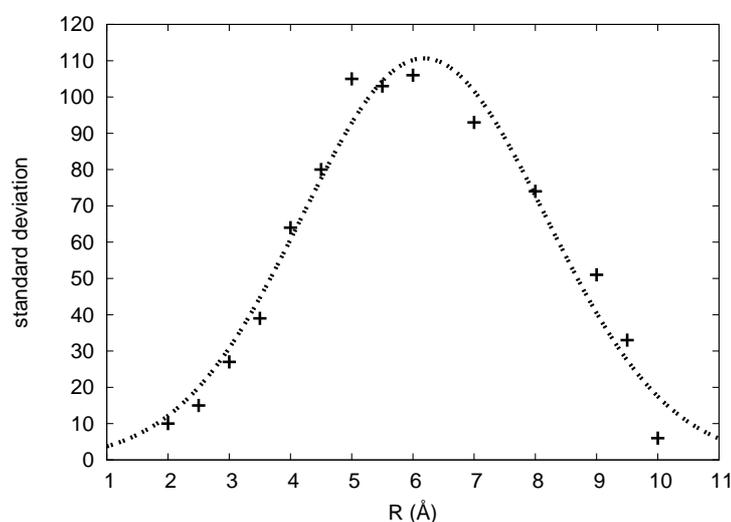


Figure 3.15: Standard deviation δ from the average number of basis functions needed for an approximate SSFC(R) calculation at a given R distance threshold for the water cluster $(\text{H}_2\text{O})_{18}$ obtained with the aug-cc-pVDZ basis set.

The curve passes the maximum at $R=6 \text{ \AA}$ and decays for smaller and greater R than 6 \AA equably and therefore may be fitted according to a Gaussian distribution. This behavior is reasonable since the number of ghost atoms covered by small R values are comparable for the n different monomers the standard deviation is therefore small. The situation is similar for bigger R values since in this case the greatest fraction of the ghost atoms employed to construct \mathbb{B}_i is already included and hence the standard deviation is also small.

In order to emphasize the computational savings we present in Fig. 3.16 a comparison between the calculation time of the proposed approximate SSFC(R) scheme at the CCSD(T)/aug-cc-pVDZ level of theory with respect to the full SSFC approach with respect to R for the water clusters with $n=8,10,\dots,18$. Note that the timings are obtained from calculations done on a PC cluster frequently used by other jobs and therefore the

results are meant to be rather a guide for the eye than a thorough investigation of the timings.

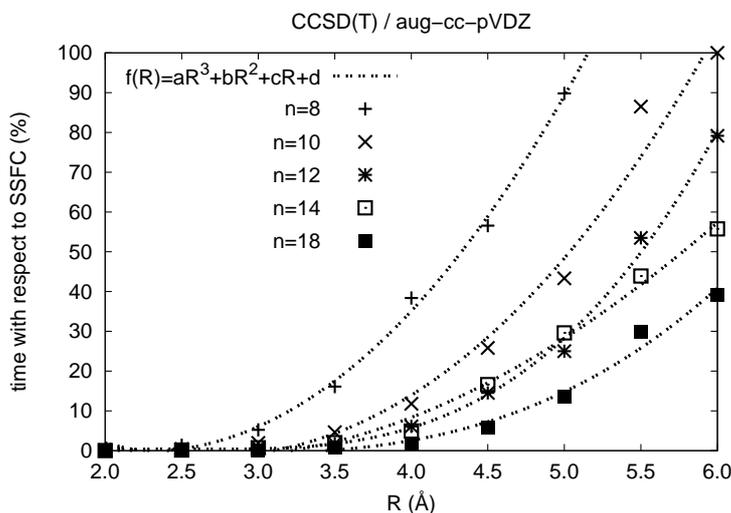


Figure 3.16: Relative computational time of SSFC(R) with respect to the SSFC scheme at CCSD(T)/aug-cc-pVDZ level for the $(\text{H}_2\text{O})_n$ $n=6,8,\dots,18$ water cluster series ($n=6$ boat structure).

For satisfactory accurate results at a cutoff parameter between 4 and 4.5 Å we need at most 15 to 40% of the computational time compared to the full SSFC scheme. The savings grow for larger cluster sizes at constant R values. Huge computational savings up to 80% are obtained for cluster sizes with $n > 10$, even when R is equal to 6 Å.

3.3.3 Approximate SSFC(R) scheme applied to methanol clusters

The approximate SSFC(R) scheme was also used to account for BSSE effect of a methanol cluster series $(\text{CH}_3\text{OH})_n$ with $n=8,12$ and 16 at the CCSD(T)/aug-cc-pVDZ level of theory. Table 3.5 summarizes the results of calculations which were still feasible to carry out. From the presented data we can firstly conclude that the SSFC(R) scheme performs also well for other types of clusters, as the already discussed water clusters. The percentage coverage of SSFC(R) with respect to SSFC follows also an exponential behavior when R increases. The employment of about 70% of the basis functions in average, for the calculation of SSFC(R=4.5) yields satisfactory accurate BSSE corrections with an error slightly above 1 kcal/mol for $n=8$.

Table 3.5: SSFC(R) CCSD(T)/aug-cc-pVDZ BSSE correction with respect to R for three different methanol clusters $(\text{CH}_3\text{OH})_n$ and the deviation (for $n=8$) from the full SSFC scheme. Avg. no bf. = average number of basis functions with respect the full SSFC scheme.

R Å	SSFC(R) kcal/mol	SSFC(R) %	error kcal/mol	avg. no bf.	
					%
n=8					
4	13.35	85.66	2.23	341	52
4.5	14.36	92.15	1.22	438	67
5	14.94	95.90	0.64	497	76
5.5	15.31	98.25	0.27	568	87
∞	15.58	100.00	0.00	656	100
n=12					
4	23.28			366	37
4.5	25.15			459	47
5	26.34			551	56
∞				984	100
n=16					
4	37.65			438	33
∞				1312	100

The important message so far is that we have an systematically improvable BSSE correction scheme which can be applied for different cluster types and we can control both, the level of desired accuracy and the computational savings. Moreover the scheme is also applicable to correct for BSSE artifacts where standard methods fail to proceed.

3.3.4 Basis set dependency on the approximate SSFC(R) scheme

The influence of the basis set quality on the accuracy of the SSFC(R) results has been investigated for a water hexamer (bag structure, Fig 3.35). Dunning's aug'-cc-pVXZ basis sets with cardinal numbers D,T,Q and 5 have been used. The prime indicates that the augmented functions have not been applied on the hydrogen atoms. This basis set has been frequently used in the computation of hydrogen bonded clusters [128, 129]. To obtain an estimate for the basis set limit we extrapolated the correlation energy according to Eq. 3.3 using the quadruple- ζ and quintuple- ζ basis sets. The CCSD(T) BSSE corrections obtained with the distance thresholds $R=2.5, 3.5, 4.5$ Å as well as for $R \rightarrow \infty$ and the corresponding BSSE corrected and uncorrected stabilization energies

are summarized in Table 3.6. The influence of the basis set size on the stabilization energy and on the BSSE correction at MP2 and CCSD level of theory, are shown in Appendix B (Table B.2).

Table 3.6: CCSD(T) BSSE corrected and uncorrected stabilization energies of the water cluster $(\text{H}_2\text{O})_6$ (bag structure) and BSSE corrections obtained with different schemes for the basis set series aug'-cc-pVXZ (X=D,T,Q,5). ($\Delta E^{\text{BSIE}} = D^{\text{CBS(Q,5)}} - D^{\text{aug'-cc-pVXZ}}$, all energies in kcal/mol, values in parentheses in %).

X	D	T	Q	5	CBS(Q,5)
D	-37.68	-39.15	-40.05	-40.08	-40.17
ΔE^{BSIE}	2.50	1.03	0.13	0.10	0
SSFC (R=2.5)	3.38 (61.50)	1.72 (69.40)	0.67 (68.28)	0.28	0
SSFC (R=3.5)	5.04 (91.19)	2.32 (93.47)	0.91 (93.21)	0.37	-0.1
SSFC (R=4.5)	5.50 (99.54)	2.47 (99.56)	0.97 (99.55)		
SSFC	5.52 (100)	2.48 (100)	0.98 (100)		
D(SSFC (R=2.5))	-34.30	-37.42	-39.38	-39.80	-40.20
D(SSFC (R=3.5))	-32.64	-36.83	-39.14	-39.71	-40.23
D(SSFC (R=4.5))	-32.18	-36.67	-39.07		
D(SSFC)	-32.16	-36.66	-39.07		

Stabilization energies D were calculated according to Eq. 3.9. The analysis of the data in Table 3.6 provides the encouraging observation that with the increase of the cardinal number X=D to T, the percentage coverage of the approximate SSFC(R) BSSE correction with respect to the SSFC scheme is growing. Exemplary the CCSD(T)/aug'-cc-pVDZ BSSE correction at R=2.5 Å covers 61.5% of the full SSFC BSSE treatment whereas with the aug'-cc-pVTZ basis set one obtains a percentage rate equal to 69.4%. A small improvement of the percentage coverage arising due to the increase from double- ζ to triple- ζ basis set is also observed when we analyze the results obtained with the 3.5 Å cutoff parameter whereas the outcomes of SSFC(R=4.5) stay rather stable among different cardinal numbers. Whereas the CCSD(T)/aug'-cc-pVXZ SSFC(R)/SSFC ratio for X=T and Q is almost equal.

The BSSE error is reduced approximately by a factor of 2.2 when going from double- ζ to triple- ζ and by a factor of 2.5 when increasing the basis sets from triple- ζ to quadruple- ζ and from quadruple- ζ to quintuple- ζ at CCSD(T) level of theory. The 4,5-extrapolation leads to correlation energies very close to the basis set limit as the

BSSE value is virtually zero for the SSFC(R=3.5,4.5) outcomes.

Considering the accuracy of the investigated R-dependent SSFC scheme we find highly accurate stabilization energies when the BSSE error is corrected with the SSFC(R=4.5) values, the difference between D(SSFC(R=4.5)) and D(SSFC) is negligible. Very good agreement with the full SSFC correction is observed for a distance threshold of 3.5 Å, with errors of just a few tenths of a kcal/mol.

In theory any BSSE correction method aims to cure the basis set imbalance in the evaluation of the interaction energy, but does not affect the so-called basis set incompleteness error (BSIE) which also occurs due to the usage of finite basis sets. In other words, the BSSE can be removed for any given basis set, whereas the BSIE can only be reduced when the basis set quality is increased. It is also clear that at the basis set limit both errors have to disappear. In Table 3.6 we present the BSIE as a difference between the uncorrected interaction energy D for aug'-cc-pVXZ with X=D,T,Q,5 and D at CBS(Q,5). For the CCSD(T) level we find a deviation from the CBS(Q,5) limit equal to 2.5, 1, 0.1 and 0.1 kcal/mol for cardinal numbers X=D,T,Q and 5 respectively. One should keep in mind the magnitude of the BSIE for a given X when deciding on how accurate an approximate BSSE calculation should be carried out. We believe that it is reasonable to account just for 92% BSSE correction as we do when applying D(SSFC(R=3.5)) at the CCSD(T) level causing only a few tenths of a kcal/mol deviation from the full correction scheme when at the same time the BSIE error is of the order of several kcal/mol.

Another important advantage of the proposed SSFC(R) approach is that one can carry out BSSE correction calculations which are no longer manageable with the standard procedure. The D(SSFC(R=4.5)) calculation for the aug'-cc-pV5Z basis set was not feasible anymore with our standard cluster nodes due to insufficient memory available. The calculated BSSE corrections at R=3.5 Å are therefore the best estimates for the large aug'-cc-pV5Z basis set. Since we observed a very good agreement for the full SSFC scheme for the basis sets with cardinal numbers X=T and Q, we may also regard the D(SSFC(R=3.5)) aug'-cc-pV5Z outcomes as highly accurate.

3.3.5 SSFC(R) corrected stabilization energies of the water cluster series

In order to investigate the magnitude of the BSSE correction on the different water cluster sizes the stabilization energies (via Eq. 3.9) of the series were also calculated. Correlation energies of the clusters have been calculated within the incremental scheme at the CCSD(T), CCSD and MP2 level of theory and with the usage of Dunning's augmented correlation consistent polarized valence double- and triple- ζ basis sets [67,68].

The incremental correlation energies at various correlation levels are listed in Table 3.7 and 3.8. The errors among the incremental CCSD(T), CCSD and MP2 results are comparable among each other. Once the computational less demanding MP2 reference calculation is used to check the accuracy of the incremental MP2 expansion, we may expect a similar accuracy also for the CCSD and CCSD(T) correlation level. Exemplary the aug-cc-pVDZ/MP2_{corr}(3) errors with respect to the canonical calculation differ from the aug-cc-pVDZ/CCSD_{corr}(3) ones for n=12,14,16 and 18 by at most 0.05 kcal/mol. This is important since not every CCSD(T) or even CCSD reference calculation could be carried out especially for larger cluster sizes. We note that very good agreement with the canonical MP2 results is already achieved with a third-order expansion with aberration around or smaller than 1 kcal/mol. For the smallest water cluster a second-order expansion yields sufficiently accurate results with errors of only a few tenths of a kcal/mol. Considering the accuracy of the results obtained with different basis set sizes we find that for larger basis sizes the results become even more accurate. For the bag and boat water hexamer clusters at all correlation levels the second-order expansion using a triple- ζ basis set is by an order of magnitude more accurate than the expansion for the double- ζ basis set.

Table 3.7: CCSD(T), CCSD and MP2 incremental correlation energies for the water cluster series $(\text{H}_2\text{O})_n$ compared to the canonical results for the aug-cc-pVXZ (X=D,T) and aug'-cc-pVXZ (X=D,T,Q,5) basis sets. ($t_{\text{con}}=3$ Bohr; dsp=4; core=n; $t_{\text{main}}=3$ Bohr; environment basis: H=STO-3G, O=6-31G; Order-dependent distance screening according to Eq. 2.66)

n	E_{thres}	X	f	\mathcal{O}	CCSD(T) _{corr} (i)		CCSD _{corr} (i)		MP2 _{corr} (i)			
	au				Bohr	au	kcal/mol	au	kcal/mol	au	kcal/mol	
6 bag	1×10^{-6}	D'	30	2	-1.391342	-0.20	-1.359579	-0.20	-1.314954	-0.20		
				3	-1.390816	0.13	-1.359109	0.10	-1.314485	0.10		
				T'	30	2	-1.694962	0.004	-1.641703	-0.03	-1.614552	-0.01
						3	-1.694966	0.002	-1.641653	0.00	-1.614521	0.01
		Q'	30	2	-1.793864		-1.735815		-1.722963			
					3	-1.793895		-1.735781		-1.722976		
				5'	30	2	-1.827250		-1.767427		-1.765951	
6 boat	1×10^{-7}	D	∞	2	-1.400281	-0.18	-1.367002	-0.16	-1.322291	-0.21		
				3	-1.399988	0.01	-1.366738	0.00	-1.321954	0.01		
				T	∞	2	-1.697929	-0.02	-1.644589	-0.04	-1.617694	-0.07
						3	-1.697876	0.01	-1.644519	0.00	-1.617565	0.01
		Q	∞	2	-1.793041		-1.735304		-1.722931			
					3	-1.793041		-1.735268		-1.722865		
				D	30	2	-1.881120		-1.834522	-0.94	-1.776090	-0.89
						3	-1.879340		-1.832804	0.14	-1.774450	0.14
T	30	2	-2.278583		-2.205291		-2.170619					
		3	-2.277865		-2.204531		-2.169990					
10	1×10^{-7}	D	30	2	-2.351846	-1.19	-2.293460	-1.14	-2.220585	-1.12		
				3	-2.349557	0.25	-2.291314	0.20	-2.218465	0.21		
				T	30	2	-2.848559		-2.756865		-2.713687	-0.44
						3	-2.847519		-2.755795		-2.712699	0.18

Table 3.8: As Table 3.7, but for the water cluster series $(\text{H}_2\text{O})_n$ $n=10,12,\dots,20$.

n	E_{thres}	X	f	\mathcal{O}	CCSD(T) _{corr(i)}		CCSD _{corr(i)}		MP2 _{corr(i)}	
	au				au	error	au	error	au	error
			Bohr			kcal/mol		kcal/mol		kcal/mol
12	1×10^{-6}	D	∞	2	-2.823154		-2.752911	-1.31	-2.665757	-1.30
				3	-2.820222		-2.750138	0.43	-2.663079	0.38
		T	∞	2	-3.419197		-3.309014		-3.257480	
				3	-3.417972		-3.307739		-3.256320	
14	1×10^{-7}	D	30	2	-3.299746		-3.216932	-1.97	-3.115388	-1.90
				3	-3.295847		-3.213199	0.37	-3.111716	0.40
		T		2	-3.995409		-3.865971		-3.806086	
				3	-3.993578		-3.864071		-3.804281	
16	1×10^{-7}	D	40	2	-3.772078		-3.677230		-3.561605	-2.11
				3	-3.766828		-3.672258		-3.556782	0.92
				4	-3.768762		-3.673904		-3.558486	-0.15
16	1×10^{-7}	T	40	2	-4.566770		-4.418721		-4.350688	
				3	-4.564910		-4.416765		-4.348796	
18	1×10^{-7}	D	30/40/40	2	-4.298669		-4.184566	-4.04	-4.062820	-3.68
				3	-4.291131		-4.176431	1.07	-4.055221	1.09
		T	25	2	-5.195707		-5.021122		-4.953708	
				3	-5.190441		-5.015435		-4.948531	
20	1×10^{-7}	D	30/40/40	2	-4.719624		-4.600392		-4.455943	
				3	-4.713477		-4.594102		-4.449776	

In Table 3.9 we present the CCSD(T)/aug-cc-VXZ (X=D,T and CBS(D,T)) BSSE corrected and uncorrected stabilization energies, as well as the BSSE error obtained according to Eq. 3.22 with a distance threshold of R=4 and 4.5 Å. For the aug-cc-pVDZ basis set the full SSFC BSSE correction is also presented in Table 3.9. The corresponding CCSD and MP2 (with n up to 20) results are presented in the Appendix B in Tables B.3-B.6.

Table 3.9: SSFC, SSFC(R=4,4.5) BSSE corrections and BSSE corrected and uncorrected stabilization energies for the water cluster series $(\text{H}_2\text{O})_n$ (n=6,8,10 and 12) at the CCSD(T)/aug-cc-VXZ (X=D,T and CBS(D,T)) level, all energies in kcal/mol, values in parentheses in %.

n	6	8	10	12
aug-cc-pVDZ				
D	-37.58	-65.95	-84.99	-104.71
SSFC(R=4)	5.39 (93.11)	11.17 (96.64)	13.98 (94.44)	17.06 (93.21)
SSFC(R=4.5)	5.60 (96.76)	11.37 (98.43)	14.40 (97.28)	17.61 (96.21)
SSFC	5.79 (100)	11.56 (100)	14.80 (100)	18.30 (100)
D(SSFC(R=4))	-32.19 (16.75)	-54.57 (20.47)	-71.01 (19.69)	-87.65 (19.46)
D(SSFC(R=4.5))	-31.98 (17.52)	-54.44 (20.89)	-70.59 (20.40)	-87.10 (20.21)
D(SSFC)	-31.79 (18.22)	-52.94 (21.83)	-70.19 (21.09)	-86.41 (21.18)
aug-cc-pVTZ				
D	-39.16	-67.39	-86.71	-106.82
SSFC(R=4)	2.64	5.48	6.92	8.43
SSFC(R=4.5)	2.73	5.58	7.10	8.68
D(SSFC(R=4))	-36.53 (7.22)	-61.91 (8.86)	-79.80 (8.67)	-98.39 (8.57)
D(SSFC(R=4.5))	-36.43 (7.50)	-61.82 (9.02)	-79.61 (8.91)	-98.14 (8.82)
CBS(D,T)				
D	-39.75	-68.58	-88.14	-108.58
SSFC(R=4)	1.95	4.23	5.34	6.50
SSFC(R=4.5)	2.02	4.30	5.48	6.69
D(SSFC(R=4))	-37.80 (5.15)	-64.35 (6.58)	-82.79 (6.45)	-102.08 (6.37)
D(SSFC(R=4.5))	-37.74 (5.35)	-64.27 (6.70)	-82.66 (6.62)	-101.90 (6.55)

Table 3.10: As Table 3.9, but for n=14,16 and 18.

n	14	16	18
aug-cc-pVDZ			
D	-128.53	-149.24	-184.47
SSFC(R=4)	20.87 (91.73)	24.29 (91.54)	36.20 (92.55)
SSFC(R=4.5)	21.68 (95.31)	25.12 (94.65)	37.48 (95.82)
SSFC	22.75 (100)	26.54 (100)	39.11(100)
D(SSFC(R=4))	-107.66 (19.38)	-124.95 (19.44)	-148.27 (24.42)
D(SSFC(R=4.5))	-106.85 (20.29)	-124.12 (20.24)	-146.99 (25.50)
D(SSFC)	-105.78 (20.98)	-122.70 (21.63)	-146.32 (26.73)
aug-cc-pVTZ			
D	-130.48	-151.91	-185.55
SSFC(R=4)	10.28	11.94	
SSFC(R=4.5)			
D(SSFC(R=4))	-120.19 (8.56)	-139.97 (8.53)	
D(SSFC(R=4.5))			
CBS(D,T)			
D	-132.63	-154.55	-188.91
SSFC(R=4)	8.00	9.24	
SSFC(R=4.5)			
D(SSFC(R=4))	-124.63 (6.42)	-145.31 (6.36)	
D(SSFC(R=4.5))			

Stabilization energies exhibit large BSSE effects when the aug-cc-pVDZ basis set is employed in the calculations. The CCSD(T) BSSE corrections obtained with the SSFC approach range in between 6 and almost 40 kcal/mol. The percentage BSSE correction with respect to D(SSFC) is in between 18 to 27% for the aug-cc-pVDZ basis set. Regarding the water clusters ranging in between n=8,10,...,16 the BSSE fraction of D(SSFC) is rather stable varying around 20%. For all investigated cluster sizes the BSSE correction is reduced by about 50% when increasing the basis set from double- ζ to triple- ζ using the approximate SSFC correction with R=4 and 4.5 Å. The employment of the 2,3-extrapolation further reduces the BSSE correction. The failure to account for the BSSE effect with the aug-cc-pVTZ basis set would lead to an overestimation of the stabilization energies by about 7 to 11% which is still considerably large and the magnitude of the BSSE generated with the aug-cc-pVTZ basis set is still between 3 and 18 kcal/mol. We note that the calculation of the BSSE within the whole n-mer basis set for single ghost calculations at aug-cc-pVTZ level was not feasible for

all investigated cluster sizes, even the less demanding calculation with $R=4.5 \text{ \AA}$ was difficult to carry out for larger cluster sizes.

Considering the accuracy of the truncated BSSE correction scheme we propose the usage of $D(SSFC(R=4))$ for the water clusters. Note that the usage of larger distance parameters for larger cluster sizes is hardly manageable for the more accurate aug-cc-pVTZ basis set. Moreover, as already pointed out in Section 3.3.4 we may even expect the results of $D(SSFC(R=4))$ to be more accurate when calculated with the aug-cc-pVTZ compared to the aug-cc-pVDZ basis set. Therefore we recommend the approximate $D(SSFC(R=4))$ BSSE correction as the optimum between a reasonable good accuracy and the saved computational time.

We emphasize that the given recommendation is valid for water clusters. In a study on BSSE corrected interaction energies using the distance approximation for other systems, one should verify the accuracy with respect to R using a small basis set first.

3.3.6 Approximate VMFC(2)(R) scheme

The accuracy of the proposed BSSE second-order correction VMFC(2)(R) with respect to R has been tested for the water clusters with $n=8,10$ and 12 at the CCSD(T)/aug-cc-pVDZ level of theory. The findings are summarized in Table 3.11. The MP2/aug-cc-pVDZ results are also presented in the chart of Fig. 3.17 since it was possible to carry out the full VMFC(2) calculation for cluster sizes up to $n=16$.

Table 3.11: Approximate second-order BSSE correction VMFC(2)(R) with respect to different R distance thresholds and the deviation from the VMFC(2) scheme for the water cluster series $(\text{H}_2\text{O})_n$ at the CCSD(T)/aug-cc-pVDZ level of theory.

R	VMFC(2)(R)		error	VMFC(2)(R)		error	VMFC(2)(R)		error
Å	kcal/mol	%	kcal/mol	kcal/mol	%	kcal/mol	kcal/mol	%	kcal/mol
	n=8			n=10			n=12		
2	0.23	25.85	0.65	0.40	36.19	0.71	0.50	37.73	0.83
2.5	0.64	72.45	0.24	0.78	70.41	0.33	0.92	69.22	0.41
3	0.77	87.50	0.11	0.97	87.48	0.14	1.14	85.67	0.19
3.5	0.83	94.80	0.05	1.02	92.42	0.08	1.21	90.38	0.13
4	0.87	98.57	0.01	1.07	96.43	0.04	1.27	94.84	0.07
4.5	0.87	99.59	0.004	1.08	98.16	0.02	1.30	97.24	0.04
5	0.88	99.79	0.002	1.10	99.51	0.01	1.32	98.65	0.02
5.5	0.88	100	0	1.10	99.93	0.001	1.33	99.56	0.01
6				1.10	100	0.00	1.33	99.87	0.002
∞	0.88	100	0	1.10	100	0	1.34	100	0

The chart in Fig. 3.17 and the percentage ratio presented in Table 3.11 clearly show that the convergence behavior of VMFC(2)(R) follows an exponential distribution with respect to R similarly to the SSFC(R) scheme. Highly accurate BSSE corrections with a percentage coverage ranging from 90 to 95% for MP2 and CCSD(T) level of theory for VMFC(R=3.5)(2) are achieved. Very small deviations of at most 0.1 kcal/mol occur for CCSD(T)/aug-cc-pVDZ second-order BSSE corrections with distance threshold of 3.5 Å (refer to Table 3.11). Interestingly the results only slightly worsen for a given R when n increases.

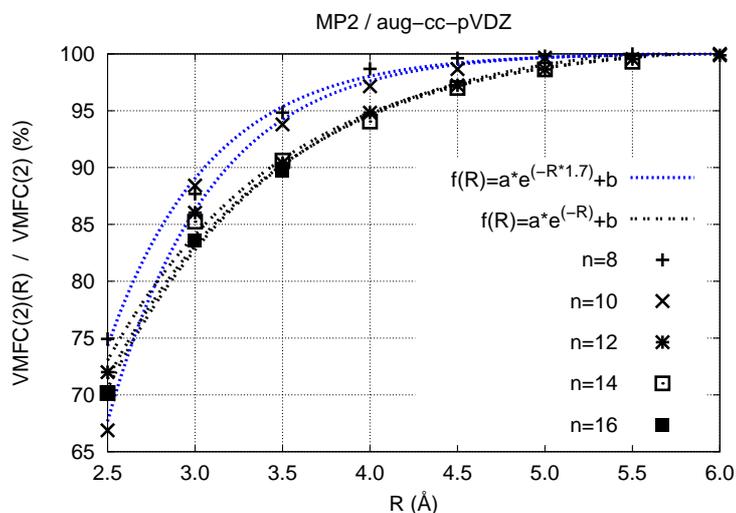


Figure 3.17: Approximate BSSE second-order VMFC correction (percentage) depending on distance threshold R (Å) with respect to the full VMFC approach at the MP2 level for the $(\text{H}_2\text{O})_n$ $n=8,10,\dots,16$ water cluster series.

The computational savings for the CCSD(T)/aug-cc-pVDZ calculations of VMFC(2)(R) with respect to VMFC(2) are presented in Fig. 3.18.

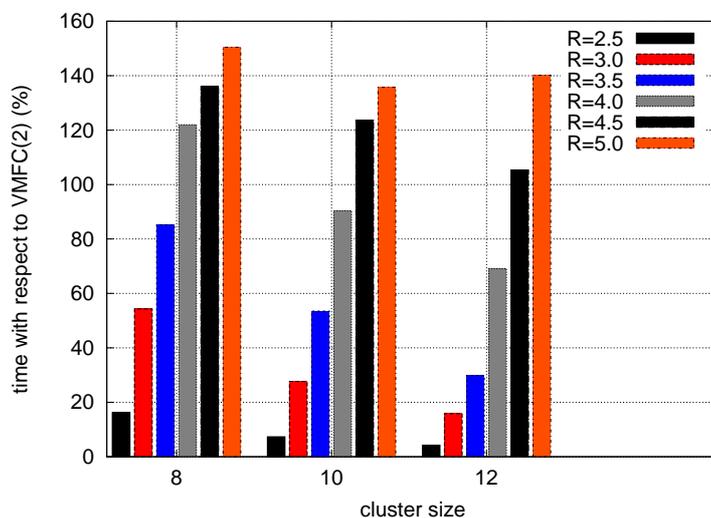


Figure 3.18: Relative computational time of VMFC(2)(R) with respect to the VMFC(2) scheme for the CCSD(T)/aug-cc-pVDZ method for the $(\text{H}_2\text{O})_n$ $n=6,8,12$ water cluster series.

For the highly accurate results obtained with a cutoff radius equal to 3.5 Å we save 15%, almost 50% and 70% computational time with respect to the full second-order

BSSE treatment for the water clusters with $n=8,10$ and 12 , respectively. The results confirm the trends already observed for the approximate first-order BSSE calculation, namely that the computational savings grow with increasing cluster size.

However note that the calculation of VMFC(2)($R \geq 4.5$) becomes slower than the full one. This is due to a different number of individual calculations which have to be performed for the VMFC(2) scheme in comparison to the approximate one. For each n -mer cluster one needs $6 \binom{n}{2}$ energy calculations in order to evaluate Eq. 3.19. We rewrite $6 \binom{n}{2}$ into the sum $4 \binom{n}{2} + 2 \binom{n}{2}$ and assign the $2 \binom{n}{2}$ energy contributions to the last two terms of Eq. 3.19, namely the monomer calculations in the whole n -mer basis set: $E_i^{ijk\dots n}$ and $E_j^{ijk\dots n}$. Obviously one can distinguish n different contributions out of $2 \binom{n}{2}$ and construct from n all the necessary $2 \binom{n}{2}$ energy terms. If no symmetry elements of the molecule itself are taken into account, the total number of individual calculations is therefore equal to $4 \binom{n}{2} + n$. Within the approximate VMFC(2)(R) scheme calculated according to Eq. 3.24 we cannot reuse these calculations, because the n -mer basis set is no longer the whole one but a truncated basis set and therefore different for every monomer fragment and also different for an union of basis functions of two monomer fragments (provided that the sum over i and j is restricted to $i < j$ as requested in Eq. 3.24). But even if we have to perform altogether $3n^2 - 3n$ instead of $2n^2 - n$ calculations we want to emphasize that the presented implementation is still already more efficient than the full VMFC(2) scheme for distance thresholds below 4 \AA exhibiting at the same time very accurate BSSE corrections.

3.3.7 Approximate $SSFC_{inc}$ scheme

The accuracy of the incremental BSSE correction scheme $SSFC_{inc}$ calculated according to Eq. 3.26 was firstly analyzed for the smallest water cluster investigated in this study namely the boat structure of the water hexamer. In Table 3.12 the MP2, CCSD and CCSD(T) BSSE corrections for the basis set series aug-cc-pVXZ (X=D,T and Q) are presented. The $SSFC_{inc}$ BSSE with respect to the expansion order i as well as the percentage coverage of $SSFC_{inc}$ with respect to SSFC are given in Table 3.12.

As already pointed out in Section 3.2.3 the first-order of the $SSFC_{inc}$ scheme is equivalent to the PAFC approach and as predicted from theory (refer to Eq. 3.31) the $SSFC_{inc}$ at highest possible expansion order i is equal the SSFC scheme. That is why $SSFC_{inc}$ BSSE correction covers exactly the SSFC value at the fifth expansion order for the $(H_2O)_6$ cluster.

Regarding the convergence behavior of the $SSFC_{inc}$ scheme with respect to i almost no difference is observed for the different correlation treatments. The basis set change from X=D to T and from X=T to Q affects the accuracy of the MP2/aug-cc-pVXZ results only slightly. As already discussed in Section 3.3.1 the increase in basis set quality from double to triple- ζ halves the BSSE irrespective to the applied correlation level when the SSFC and PAFC schemes are considered. An increase in cardinal number X from T to Q (refer to results in Table 3.12) has a similar influence on the BSSE at MP2 level of theory and causes almost 60% reduction of the CCSD(T) and CCSD BSSE values obtained with the PAFC approach.

Considering the accuracy of the approximate $SSFC_{inc}$ scheme we find that the addition of the fragmental second-order BSSE corrections to the first-order yields $SSFC_{inc}$ values which are in excellent agreement with the SSFC scheme for all presented levels of theory. The aberration of $SSFC_{inc}(i=2)$ with respect to SSFC is at most 0.04 kcal/mol and hence negligibly small. Therefore we may conclude that for the $(H_2O)_6$ cluster the $SSFC_{inc}$ scheme works very good.

To draw a general conclusion we need to broaden the data set. We observed the convergence behavior to be similar for CC and MP theory, therefore the $SSFC_{inc}$ correction scheme at the MP2/aug-cc-pVDZ level of theory was applied to account for BSSE effects of the total set of $(H_2O)_n$ water clusters ($n=8,10,\dots,20$).

The percentage coverage of $SSFC_{inc}$ scheme with respect to SSFC as a function of the expansion order is presented in Fig. 3.19. For the water clusters with $n=8$ and 10 the expansion is carried to the highest order. For the $(H_2O)_{12}$ water cluster the $SSFC_{inc}$ is calculated up to the fourth-order and the fragmental BSSE expansion for the remaining water clusters is truncated after the third-order.

Table 3.12: Incremental $SSFC_{inc}$ BSSE correction as a function of the expansion order i with respect to the SSFC scheme for the water hexamer (boat structure) at various levels of theory.

	i-th order corr. i	$SSFC_{inc}$ kcal/mol	$\frac{SSFC_{inc}}{SSFC}$ %	i-th order corr. kcal/mol	$SSFC_{inc}$ kcal/mol	$\frac{SSFC_{inc}}{SSFC}$ %	i-th order corr. kcal/mol	$SSFC_{inc}$ kcal/mol	$\frac{SSFC_{inc}}{SSFC}$ %	
MP2		aug-cc-pVDZ			aug-cc-pVTZ			aug-cc-pVQZ		
	1	5.78	106.92	2.98	2.98	104.78	1.50	1.50	107.26	
	2	-0.39	99.71	-0.17	2.81	98.72	-0.11	1.39	99.35	
	3	-0.01	99.56	0.05	2.86	100.51	0.02	1.41	100.50	
	4	0.03	100.14	-0.02	2.84	99.88				
	5	-0.01	100.00	0.00	2.85	100.00				
CCSD										
	1	5.56	106.89	2.77	2.77	104.90	1.17	1.17		
	2	-0.36	99.91	-0.15	2.61	99.03	-0.09	1.08		
	3	-0.02	99.47	0.04	2.65	100.45	0.01	1.08		
	4	0.04	100.15	-0.02	2.63	99.87				
	5	-0.01	100	0.00	2.64	100				
CCSD(T)										
	1	6.22	107.41	2.95	2.95	105.40	1.25	1.25		
	2	-0.44	99.78	-0.18	2.77	98.84	-0.11	1.14		
	3	-0.02	99.44	0.05	2.82	100.48	0.01	1.16		
	4	0.04	100.17	-0.02	2.80	99.88				
	5	-0.01	100	0.00	2.80	100				

Considering the first expansion order we find that the percentage coverage of $SSFC_{inc}$ is very similar for the different cluster sizes varying around 110% with respect to SSFC. At the second-order expansion the $SSFC_{inc}$ scheme underestimates the SSFC approach by up to 10% for different n . The situation changes dramatically when we refer to the third expansion order, for which a large data spreading for the different cluster sizes is observed. At $i=3$ the deviation from the SSFC scheme grows constantly with the increase of the cluster size. The $SSFC_{inc}(i=3)$ approach overestimates the SSFC BSSE corrections for the $(H_2O)_8$ cluster by 9% but already for $n=14$ the overestimation equals 20% and for the $(H_2O)_{20}$ cluster almost 40% overestimation is reached.

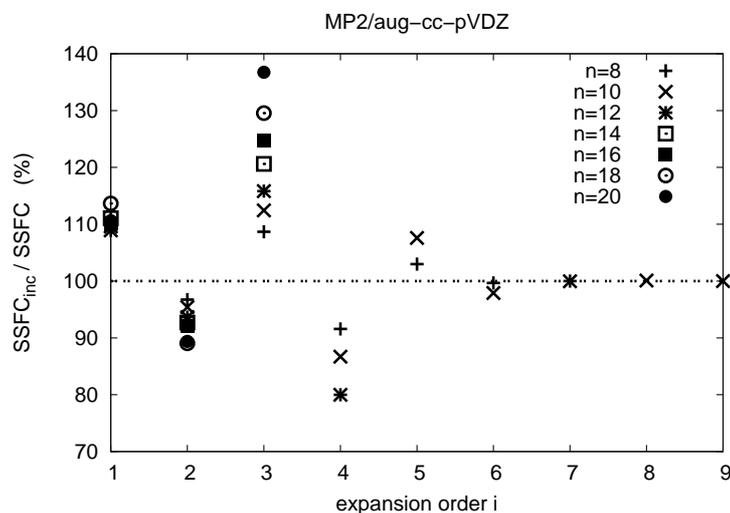


Figure 3.19: MP2/aug-cc-pVDZ $SSFC_{inc}$ correction as a function of the expansion order i with respect to SSFC in percentage for the water cluster series $(H_2O)_n$.

Note that at the fourth expansion order the fragmental BSSE correction could not be calculated for the whole cluster series, mostly because the calculation wall time exceeded the available time limit. Exemplary at the SuGi (Sustainable Grid Infrastructure) Cluster from the the University of Cologne the limitation with respect to the calculation time for a single node is limited to 30 days. But already the calculation time of the $SSFC_{inc}(i=4)$ correction for the $(H_2O)_{12}$ cluster took 13 days and even 19 days were necessary to account for $SSFC_{inc}(i=3)$ correction for $(H_2O)_{20}$ cluster. Since the calculation time is extremely long, we emphasize once again the importance of a low order truncation as a basic prerequisite for an efficient approximation of the $SSFC_{inc}$ scheme.

Nevertheless the data set is not complete at the fourth expansion order, from the avail-

able results we may conclude that the data spreading observed at the third-order is also present at the fourth-order. The $SSFC_{inc}(i=4)$ values underestimate the SSFC scheme and we also expect a strong deviation of $SSFC_{inc}/SSFC$ with respect to growing cluster sizes, as already observed for cluster sizes 6, 8, 10 and 12.

These observations lead to the conclusion that the fragmental BSSE correction diverges with the increase of the cluster size n . The results also show that the series in fact converges at high expansion order, refer to the examples with $n=6,8$ and 10, but unfortunately satisfactory accurate results are not achieved with a low order truncation. Table 3.13 summarizes the absolute i -th order $SSFC_{inc}$ correction (in kcal/mol) as well as the reference MP2/aug-cc-pVDZ SSFC correction for the cluster sizes considered in the chart of Fig. 3.19. The oscillation of the i -th order correction around the reference SSFC value is unacceptable high and does not terminate even before the third expansion order is reached. For smaller clusters, like $(H_2O)_8$ and $(H_2O)_{10}$, the $SSFC_{inc}$ series must be even calculated to at least 6-th and 7-th order in order to obtain highly accurate BSSE estimates.

Table 3.13: Incremental $SSFC_{inc}$ BSSE correction with respect to the expansion order i for the water cluster series $(H_2O)_n$ at MP2/aug-cc-pVDZ level of theory.

i	i-th order correction kcal/mol						
n	8	10	12	14	16	18	20
PAFC	12.02	15.22	18.71	23.75	27.50	42.01	36.20
2	-1.53	-1.97	-2.62	-3.93	-4.53	-9.10	-6.95
3	1.29	2.36	3.81	5.97	8.15	14.97	15.50
4	-1.85	-3.57	-6.16				
5	1.24	2.90					
6	-0.36	-1.35					
7	0.04	0.29					
8		0.02					
9		-0.01					
SSFC	10.84	13.89	17.18	21.39	24.95	36.96	32.72

The unfavorable convergence behavior to a certain extent may occur due to numerical noises which in turn may be caused if a huge number of negligibly small BSSE corrections are summed up. In order to gain insight into this issue, calculations with more tight convergence thresholds than the default settings in the MOLPRO program have been performed for the clusters, $n=8$ and 10 at MP2 and CCSD/aug-cc-pVDZ level

of theory. The accuracy of the SSFC_{inc} expansion with respect to default and tight convergence thresholds for the SCF energy, density matrix and CCSD energy and coefficient threshold is summarized in Table 3.14 and Table 3.15 for $(\text{H}_2\text{O})_{10}$ and $(\text{H}_2\text{O})_8$ respectively, (the values of the thresholds are listed on top of each Table). In Table 3.15 the series has also been investigated with the Pople-type [126, 130] polarized valence double- ζ basis set, with the addition of diffuse functions, 6-31++G** at MP2 level of theory.

The convergence behavior is indeed affected by the convergence thresholds but the influence is rather small. The MP2/aug-cc-pVDZ^b second- and third-order BSSE correction for $n=10$ are slightly decreased with respect to MP2/aug-cc-pVDZ^a and similar findings are observed also for $n=8$ where up to the 7-th expansion order every i -th order correction is found to be smaller by a few tenths of kcal/mol when the tight convergence thresholds are utilized in the calculation. The impact is smaller when we compare the MP2/6-31++G** results for $n=8$ (Table 3.15) and also at the CCSD level for $n=10$ (refer to Table 3.14). In general we find that the incremental BSSE series is not very sensitive with the respect to more tight convergence thresholds. No significant improvement of the convergence properties of the SSFC_{inc} series could be achieved with tight convergence thresholds, at least not for the two studied water clusters $(\text{H}_2\text{O})_{n=8,10}$.

Table 3.14: Incremental SSFC_{inc} BSSE correction as a function of the expansion order *i* with respect to the SSFC scheme for the (H₂O)₁₀ water cluster at different level of theory. Comparison of the influence of default^a and tight convergence thresholds on the accuracy.

MOLPRO thresholds: SCF energy: (default^a 1.00E-07 and tight^b 1.00E-11); density matrix (default^a 1.00E-05 and tight^b 1.00E-06); CCSD energy (default^a 1.00D-06 and tight^b 1.00D-08); CCSD coefficient (default^a 1.00D-4 and tight^b 1.00D-05).

	i-th order corr.	SSFC _{inc}	$\frac{\text{SSFC}_{\text{inc}}}{\text{SSFC}}$	i-th order corr.	SSFC _{inc}	$\frac{\text{SSFC}_{\text{inc}}}{\text{SSFC}}$	i-th order corr.	SSFC _{inc}	$\frac{\text{SSFC}_{\text{inc}}}{\text{SSFC}}$
	<i>i</i>	kcal/mol	%	kcal/mol	kcal/mol	%	kcal/mol	kcal/mol	%
aug-cc-pVDZ ^a									
					MP2		CCSD	CCSD(T)	
	1	15.22	109.59	14.66	15.22	109.48	16.45	16.45	111.13
	2	-1.97	95.40	-1.81	13.25	95.92	-2.35	14.10	95.23
	3	2.36	112.43	2.07	15.61	111.38	2.41	16.51	111.53
	4	-3.57	86.70	-3.11	12.04	88.14	-3.50	13.01	87.88
	5	2.90	107.59	2.45	14.94	106.40	2.78	15.79	106.67
	6	-1.35	97.90	-1.06	13.60	98.50	-1.23	14.55	98.33
	7	0.29	99.96	0.17	13.88	99.78	0.23	14.78	99.86
	8	0.02	100.08	0.04	13.90	100.11	0.03	14.82	100.09
	9	-0.01	100	-0.01	13.89	100	-0.01	14.80	100
aug-cc-pVDZ ^b									
	1	15.20	109.42	14.66	15.20	109.50			
	2	-1.88	95.89	-1.82	13.32	95.88			
	3	2.15	111.39	2.09	15.47	111.47			
aug-cc-pVTZ ^b									
	1	7.90	106.70		7.90				
	2	-0.86	95.12		7.05				
	3	0.89	107.14		7.94				

Table 3.15: Incremental SSFC_{inc} BSSE correction as a function of the expansion order i with respect to the SSFC scheme for the $(\text{H}_2\text{O})_8$ water cluster at MP2 level calculated with different basis sets. Comparison of the influence of default^a and tight convergence thresholds on the accuracy.

MOLPRO thresholds: SCF energy (default^a 1.00E-07 and tight^b 1.00E-11) and density matrix threshold (default^a 1.00E-05 and tight^b 1.00E-06).

i	i-th order corr. kcal/mol	SSFC_{inc} kcal/mol	$\frac{\text{SSFC}_{\text{inc}}}{\text{SSFC}}$ %	i-th order corr. kcal/mol	SSFC_{inc} kcal/mol	$\frac{\text{SSFC}_{\text{inc}}}{\text{SSFC}}$ %
aug-cc-pVDZ ^a						
1	12.02	12.02	110.88	12.00	12.00	110.75
2	-1.53	10.49	96.74	-1.49	10.51	97.00
3	1.29	11.78	108.69	1.22	11.74	108.30
4	-1.85	9.93	91.58	-1.78	9.95	91.84
5	1.24	11.16	102.98	1.19	11.15	102.85
6	-0.36	10.80	99.66	-0.35	10.80	99.66
7	0.04	10.84	100.00	0.04	10.84	99.98
6-31++G** ^a						
1	20.11	20.11	117.45	20.11	20.11	117.43
2	-2.92	17.19	100.40	-2.91	17.20	100.43
3	0.97	18.16	106.07	0.96	18.15	106.02
4	-1.83	16.33	95.38	-1.82	16.34	95.41
5	0.82	17.15	100.17	0.81	17.15	100.15
6	-0.01	17.14	100.10	-0.01	17.14	100.10
7	-0.02	17.12	100	-0.02	17.12	100

Interestingly we found the accuracy of the BSSE incremental scheme to be affected by the geometrical properties of the investigated system. The SSFC_{inc} BSSE correction calculated for a linear conformer of hydrogen fluoride cluster $(\text{HF})_{10}$ at the MP2/aug-cc-pVXZ (X=D and T) level of theory is given in Table 3.16. As already observed for the water hexamer also the $\text{SSFC}_{\text{inc}}(i=2)$ BSSE correction for the $(\text{HF})_{10}$ cluster is in excellent agreement with the SSFC scheme. The more tight convergence thresholds applied for this cluster definitely eliminate any numerical noises. The third- and fourth-order BSSE correction calculated with the tight convergence thresholds does not contribute significantly to the expansion and hence the series is converged after the second-order. The fragmental BSSE corrections obtained with the triple- ζ basis set are also in excellent agreement with the SSFC scheme at the first-order and what is even more important, the further expansion of the SSFC_{inc} series does not worsen the

accuracy.

Table 3.16: Incremental SSFC_{inc} BSSE correction as a function of the expansion order i with respect to the SSFC scheme for the $(\text{HF})_{10}$ cluster at MP2/aug-cc-pVXZ (X=D and T) level of theory. Comparison of the influence of default^a and tight^b convergence thresholds on the accuracy, MOLPRO thresholds as in Table 3.15.

$(\text{HF})_{10}$ linear structure; geometrical parameters taken from a HF crystal structure [131, 132]: $r_{(H-F)}=0.92 \text{ \AA}$; $R_{(F-F)}=2.50 \text{ \AA}$ and $\angle_{(HFH)}=120^\circ$.

i-th order	corr.	SSFC_{inc}	$\frac{\text{SSFC}_{\text{inc}}}{\text{SSFC}}$	i-th order	corr.	SSFC_{inc}	$\frac{\text{SSFC}_{\text{inc}}}{\text{SSFC}}$
i	kcal/mol	kcal/mol	%	kcal/mol	kcal/mol	kcal/mol	%
aug-cc-pVDZ ^a				aug-cc-pVDZ ^b			
1	10.99	10.99	106.92	10.99	10.99	106.92	
2	-0.77	10.23	99.47	-0.79	10.20	99.21	
3	0.01	10.23	99.55	0.13	10.33	100.50	
4	0.23	10.46	101.78	-0.05	10.28	100.00	
5	-0.39	10.07	97.98	-0.02	10.26	99.84	
6	0.32	10.40	101.14				
7	-0.16	10.24	99.62				
8	0.05	10.29	100.06				
9	-0.01	10.28	100				
aug-cc-pVTZ ^b							
1	6.32	6.32	100.53				
2	-0.002	6.32	100.49				
3	-0.04	6.28	99.87				

Finally we can conclude that the proposed approximation to the SSFC scheme based on an incremental design cannot be recommended to account for BSSE effects in large n -body clusters. In order to gain satisfactory accurate BSSE estimates one needs to carry out the expansion series to high order which in turn causes the calculation to be more expensive than the reference one. The convergence behavior of the approximate SSFC_{inc} scheme is not only found to be very sensitive with respect to the size and type of the investigated systems but the oscillating behavior also prevents the SSFC_{inc} approach to account for BSSE efficiently.

3.3.8 Approximate SSFC(S) scheme

The approximate SSFC(S) scheme has been tested at the CCSD and MP2 level of theory for the entire water cluster series presented in Fig. 3.35. The calculations were performed with the Ahlrichs-type SV(P) and TZVP [133, 134] basis sets available in the TURBOMOLE [125] program package and in addition with Dunning's correlation consistent aug'-cc-pVDZ basis set (which is aug-cc-pVDZ without diffuse functions on hydrogen). Since the convergence behavior of the SSFC(S) approach, similar to the SSFC(R) scheme, is not affected by different theory level, the MP2 results as well as the HF contributions are given in the Appendix B in Tables B.3, B.4 and B.5, and we will limit the discussion to the CCSD outcomes only.

The values of the applied S threshold were chosen in a large interval in order to ensure a detailed investigation of the SSFC(S) scheme. The overlap threshold S runs from 0.1 to 0.01 and from 0.009 to 0.001 in 0.01 and 0.001 steps, respectively. Furthermore the SSFC(S) correction has also been calculated for $S > 0.001$ and $S < 0.1$. Due to the rather large data range we choose i.a. a logarithmic scale for the presentation of the SSFC(S) results. The percentage coverage of SSFC(S) with respect to SSFC is plotted against $-\log(S)$ in the interval of 0.3 to 8 for the three different basis sets SV(P), TZVP and aug'-cc-pVDZ applied at the CCSD level of theory in Fig. 3.20.

An exponential increase of the SSFC(S)/SSFC ratio with respect to the chosen range of S values is observed for the Ahlrichs-type double and triple- ζ basis sets as well as for Dunning's double- ζ basis set. Regarding the accuracy of the SSFC(S) correction, measured by the percentage coverage, we can divide the range of the logarithmic scale into three categories. For $-\log(S) < 1$, the SSFC(S)/SSFC ratio falls below 70% and yields therefore BSSE corrections which are too inaccurate. In the range above 3 on the logarithmic scale the SSFC(S) percentage coverage goes towards 100% for all the basis sets and all the different cluster sizes. The most interesting interval seems to be $1 < -\log(S) < 3$. This region is therefore shown in an enlarged fashion in the charts of Fig. 3.20.

The slope of the curves is different in the range from $-\log(0.1)=1$ to $-\log(0.01)=2$ when we compare the SV(P) and TZVP results. In this interval the basis set extension from double to triple- ζ causes the SSFC(S)/SSFC curves to become smoother and at the same time the SSFC(S)/SSFC ratio significantly higher. In contrast to this, the SSFC(S) data sets between 2 and 3 on the logarithmic scale show very similar convergence behavior for the Ahlrichs-type basis sets, yielding percentage coverage rates beyond 96%.

The convergence behavior of the aug'-cc-pVDZ/CCSD SSFC(S) curves are somehow different from the TZVP or SV(P) ones. The SSFC(S) percentage coverage grows systematically with the decrease in cluster size n for the aug'-cc-pVDZ results what is

neither observed for the SV(P) nor for the TZVP basis set.

The charts of Fig. 3.21 present the average number of basis functions for the three discussed basis sets with respect to the discussed S range. The upper two charts present the average number of basis functions for SV(P) and TVZP basis sets separately for odd and even number of n in $(\text{H}_2\text{O})_n$. The iterate sequence of the orange and black data points with respect to n indicates, that for a given S value more basis functions are omitted on average with a double- than a triple- ζ basis set. We note that the differences are rather small, as the orange and black data points lie almost on top of each other. Generally the shape of the curves looks very similar. We also find, that for all basis sets the percentage average number of basis function is growing as the cluster size decreases with respect to S.

The basis set series introduced by Dunning is known to provide at the correlated level a highly reliable pathway to the CBS limit also for interaction energies calculated with basis sets of increased quality. Therefore we want to focus the analysis, regarding the influence of basis sets on the SSFC(S) scheme, on the family of Dunning's correlation consistent basis sets.

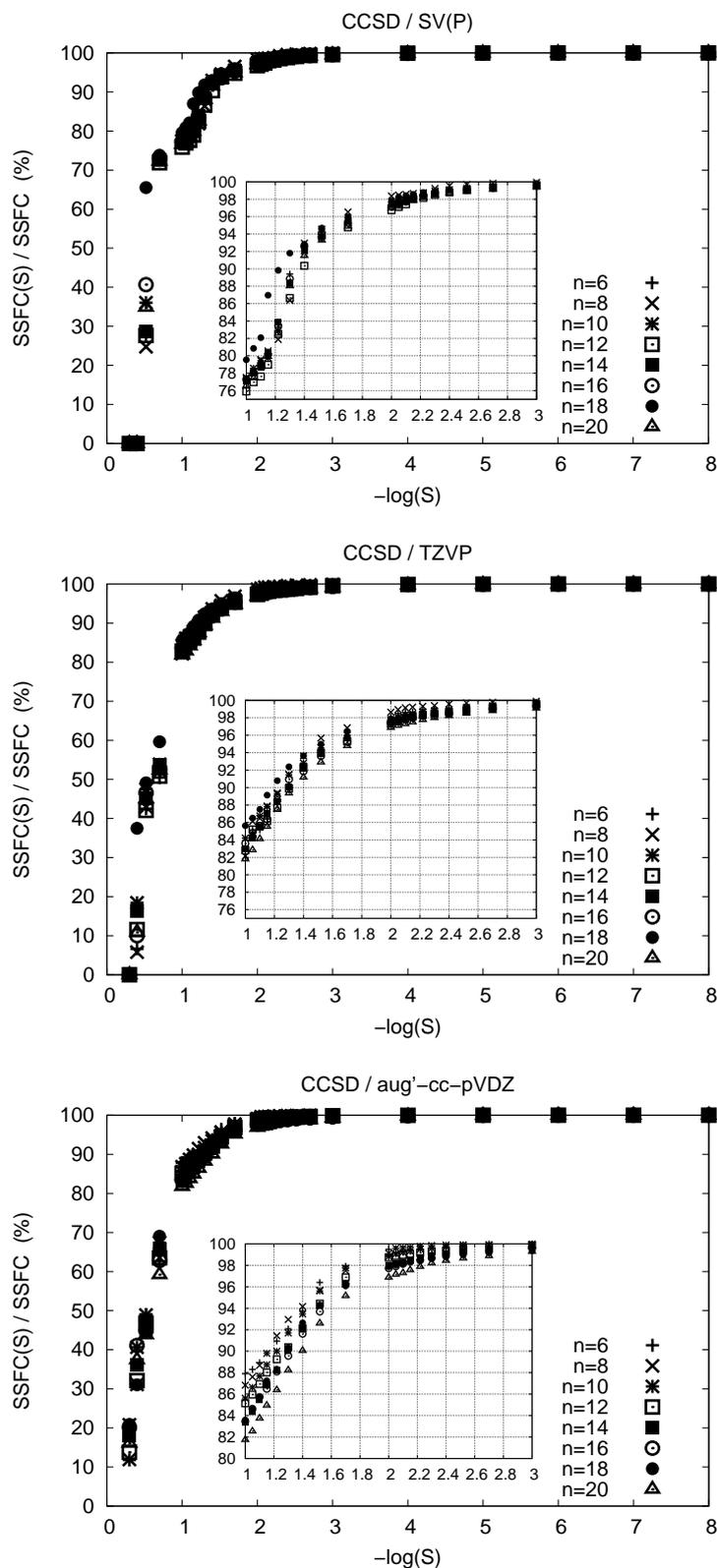


Figure 3.20: Percentage coverage of the approximate SSFC(S) scheme with respect to the full SSFC approach as a function of $-\log(S)$ at the CCSD level of theory calculated with SV(P), TZVP and aug'-cc-pVDZ basis sets for the $(\text{H}_2\text{O})_n$ water cluster series.

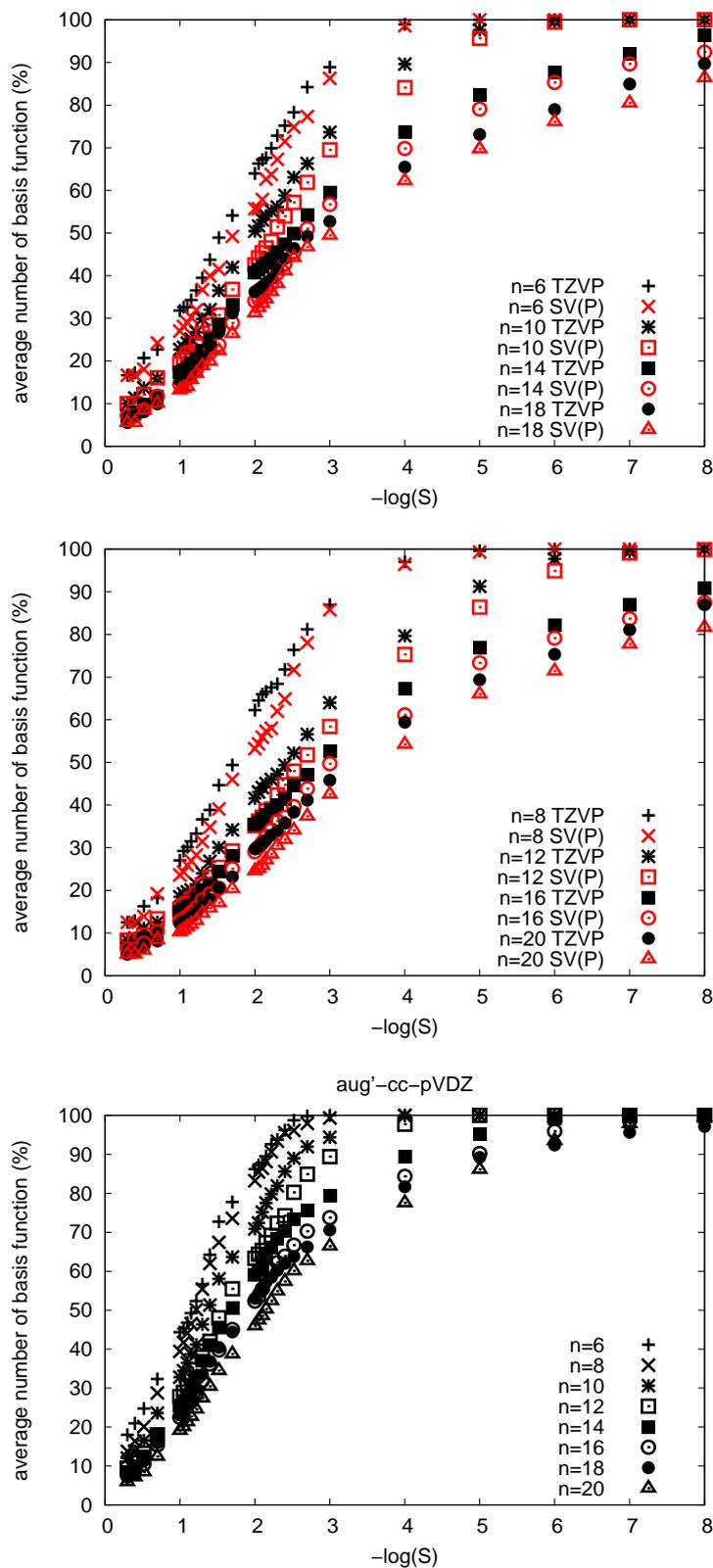


Figure 3.21: Average number of contractions for the SV(P), TZVP and the $\text{aug}'\text{-cc-pVDZ}$ basis sets applied in the calculation of SSFC(S) for the $(\text{H}_2\text{O})_n$ water clusters plotted against $-\log(S)$.

Since the threshold which controls the exclusion of basis functions utilizes the magnitude of the overlap integral between two basis functions, hence the amount and kind of omitted basis functions will depend upon the size and nature of the applied basis set as well as the distance between the atoms at which the basis functions are put at. The most simple model system simulating how many and what kind of basis functions are excluded during a SSFC(S) calculation with respect to different S values is a water dimer with a variety of distances between the water monomers covering the range of distances occurring between the nearest and farthest water monomers in the regarded water cluster series.

Therefore we analyzed the SSFC(S) scheme in detail with respect to the amount and kind of incorporated basis functions at the counterpoise oxygen and hydrogen atoms of a water dimer. This analysis has also advantage that larger basis sets can be very fast investigated as the regarded system is small enough. And in addition as the chosen distances between the two monomers from the dimer are increased in well defined steps, the results enable us to clearly identify the trends within a SSFC(S) calculation with respect to the omitted basis functions. The oxygen-oxygen distances R were chosen between 3.0 to 8.0 Bohr in 0.5 steps and the SSFC(S) calculation were carried out within Dunning's aug'-cc-pVXZ (X=D,T and Q) basis sets at the CCSD level of theory. Since within this model system only one water monomer needs to be treated with ghost functions, we may analyze the percentage number of the employed basis functions for the aug'-cc-pVXZ (X=D,T and Q) directly and not in an average way. The results with respect to $-\log(S)$ are presented in Figs. 3.22 and 3.23, where one chart corresponds to a given oxygen-oxygen distance R. Analyzing these charts we find the following dependence of the results with respect to the three influencing parameters S, R and X:

- The smaller the S threshold (the larger on the logarithmic scale $-\log(S)$) the more basis functions survive the screening,
- pulling the water molecules apart from one another cause the percentage coverage of basis functions with respect to the full CP calculation to become systematically smaller,
- and with the basis set increase of X we find that the percentage of basis functions slightly drops down for different R values, but this arrangement is not generally valid as exceptions are observed exemplary in between $-\log(S)=2$ and $-\log(S)=2.5$ for $R \geq 5$ Bohr.

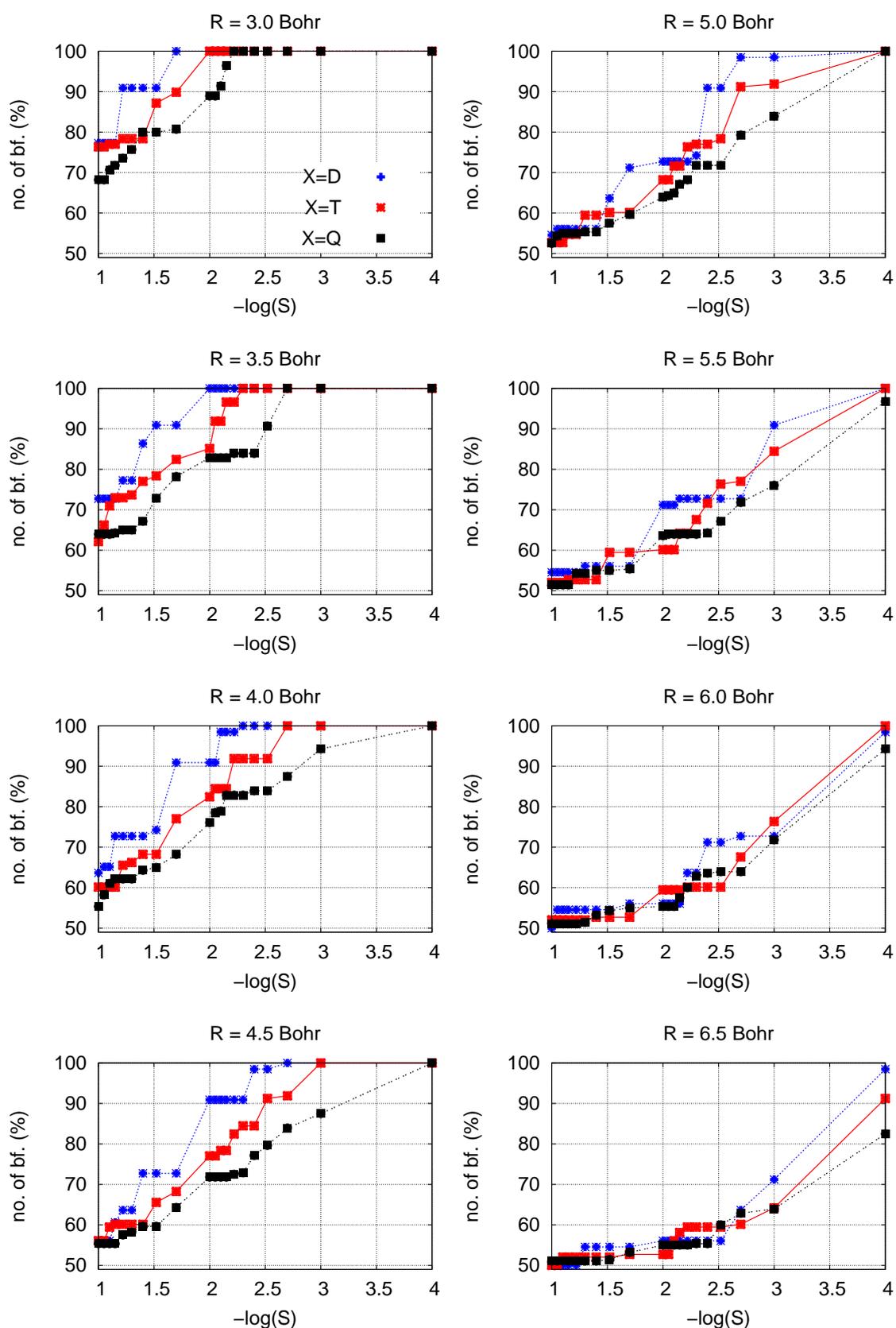


Figure 3.22: Percentage number of aug'-cc-pVXZ (X=D,T,Q) basis functions employed in the counterpoise calculation of a water monomer from a dimer with respect to the different oxygen-oxygen distances R plotted against $-\log(S)$.

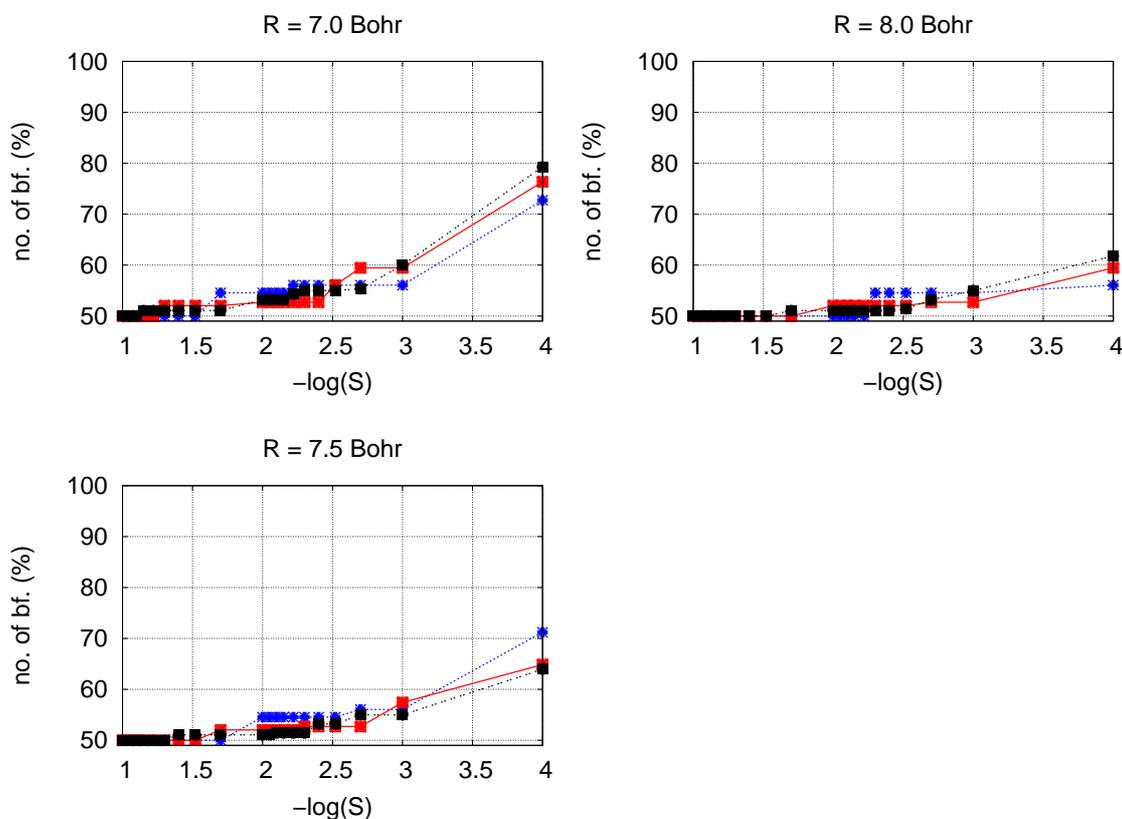


Figure 3.23: As Fig. 3.22, but for charts corresponding to oxygen-oxygen distances from $R=7.0$ to 8.0 Bohr.

Before we proceed to evaluate the data in-depth let us first shortly recapitulate the composition of the correlation consistent basis set series cc-pVXZ ($X=D,T$ and Q and the augmented counterpart of them) for hydrogen and the first row atoms boron through neon. In order to determine the optimum requirements for basis sets at the correlated level, Dunning used as a measure of impact, which a given basis function has on the electronic wave function, the correlation energy lowering caused by this function [135]. He investigated the energy lowering caused by the systematical addition of basis functions of certain angular momentum symmetry l and found that basis functions of different symmetry type may be grouped together as they cause a comparable decrease in the correlation energy. The different sets of basis functions, with its members having individually a comparable energetic impact on the correction energy, are then systematically added to the different basis sets as X increases. The set with the largest energetic influence on the correlation energy is the first one which is added. These function are called the correlation set. Their exponents ζ_μ are optimized in an so-called even-tempered expansion. In this scheme instead of optimizing for a given

l value the exponents in the Gaussian expansion individually, one uses the relationship $\zeta_\mu = \alpha\beta^{\mu-1}$ where μ describes the contraction length. Therefore - irrespective to the amount of primitives used in the expansion - one optimizes only the two parameters α and β . The correlation consistent basis sets consist also of a so-called HF set which in turn is a set of general contracted Gaussians with contraction coefficients and exponents determined from atomic HF calculations. Summarizing the above one finds for the oxygen atom that the smallest basis set, the cc-pVDZ one, to consist of the contracted HF (16s3p) / [2s1p] functions plus a single s-,p- and d-function. The triple- ζ is composed of the HF (16s3p) / [2s1p] functions and an optimized set of two primitive s-, p-, and d-functions and a single f-function. Finally the cc-pVQZ basis set has the (3s3p3d2f1g) functions as the correlation set and in addition the HF (18s3p) / [2s1p] set. For the hydrogen atom a similar set up occurs, with the HF (3s) / [1s] set of functions plus the (1s1p), (2s2p1d) and (3s3p2d1f) correlation set for X=D,T and Q, respectively. On top of that, all these basis sets may be further augmented with diffuse functions, whose exponents have been optimized in HF calculations for anions when the s-, p- exponents were considered and in CI calculations for d-,f- and g- exponents. Keeping this knowledge in mind we now turn back into the discussion regarding the type of omitted basis functions with respect to the overlap threshold and the oxygen-oxygen distance R of the water dimer. The findings will be discussed on the basis of the results presented in the graphs of Figs. 3.24, 3.25, 3.26 and 3.27. In these charts we present the appearance order of basis functions grouped according to their angular momentum symmetry with respect to R and S parameters and for a given basis set X. The mentioned basis functions are those from the counterpoise oxygen atom of the water dimer. In Fig. 3.24 all s-functions from the correlation and HF set are investigated separately for the different basis sets, Fig. 3.24 presents the p-functions analysis for the X=D,T and Q, Fig. 3.26 refers to the d-functions and Fig. 3.27 shows the behavior of f-functions for X=T and Q and g-functions for X=Q with respect to S and R. The S threshold is set on the x-axis in a logarithmic scale, whereas the different oxygen-oxygen distances R are marked by different token, as indicated in the legend. The position of the data points with respect to the y-axis only informs which basis function has been added. All data points below a colored horizontal line have achieved this certain configuration of basis functions defined in the legend. Note that the displacement of the data set in between two colored lines with respect to the y-axis has no physical meaning and is only performed in order to ensure a more clear view of the huge amount of data.

We begin the analysis with an overview introducing the sort of informations disclosed in the charts, we may distinguish at least five different points:

- the order of the type of basis function systematically employed with respect to

S can be read from the charts as one examines the appearance of the different horizontal, colored lines and their explanation in the legend,

- when we analyze the data points with respect to fixed S values, we can see what type of basis function occurs at which distance R,
- from the raise of the lines, from one colored line to another for different X, we can analyze, at which distances R what value of S is needed to incorporate a certain basis function
- how often a certain basis function combination occurs for a given R when S increases can be interpreted when we analyze the length of data points for a certain R series
- when regarding the amount of data points in between two colored lines we gain informations about how often a certain type of configuration occur in general.

We will proceed to analyze the charts from Figs. 3.24 to 3.27 with respect to the pointed out type of information all together. The appearance of the s-, p-, d-, f- and g-functions successively surviving the overlap screening procedure is besides one exception found to behave highly systematic. So we first consider the systematic behavior and then the exception which is found for the order of the s-functions from the aug-cc-pVTZ basis set. For any symmetry type, irrespective to the basis set size, it is always a basis function from the augmented set of basis functions, which is the first one occurring when S increases, refer to the legend of all turquoise horizontal lines in the charts from Figs. 3.24 to 3.27. The next functions following are always from the correlation set of basis function. Whereas the most tight ones, those from the contracted HF sets, are the last, which are added with the increase of $-\log(S)$. For the different primitive functions from the correlation set a systematic order of appearance is found as well. Whenever more than one basis function in a certain symmetry type occurs, as in the (2s2p2d) and (3s3p3d2f) set of functions from the correlation set of the cc-pVTZ and cc-pVQZ basis sets, respectively, we find their order of occurrence depending on the basis function exponent. The more diffuse the s-, p-, d- or f-functions are, the smaller their exponent is, the earlier this functions occur with respect to increase of $-\log(S)$ when a given R is considered. Referring exemplary to the results from the aug-cc-pVQZ chart of Fig. 3.24, with respect to our considerations, we can match the blue, red and green line which correspond to the 1s, 2s and 3s basis function to the exponents of: 0.2067, 0.5547 and 1.428, respectively.

This systematic behavior is also found for the two distinguishable s-functions from the triple- ζ basis set. Here again the first one - which survives the screening - is the one with the smaller exponent when comparing both, even though here we find the

general order of basis functions to be slightly disturbed. This exception was already mentioned in the beginning, as we find with the decrease of S the following sequence of s -functions: $(1s)^{\text{aug}}/[1s] + (1s)/[1s] + (8s)/[1s] + (1s)/[1s] + (8s)/[1s]$. So with respect to the order found in general, in this case the third and fourth basis function are interchanged. However note that no such disruption is found when we refer to the order of basis functions with higher angular momentum symmetry.

When we analyze the sequence of data points along lines parallel to the y -axis, so for constant S values, we find the more basis functions on an counterpoise oxygen atom the closer the two monomers approach each other. From this order no exception is found, neither when we analyze different S values nor when we refer to different basis set sizes X or different type of basis function angular momentum symmetry.

The inspection of the charts with respect to the third point listed in the beginning reveals the following findings. Apparently the larger the distances between two oxygen atoms become the larger the $-\log(S)$ value has to be chosen to gain a certain basis function, irrespective to the type of basis function as well as the basis set size. There is even more information hidden in the charts as in the third point of the overview indicated, as we also can compare the appearance of basis functions of certain symmetry type among the different basis sets X . For the s -, p - and d -functions from the augmented set of basis functions, we find that with the increase of $-\log(S)$ the data points for different R values are shifted to the left side with respect to the x -axis. Exemplary the $(1s)^{\text{aug}}$ function of the double- ζ basis survives the screening for $-\log(S) < 1$ when $R = 3.0, 3.5, 4.0$ and 4.5 Bohr (refer to Fig. 3.24). Whereas for the triple- and quadruple- ζ basis sets we find one and two more counterpoise oxygen atoms (with larger R than 4.5 Bohr), with the $(1s)^{\text{aug}}$ function put at. Compare in second and third charts of Fig. 3.24 data points in the range $-\log(S) < 1$, for the $R = 5.5$ Bohr, and $R = 5.5, 6.0$ Bohr for $X = T$ and Q , respectively. It is also clear that the reason for this order is again easily explainable with the size of the exponent of these basis functions. The augmented s -, p - and d - basis function exponents become smaller with the increase of X . Therefore the arrangement is observed, that for a given R a smaller S value (or larger on the $-\log(S)$ scale) is required for the diffuse s -, p - or d -function from the augmented set of basis function to be incorporated on the oxygen atom. That is why we observe the left shifting of data points, which stand for the augmented basis function when we compare the $X = D, T$ and Q charts for a given symmetry type. Let us again consider one example from the charts of Fig. 3.26. The $(1d)^{\text{aug}}$ basis function is set on the counterpoise oxygen atom when we refer to the $R = 5.5$ series for the first time for $-\log(S) = 1.4, 1.1$ and 0.6 when $X = D, T$ and Q respectively. This characteristics is not only subject to the augmented set of basis function, it is also found for the s -, p -, d - and f -functions from the correlation set. Whenever we find the exponent for a given symmetry type to become smaller with the increase of X , we observe the left shifting of all data points below a certain

colored line. This behavior is found for all data points below:

- the blue horizontal lines in all charts of Fig. 3.24, as the exponents for the three different primitive s-basis functions are equal to 0.3026, 0.2384 and 0.2067 for $X=D, T$ and Q , respectively,
- the blue horizontal lines in the aug-cc-pVDZ and aug-cc-pVTZ charts of Fig. 3.25, as the exponents for the two primitive p-basis functions are equal to 0.2753 and 0.2140 for $X=D$ and T , respectively,
- the red horizontal lines of the aug-cc-pVTZ and aug-cc-pVQZ charts of Fig. 3.25, as the exponents for the two primitive p-basis functions are equal to 0.7156 and 0.5302 for $X=T$ and Q , respectively,
- the blue horizontal lines in all charts of Fig. 3.26, as the exponents for the three different primitive d-basis functions are equal to 0.7156 and 0.5302 for $X=D, T$ and Q , respectively,
- the red horizontal lines of the aug-cc-pVTZ and aug-cc-pVQZ charts of Fig. 3.26, as the exponents for the two different primitive d-basis functions are equal to 2.3140 and 1.30 for $X=T$ and Q , respectively,
- and below the blue horizontal lines of the aug-cc-pVTZ and aug-cc-pVQZ (second chart from above) charts of Fig. 3.27, as the exponents for the two different primitive f-basis functions are equal to 1.428 and 0.859 for $X=T$ and Q , respectively.

The analysis of the stepwise shape of the different R-series with respect to $-\log(S)$ does not follow any general pattern scheme. However there are a few interesting points. Irrespective to the basis set size X , we find for the results obtained with $-\log(S) < 1.5$ and the three smallest R values, that the data points proceed rather fast from one occupation to another. Whereas at the middle range of $-\log(S)$ and R, the data points stay more often at a given basis function configuration when $-\log(S)$ increases. How often a certain basis function composition is hit with the increase of $-\log(S)$ is nicely indicated by the length of data points for a given R in between two colored lines. For the results between 1 and 3 on the logarithmic scale the length of the data with respect to different R values are rather comparable, although of course shifted to the right side as larger $-\log(S)$ values are needed to incorporate a certain basis function when R increases.

The investigation of the last point, from the key issues introduced in the beginning, also does not allow us to formulate any general trend. Considering the amount of data points between two colored lines for any basis function symmetry type or for any

basis set size we find mostly no preference with respect to a certain basis function configuration. But we also note that besides the two exceptions which we will discuss shortly, that the amount of data between two colored lines is comparable. One of the two mentioned exceptions reveals a configuration of basis function which occurs very seldom with respect to R and S, namely the (13s) / [5s] {91111} set of basis function from the aug-cc-pVQZ basis set. But also the (10s) / [3s] {811} basis function configuration, from the triple- ζ basis set, which is the one disrupting the expected order of basis function when $-\log(S)$ increases, is found to be little less frequently matched.

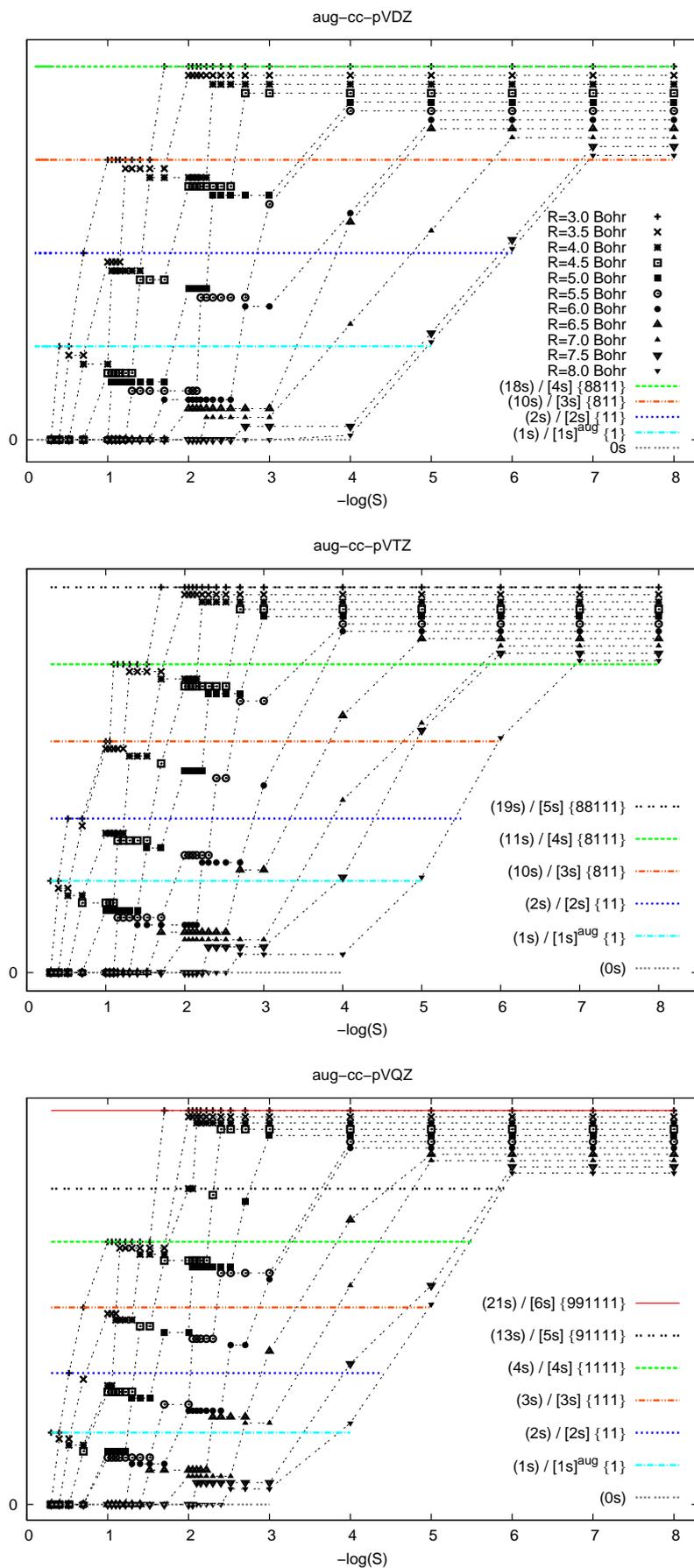


Figure 3.24: Inclusion of s-symmetry bfs. from aug-cc-pVXZ (X=D, T and Q) basis sets on one counterpoise oxygen atom from a water dimer with respect to the overlap threshold S and different oxygen-oxygen distances R .

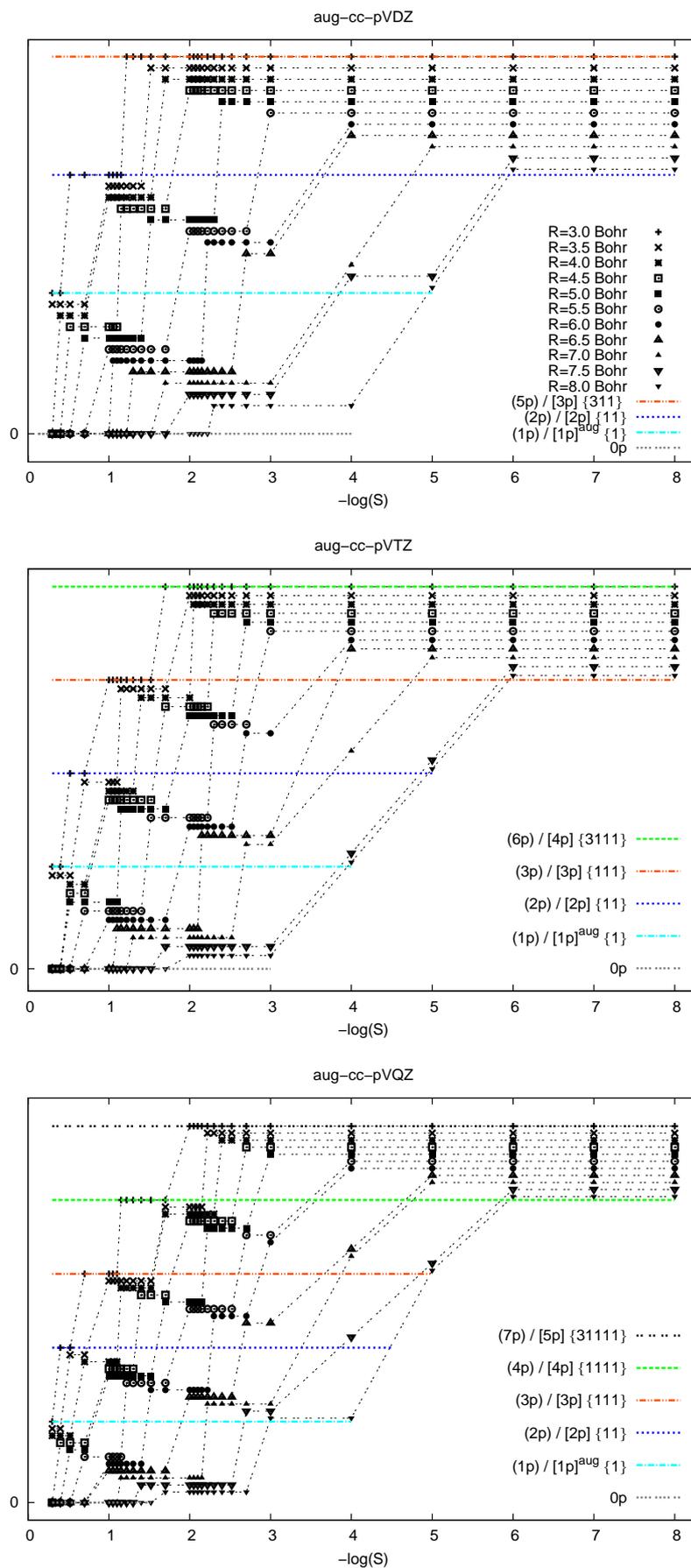


Figure 3.25: Inclusion of p-symmetry bfs. from aug-cc-pVXZ ($X=D, T$ and Q) basis sets on one counterpoise oxygen atom from a water dimer with respect to the overlap threshold S and different oxygen-oxygen distances R .

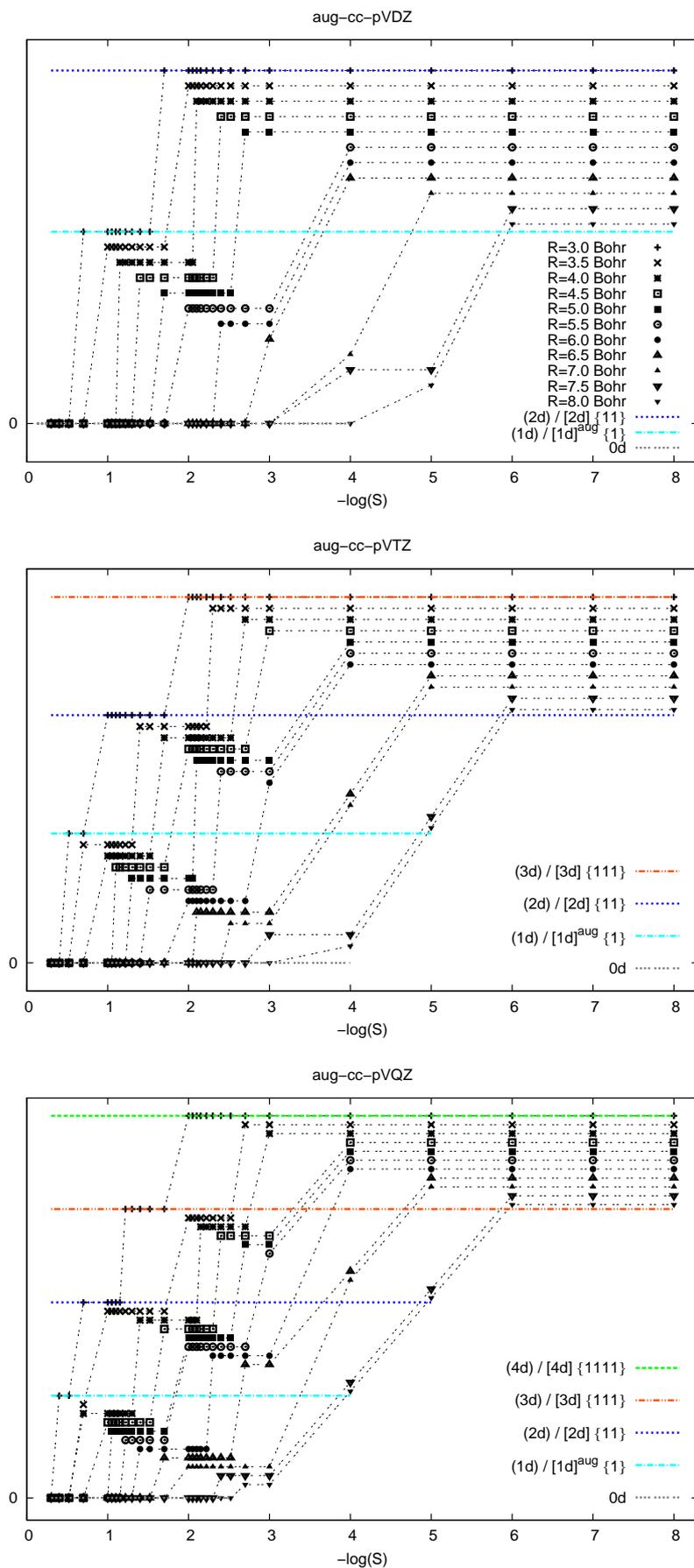


Figure 3.26: Inclusion of d-symmetry bfs. from aug-cc-pVXZ (X=D,T,Q) basis sets on one counterpoise oxygen atom from a water dimer with respect to the overlap threshold S and different oxygen-oxygen distances R .

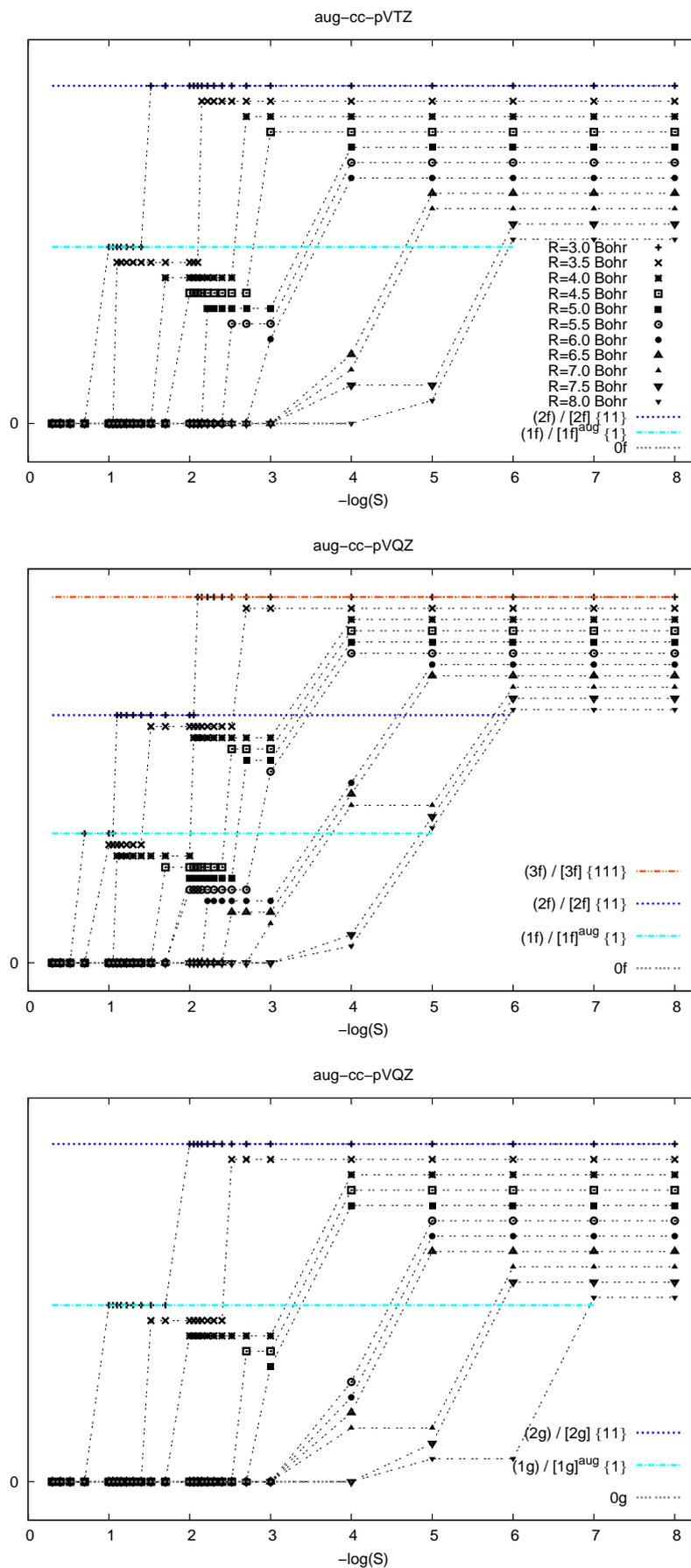


Figure 3.27: Inclusion of f- and g-symmetry bfs. from aug-cc-pVXZ ($X=T,Q$) basis sets on one counterpoise oxygen atom from a water dimer with respect to the overlap threshold S and different oxygen-oxygen distances R .

The results presented so far, grouped according to the type of basis functions, allowed to confirm the expected systematic behavior in the basis function elimination procedure of the SSFC(S) scheme with respect to S and R. But this way of presentation enables us to consider another important issue, namely the overall composition of the resulting basis sets. The presentation of the results from Fig. 3.28 allows us to consider also this issue. In contrast to the type of charts considered previously, the two charts from Fig. 3.28 carry the overall number of contractions on the y-axis. The data are plotted with respect to $-\log(S)$, again for the different oxygen-oxygen distances R, labeled by different token. Thus the charts present the distribution of the different basis function configurations, whereas in the Table below, the contraction pattern of these different configurations are explained. The separate presentation of the R-series into two charts from Fig. 3.28 shall just ensure a clear overview.

For $R \geq 6.0$ Bohr (right hand side chart) we do not find a basis function composition which match the 17 or 14 number of contractions and the configurations corresponding to 7, 12 and 13 are seldom found. Moreover in the interval between 1 and 2 with respect to the x-axis either only the augmented (1s) basis function, or that one together with the augmented (1p) basis function is found on the oxygen atom. In the interval $\{-\log(S) > 2 \wedge -\log(S) < 3\}$, also mainly the (1s1p) functions survive the screening, even though at closer distances R also the augmented (1d) function and primitive (1p) function occurs. For the distribution of different basis function configurations for the $R < 6.0$ Bohr series (left hand side chart), we find for the $-\log(S)$ interval between 1 and 2 the configurations between 4 and 17 contractions frequently hit. Beyond 2 and below 3 on the logarithmic scale, almost all configuration correspond to 12 contractions, this is the region, where to the correlation and augmented set of functions also the most dense ones from the HF set are added.

From the results of Fig. 3.28 we find the following order of basis function occurring on the oxygen atom with the increase of $-\log(S)$: $(1s)^{\text{aug}} + (1p)^{\text{aug}} + (1p) + (1d)^{\text{aug}} + (1s) + (8s) / [1s] + (3p) / [1p] + (1d) + (8s) / [1s]$. So the successively survival of basis functions is also coupled to the angular momentum, as we find the basis functions of s-, p- and d-symmetry being added consecutively. This shows a well-balanced way of putting different basis function on the counterpoise oxygen atom.

From the order in the series above, one exception seems to occur, when we compare this order with all the possible contraction schemes from the Table of Fig. 3.28. Starting from the (1s2p)/[1s2p] configuration we find two different possibilities for the addition of the next basis function. The primitive (1d) from the correlation set competes with the primitive (1s) basis function, to yield whether the (1s2p1d) or the (2s2p) configuration. But the later is found only twice at $-\log(0.09)=1.05$ and $-\log(0.08)=1.10$ for $R=4.0$ Bohr, and the basis set composition following the (2s2p) at $R=4.0$ series is the (2s2p1d) one.

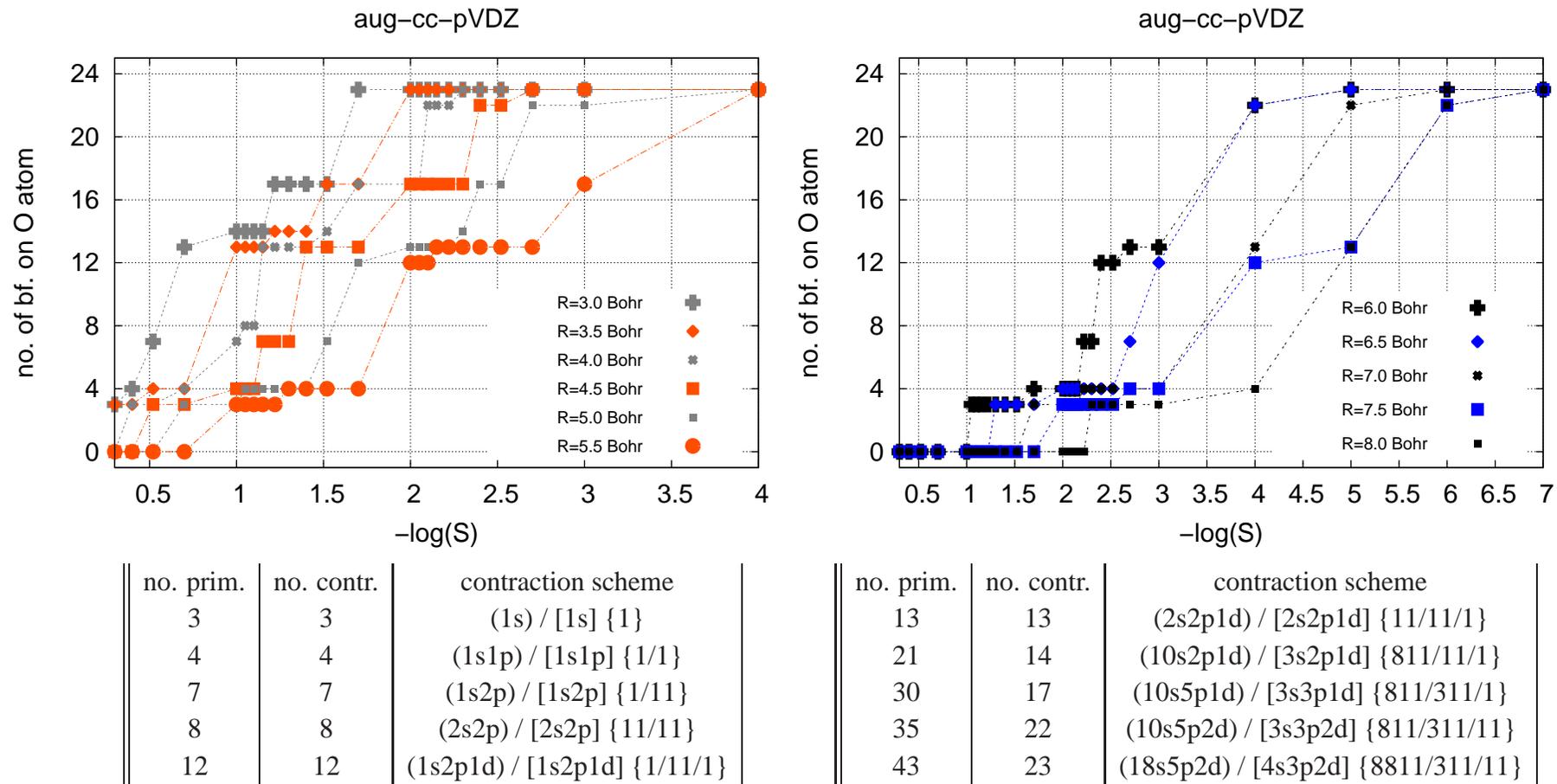


Figure 3.28: Number of contractions put on the counterpoise oxygen atom from a water dimer with respect to the overlap threshold S and for different oxygen-oxygen distances. The analysis is performed for the aug-cc-pVDZ basis set. Below the charts the contraction patterns for the possible number of basis functions are presented.

What we want to point out with these considerations is that, when once a basis function survives the screening, we do not find it deleted later when $-\log(S)$ increases, as one might wrongly interpret when analyzing just the contraction schemes which refer to 8,12 and 13 basis function from the Table of Fig. 3.28.

In order to complete our investigation we also want to consider the basis functions appearance on the counterpoise hydrogen atom shortly since the findings do not differ from the already discussed ones. The results of the analysis are shown in the fashion of Fig. 3.28 for the cc-pVDZ, cc-pVTZ and cc-pVQZ basis sets in Figs. 3.28, 3.30 and 3.31, respectively.

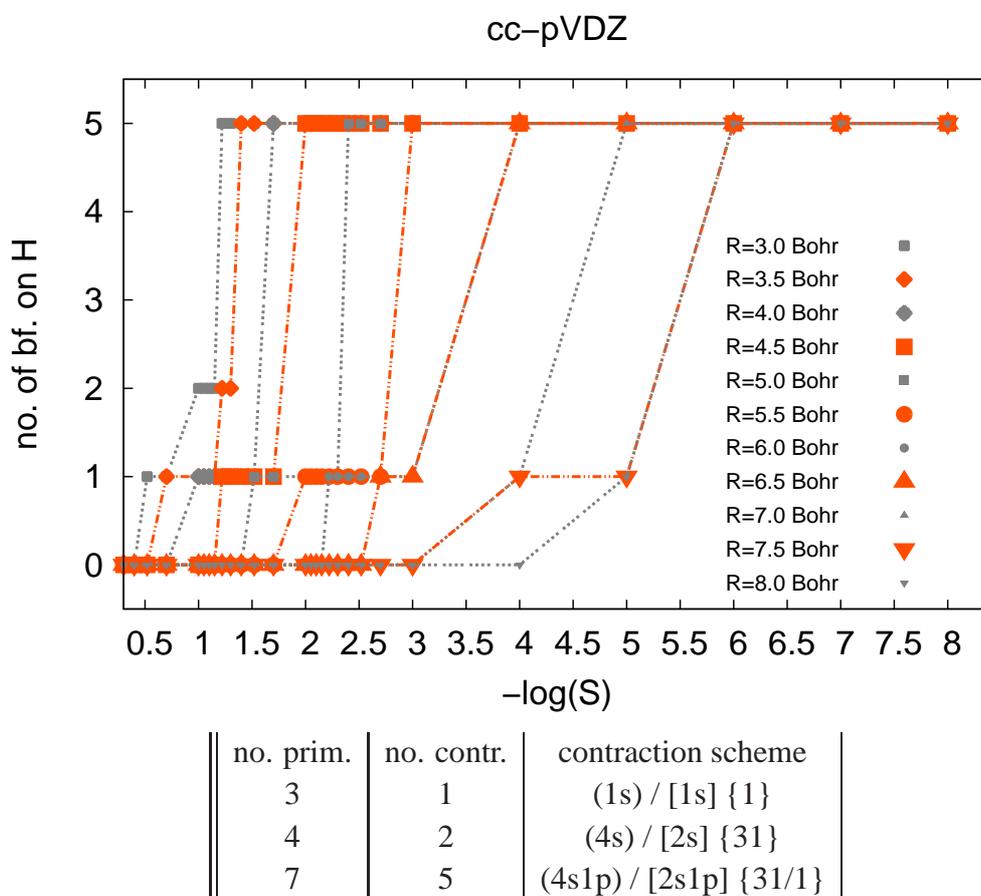


Figure 3.29: Number of contractions put on one counterpoise hydrogen atom from a water dimer with respect to the overlap threshold S and for different oxygen-oxygen distances. The analysis is performed for the cc-pVDZ basis set. Below the chart the contraction patterns for the possible number of basis functions are presented.

For the different (2s), (2p) and (3s), (3p) and (2d) primitives from the cc-pVTZ and cc-pVQZ basis sets, again the order of appearance in the screening procedure depends

upon the size of the exponent when referring to the same symmetry type. Since we did not augment the cc-pVDZ basis set with diffuse functions, hence we have a rather small set of basis functions to be considered, namely only (4s1p) / [2s1p]. The three contracted s- and p-basis functions survive the screening in the order: (1s) / [1s] + (3s) / [1s] + (1p) / [1p] when $-\log(S)$ increases. From the chart of Fig. 3.29 we note, that the (4s) / [2s] configuration is matched only for $R=3.0$ and 3.5 Bohr distances within $\{-\log(S) > 1 \wedge -\log(S) < 1.3\}$. All the remaining data points correspond to the situation where either no basis functions are added, or only the first one surviving the screening, the (1s) occurs, or all the available ones are found at the CP H atom. All data points with respect to the different R-series always run through the (1s) configuration before occupying the highest possible, the (1s) basis function is therefore never skipped.

The possible cc-pVTZ contraction pattern together with the graphical illustration of the distribution among different R and with respect to $-\log(S)$ values for the CP H atom are summarized in Fig. 3.30. Beginning with the (1s) basis function two competitive configurations occur, either the (2s) or the (1s1p), corresponding to the addition of another primitive (1s) or the first primitive (1p). However both configuration occur only four times and it is instead found that both basis functions (1s1p) are added simultaneously to yield the (2s1p) set of functions, counted to appear 24 times in the whole series with respect to R and S. From these considerations the overall order of basis functions which occur at the CP H atom for the triple- ζ basis set is: (1s) + (1s1p) + (3s)/[1s] + (1p) + (1d), and we once again face with the finding that the order depends upon the basis function exponent as well as the angular momentum of the basis functions, where we have the successively addition of basis function of higher angular momentum symmetry.

The basis function systematically incorporated on the counterpoise H atom from the cc-pVQZ basis set are presented in Fig. 3.31. We can distinguish between 14 different configurations of basis functions and therefore at the first glance it is not obviously clear whether we can derive a general order according to which the basis functions survive the screening. But within these 14 different composition four appear just once, that are the ones corresponding to 2, 13, 17 and 25 contractions and the configuration with 23 basis function in sum occurs also only twice. When we neglect these very rarely emerging configurations, we can set up a general order the following series: (1s) + (1p) + (1s) + (1s) + (1p) + (1d) + (3s)/[1s] + (1p) + (1d1f).

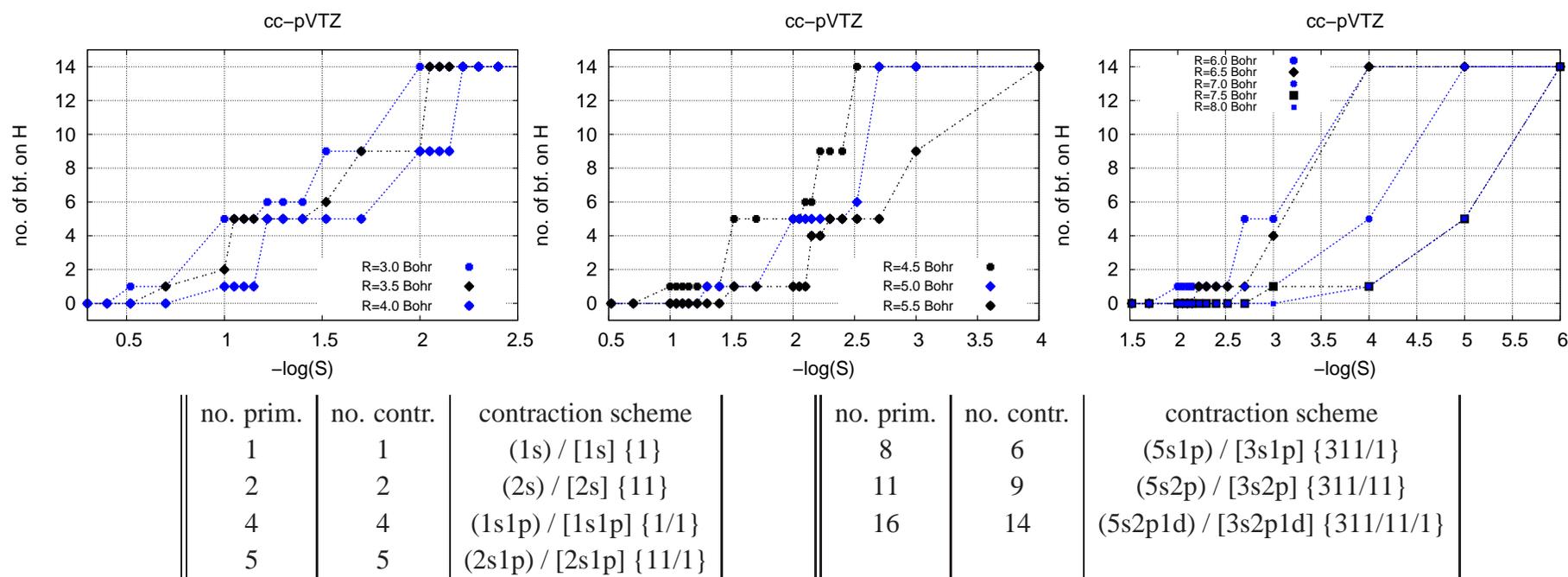


Figure 3.30: Number of contractions put on one counterpoise hydrogen atom from a water dimer with respect to the overlap threshold S and for different oxygen-oxygen distances. The analysis is performed for the cc-pVDZ basis set. Below the chart the contraction patterns for the possible number of basis functions are presented.

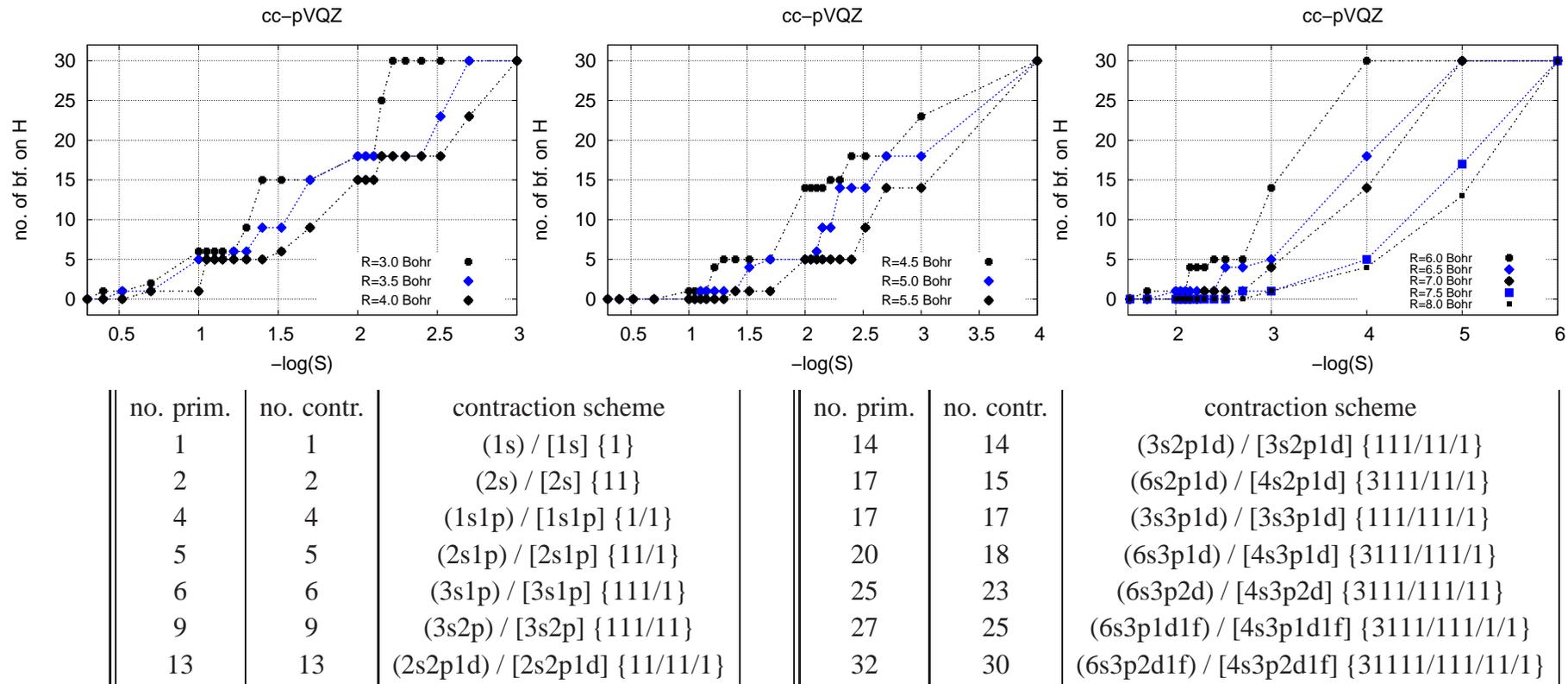


Figure 3.31: As Fig. 3.30, but for the cc-pVQZ basis set.

Even though we keep in mind that both schemes account in contrarious manner for the BSSE corrections, nevertheless they have one important key issue in common. Both schemes aim at the omission of a certain amount of basis functions in order to reduce the computational cost, and - at the same time - at a controlled and as small as possible loss in accuracy. Therefore we want to pose the question, which of the two presented schemes SSFC(R) or SSFC(S) delivers a more efficient way to account for BSSE effects in large n-body clusters. A reasonable comparison between both approaches is assured, when the calculations are performed within the same quantum chemical program. Therefore we recalculated the SSFC(R) BSSE correction, for the entire water cluster series within the TURBOMOLE program using the aug'-cc-pVDZ basis set at the CCSD level of theory.

The results are arranged according to a comparable range of the percentage coverage SSFC(X)/SSFC among X=R and S. This leads to a comparison of the SSFC(S) data obtained with $\{S > 0.1 \wedge S < 0.01\}$ and SSFC(R) corrections calculated within $\{R > 6 \text{ Bohr} \wedge R < 10 \text{ Bohr}\}$. The appropriate juxtaposition in the results is presented in Fig. 3.32. An analog comparison for the interval $\{S > 0.01 \wedge S < 0.001\}$ and $\{R > 10 \text{ Bohr} \wedge R < 15 \text{ Bohr}\}$ is summarized in Fig. 3.33.

Referring to the comparison from Fig. 3.32 the following remarks are important. The SSFC(X)/SSFC ratio differs among the schemes, controlled by either X=S or X=R more significantly for smaller cluster sizes. Whereas for larger n, we find the SSFC(R)/SSFC and SSFC(S)/SSFC curves to resemble more and more each other. The SSFC(R) data are also more spread out for a given R value than the SSFC(S) ones, which are more compact with respect to a fixed S threshold. Since we can not claim to deal with an analogous convergence behavior when we consider SSFC(S)/SSFC and SSFC(R)/SSFC, we should carefully consider the comparison of the advantages both schemes deliver with respect to the saved computational resources, measured by the omitted basis functions in average. As already discussed in Section 3.3.2, where the SSFC(R) scheme for an almost equal interval with respect to R was considered, the percentage average number of basis functions follows a linear fit, with curves shifted parallel among each other and which are ordered accordingly to the cluster size growth. The percentage average number of basis functions, within the SSFC(S) approach on the other hand, can be fitted to a polynomial shape, again with curves covering the more basis functions in average the smaller the cluster size becomes and vice versa. This allows us to conclude, that within the SSFC(S) scheme almost as accurate results are obtained as for the SSFC(R) calculations, especially when we consider larger cluster sizes. Since more basis functions are neglected in average within the overlap threshold approach, we can regard the SSFC(S) scheme as more efficient at least in terms of computational savings, when we refer to the range of R and S thresholds chosen for the calculations in Fig. 3.32.

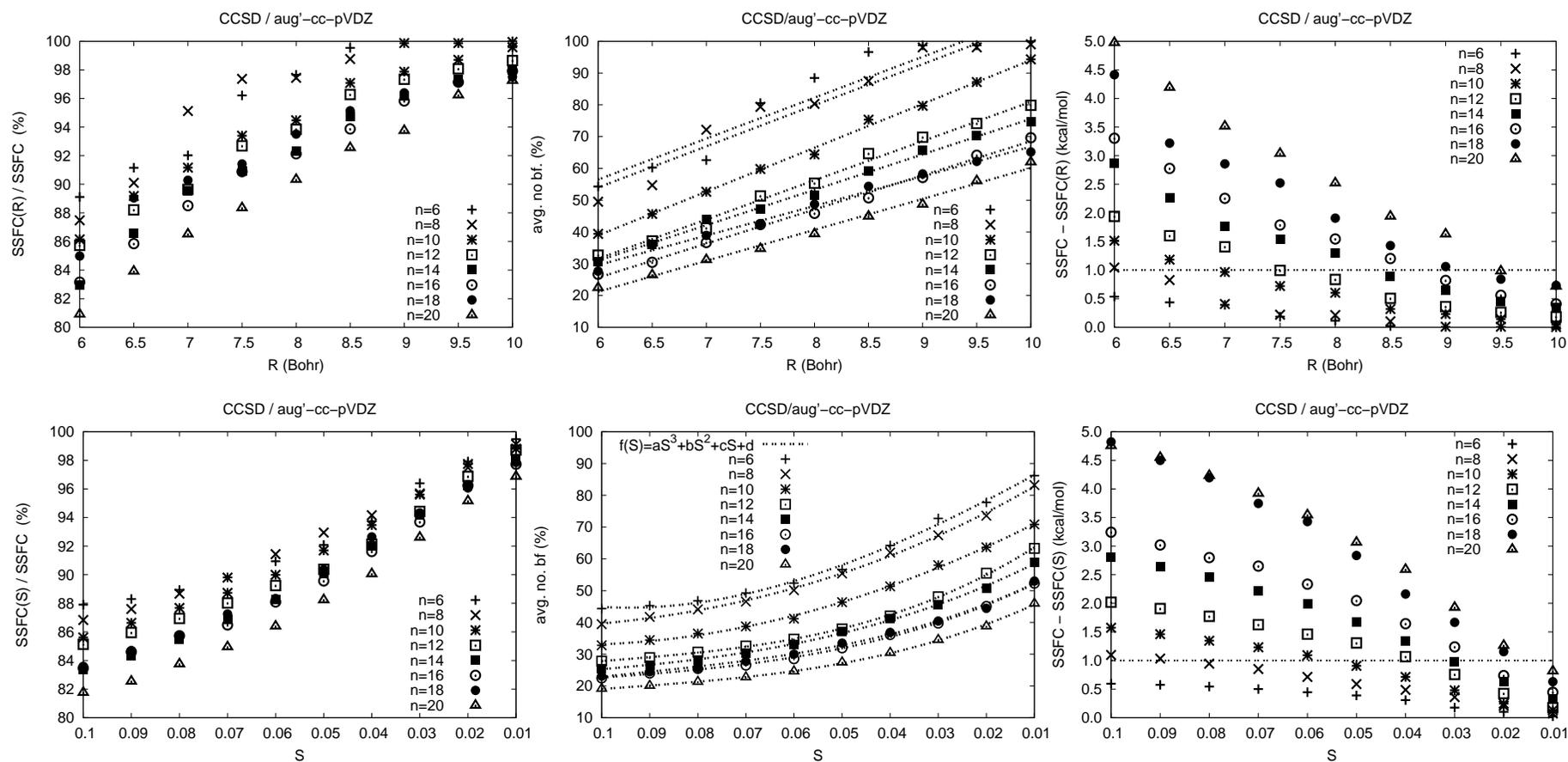


Figure 3.32: Performance of SSFC(X) for X=R with $R \in (6, \dots, 10)$ and for X=S with $S \in (0.1, \dots, 0.01)$ measured by SSFC(X)/SSFC (in %), the average number of basis functions (in %) and the aberration SSFC-SSFC(X) (in kcal/mol). BSSE corrections are calculated at the CCSD/aug'-cc-pVDZ level of theory for the $(\text{H}_2\text{O})_n$ water cluster series.

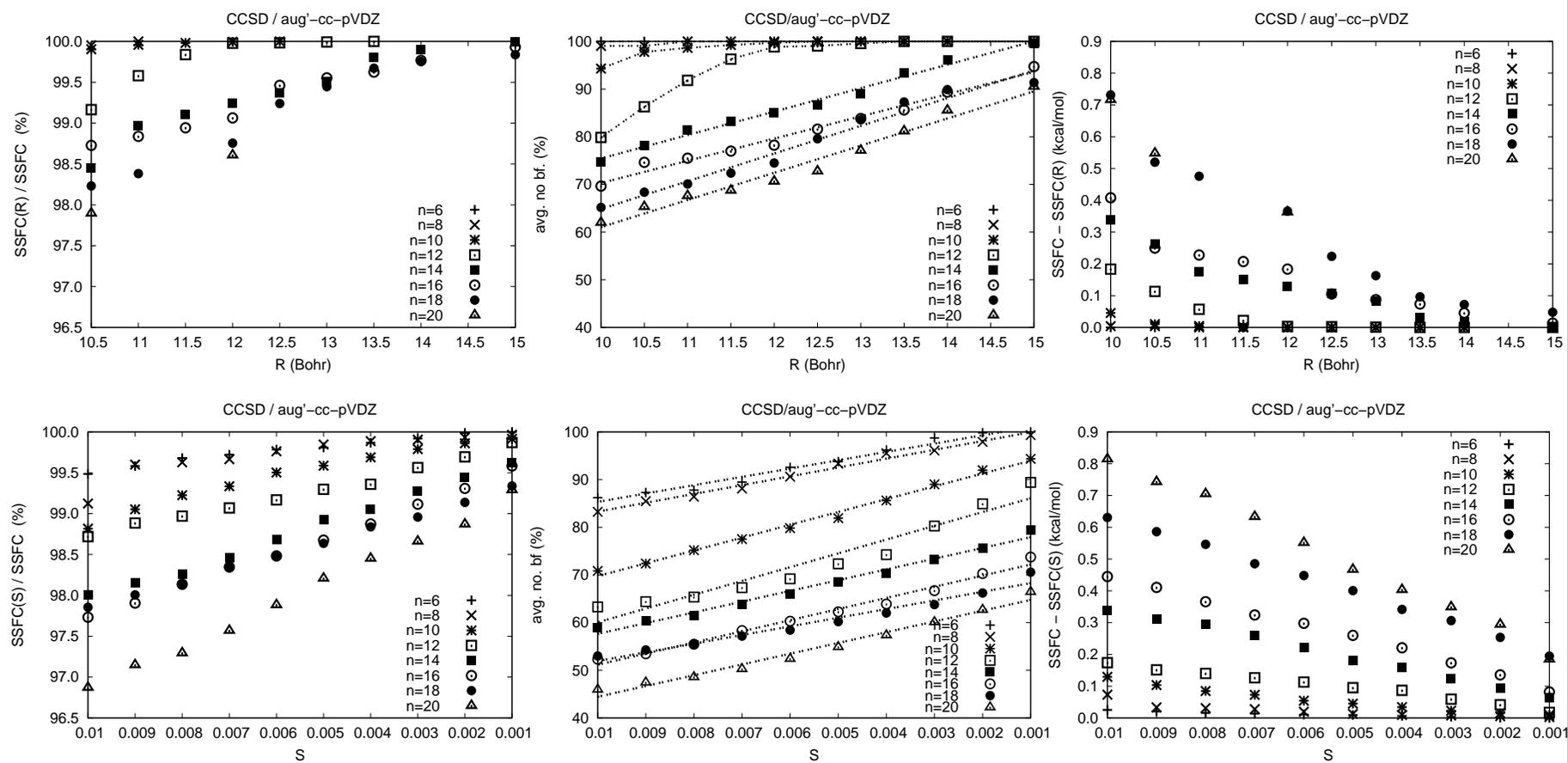


Figure 3.33: As Fig. 3.32, but for: X=R with $R \in (10, \dots, 15)$ and for X=S with $S \in (0.01, \dots, 0.001)$.

However a substantial difference occurs in the region where both schemes approach the 100% limit of the SSFC amount, what is discussed on the basis of the comparison from the results of Fig. 3.33. The average number of basis functions follows for both schemes in the regarded interval (refer to Fig. 3.33) a linear fit. We observe as well, that the SSFC(S) lines are remarkably shifted to smaller number of basis function when compared with the corresponding lines from the SSFC(R) approach. We find the distance-dependent BSSE correction to be faster converged to the full SSFC approach than the overlap driven one. Exemplary for the water hexamer, octamer and decamer 100% of the full SSFC correction is already reached at distance threshold $R=8.5, 9$ and 12 Bohr respectively. The SSFC(S) scheme provides much more flexibility, for the examined interval $\{S > 0.01 \wedge S < 0.001\}$ of S thresholds, which account for 80 to 100% of the SSFC value, we gain a significantly larger range of systematically improved BSSE corrections as in case of the SSFC(R) scheme. What is a direct consequence of the different methodology employed in the construction of the ghost orbital basis sets. The water hexamer exemplary offers just 15 ghost atoms and we can distinguish only between the two extreme possibilities either to choose all basis functions from a regarded basis set on one CP atom or none. This in consequence leads to a smaller range of R -dependent BSSE corrections in contrast to the SSFC(S) scheme, where we also may distinguish in between the set of basis functions from one CP atom. The additional flexibility of the SSFC(S) scheme can be especially seen when referring to the results in the area beyond $-\log(S) > 3$, as presented in the charts of Fig. 3.20. Indeed when we zoom-in into these region, as shown in Fig. 3.34 for the SV(P), TZVP and aug'-cc-pVDZ basis sets, we notice that the SSFC($S > 1e-4 \wedge S < 1e-8$) results are still not converged to the full SSFC approach. And since we discuss herein a percentage coverage of beyond 99.70 for all basis sets and all cluster sizes, the resulting aberration with respect to the SSFC is negligible. This highly accurate results are obtained with still considerable large to even huge computational savings, as indicated in the charts of Fig. 3.21. Exemplary the $-\log(0.0001)$ BSSE correction for the $(\text{H}_2\text{O})_{20}$ cluster at CCSD/aug'-cc-pVDZ level of theory is obtained with 22% basis functions neglected in average, and the aberration from the full SSFC treatment of 26.05 kcal/mol is just equal to 0.06 kcal/mol.

So far we still did not consider all pros and cons of both schemes and should not claim too early the SSFC(S) to be the better approach just on the basis, that it provides a generally more efficient way with respect to the saved number of basis function. The advantage of the SSFC(R) scheme is, that the accuracy of the results are not coupled to the usage of different basis sets. Indeed the SSFC(S) method eliminates the basis functions from the regarded basis sets in a predictable manner. But further investigations are necessary to check, whether the results obtained with different basis sets, may be used to extrapolate the BSSE corrected correlation energies according to the

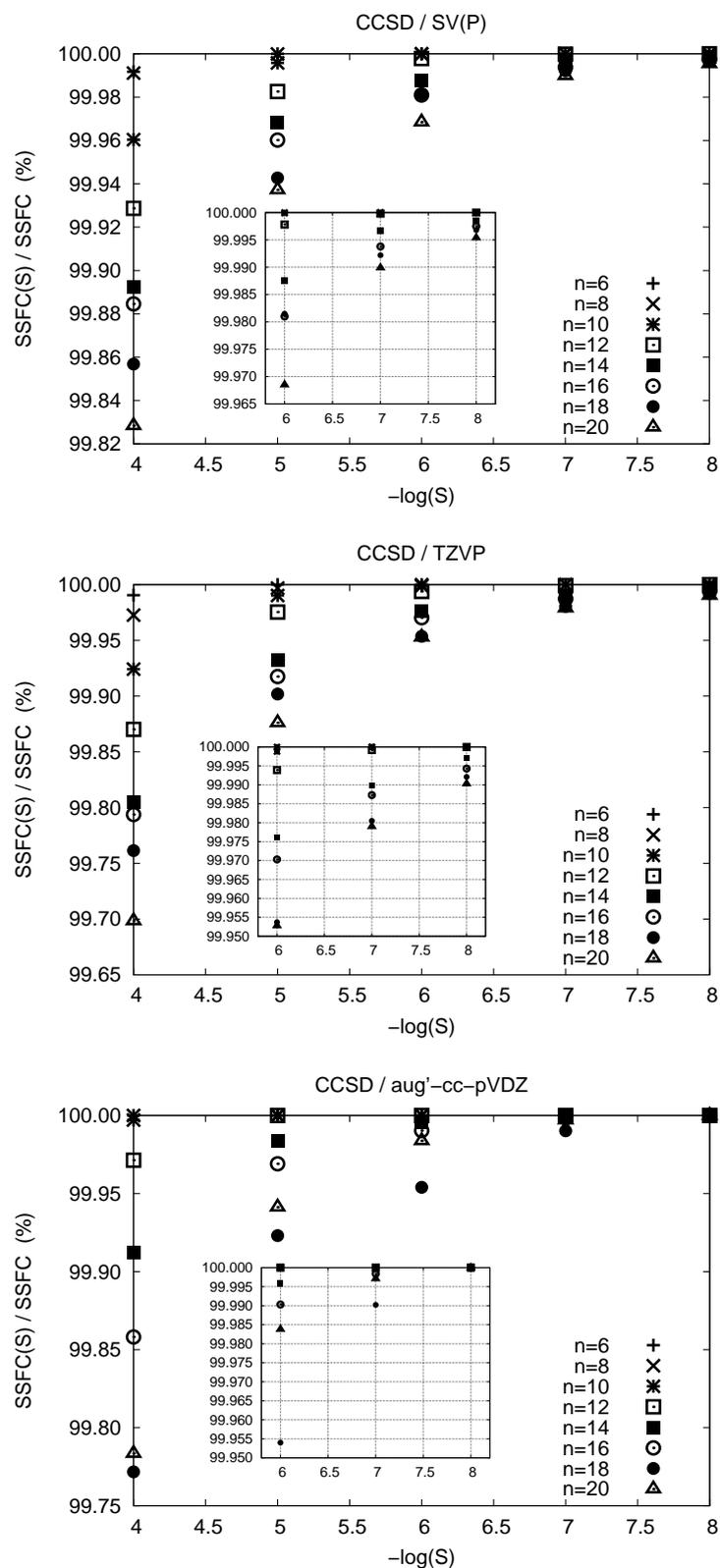


Figure 3.34: Percentage coverage of the SSFC(S) scheme plotted against $-\log(S)$ in the range of 4 to 8 for the $(\text{H}_2\text{O})_n$ water cluster at the CCSD level of theory for the SV(P), TZVP and aug'-cc-pVDZ basis sets.

known extrapolation schemes. The fear, that inconsistent results may occur, is based on the observation that, the incorporated amount and kind of basis functions differ for a chosen S value among the cardinal numbers $X=D,T$ and Q . For the SSFC(R) scheme in contrast we already could prove, that the two-point extrapolation worked fine for SSFC($R=4.0, 4.5\text{\AA}$), cf. the results of Figs. 3.9 and 3.10.

Let us finally consider the last concerning issue. We observed frequently that with both schemes a comparable percentage of BSSE is gained but with a different set of basis functions. We shall therefore also invoke the question, how to compare these two values with respect to a physical interpretation. In other words, we should also judge the physical motivation behind both schemes. We can justify the R -dependent approach with the well known behavior, that the BSSE between two interacting systems decreases as the distance separating them increases. This is a reasonable justification, with the additional advantage, that it even becomes more and more significant as the cluster size increases. Within the SSFC(S) scheme we incorporate the most diffuse basis functions from a given symmetry type and the dense ones are neglected. So we basically made the assumption, that since a set of diffuse basis functions is mandatory to describe weak interactions reasonably, we mainly utilize them to account for the BSSE effect. It is rather difficult to judge the validity of this interpretation, so let us consider some valuable hints from the literature. Estarellas et al. [136] examined the interaction energies of several complexes of benzene with alkaline cations and *s*-triazine with halide anions. Whenever they performed all-electron calculations without incorporating appropriate core-valence basis functions, the estimated BSSE was enormously in comparison with the FC calculations. They also found in contrast to the expected behavior, that the BSSE even increased with the increase of basis set quality. The reported counterpoise BSSE estimates of MP2(FC)/aug-cc-pVTZ vs. MP2(all-electron)/aug-cc-pVTZ are exemplary for the Na^+ -benzene complex: 0.6 kcal/mol (2.5%) vs. 10 kcal/mol (32%) and for the *s*-triazine- Br^- interaction: 1.9 kcal/mol (21%) vs 4 kcal/mol (38%). Further investigations reported by Tzeli et al. [137] also confirmed this erroneous behavior for strongly bound systems and also weak and medium bond strengths. The important message relevant to our studies is, that inappropriate balanced basis sets may affect the amount of BSSE and we should therefore carefully handle the SSFC(S) results. The second important conclusion from the reported papers is, that core-correlation also has an influence on interaction energies, but we will discuss this issue in Section 3.4.

Summarizing these considerations we can say that both schemes have their strength and weakness. Rather than recommend one or another approach we believe that the optimum would be to combine both schemes. So within a certain sphere, defined by the R -distance approach all CP atoms would get all the available basis functions from a regarded basis set and beyond that sphere one could use the overlap threshold scheme.

For the CP atoms beyond the R threshold, a rigid S-value would be favorable, since with almost no loss in accuracy an additional speed-up of the calculations would be achieved.

3.4 Incremental evaluation of core, core-valence and valence correlation energies

3.4.1 Introduction

The scope of the presented study here is twofold, firstly the accuracy of the approximate incremental core-valence treatment as introduced in Section 2.2 is checked and secondly an investigation regarding the importance of core correlation effects in the evaluation of interaction energies is carried out. Therefore Eq. 2.68 is applied to three test systems: a diallylmagnesium complex, a binuclear titanium complex and a tetrahedral mercury cluster, their optimized structures are shown in Fig. 3.35.

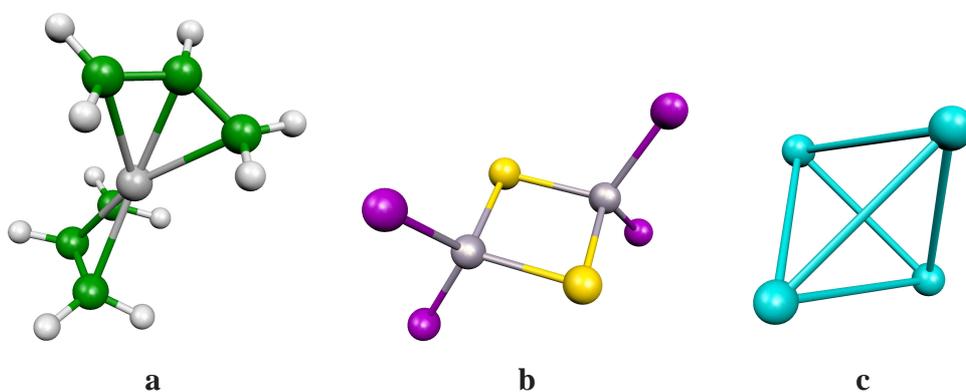


Figure 3.35: Optimized structures of the diallylmagnesium complex $\text{Mg}(\text{C}_3\text{H}_5)_2$ (a), titanium complex $\text{Ti}_2\text{Cl}_4\text{S}_2$ (b), the mercury cluster Hg_4 (c).

All of the test systems except the titanium complex presented here have been chosen in such a way that a standard reference CCSD(T) calculation correlating also the cores was still feasible on the hardware available at the Theoretical Chemistry Institute at the University of Cologne in 2010 (cluster equipped with a 2.4 GHz Core 2 Quad processor, 8 GByte RAM and 200 GByte disk space per node, PCs are connected with 1 GBit ethernet). The computational more demanding calculation on the $\text{Ti}_2\text{Cl}_4\text{S}_2$ complex has required at the CCSD(T)/cc-pwCVTZ theory level correlation of 112 electrons in 558 basis functions, and was no longer manageable due to insufficient memory. The increase of the RAM up to 16 Gb (which was at the time also possible) would make this calculations feasible but the estimated calculation time was of the order of several months. In the next step the incremental core-valence scheme is used to evaluate interaction energies in the formation of sodium cation water complexes and the corresponding sequential interaction energies. For all presented incremental results a comparison with the reference canonical results is made. Therefore for each investi-

gated system two additional canonical calculations are performed, one correlating all orbitals and one using the frozen-core approximation. The difference between both outcomes is taken as the correlation contribution of the core electrons.

The prerequisite for an accurate description of the core correlation is the usage of appropriate tailored basis sets, since most of the standard basis sets have been designed to treat valence correlation only. Suitable basis sets are particularly the core-valence basis sets of Peterson and Dunning [138–140] denoted as cc-pCVXZ and the cc-pwCVXZ basis set series [141] with a weighted core-valence prescription applied.

3.4.2 Applications

The optimized coordinates of the $\text{Mg}(\text{C}_3\text{H}_5)_2$ complex are taken from [142] and have been obtained with the B3PW91 method, using the aug-cc-pVTZ basis set for C and H and the cc-pCVTZ for Mg. From the SCF results for the calculation on $\text{Mg}(\text{C}_3\text{H}_5)_2$ one obtains 29 occupied orbitals. The 11 energetically lowest, corresponding to the 1s carbon orbitals and the 1s2s2p magnesium orbitals are grouped into 7 core domains. The remaining 18 valence orbitals are partitioned into 6 valence domains. The results of the incremental CCSD(T)/cc-pCVDZ and MP2/cc-pCVDZ calculations are given in Table 3.17. Therein the following notation:

$$E_{\text{corr}}^{\text{total}}(i) = E_{\text{corr}}^{\text{c}}(1) + E_{\text{corr}}^{\text{cv}}(i) + E_{\text{corr}}^{\text{v}}(i), \quad (3.35)$$

is used to describe the sum up to a given expansion order i of the incremental core (c), core-valence (cv) and valence (v) correlation energy contributions as introduced in Eq. 2.68. The presented results are grouped into three blocks illustrating the accuracy of the proposed expansion for core, core-valence, valence-only and total correlation energy contributions separately.

Table 3.17: Convergence of the incremental series for core (c), core-valence (cv), valence (v) and total CCSD(T) and MP2 correlation energies of the diallylmagnesium complex $\text{Mg}(\text{C}_3\text{H}_5)_2$ obtained with the cc-pCVDZ basis set. The results are compared with the canonical CCSD(T) and MP2 correlation calculation. (0 frozen orbitals, 11 correlated core orbitals, 7 core domains, 18 valence orbitals, 6 valence domains, $t_{\text{con}}=3$ Bohr, $\text{dsp}=3$, $E_{\text{thres}}=1\text{e-}5$ au)

\mathcal{O}	E_{corr}^c au	$E_{\text{corr}}^{\text{cv}}$ au	$E_{\text{corr}}^{\text{c,cv}}$ (i) au	error kcal/mol	$E_{\text{corr}}^{\text{c,cv}}/E_{\text{can.}}$ %	E_{corr}^v (i) au	error kcal/mol	$E_{\text{corr}}^v/E_{\text{can.}}$ %	$E_{\text{corr}}^{\text{total}}$ (i) au	error kcal/mol	$E_{\text{corr}}^{\text{total}}/E_{\text{can.}}$ %
1	-0.361780	–	-0.361780	26.20	89.65	-0.683126	184.88	69.87	-1.044907	211.08	75.65
2	0	-0.043069	-0.404850	-0.82	100.33	-0.982418	-2.93	100.48	-1.387268	-3.75	100.43
3	0	0.001312	-0.403538	0.00	100.00	-0.977609	0.09	99.99	-1.381147	0.09	99.99
canonical CCSD(T)			-0.403537			-0.977755			-1.381291		
1	-0.344785	–	-0.344785	26.71	89.01	-0.562545	196.88	64.20	-0.907329	223.59	71.80
2	0	-0.042541	-0.387326	0.01	99.99	-0.872003	2.69	99.51	-1.259329	2.70	99.66
3	0	-0.000009	-0.387335	0.01	100.00	-0.876210	0.05	99.99	-1.263546	0.06	99.99
canonical MP2			-0.387349			-0.876287			-1.263636		

For the sum of the core and core-valence CCSD(T) correlation energies $E_{\text{corr}}^{\text{c-cv}}(i)$ at second-order an aberration of -0.82 kcal/mol from the canonical result is achieved, being already within chemical accuracy of 1 kcal/mol (see results from Table 3.17). The accuracy is significantly improved at third-order, where virtually exact agreement with the canonical reference data is obtained. For the MP2 incremental treatment of $\text{Mg}(\text{C}_3\text{H}_5)_2$ a comparable high accuracy is even observed at the second-order of the expansion. The third-order expansion does not deteriorate the accuracy. Both MP2 and CCSD(T) convergence for $E_{\text{corr}}^{\text{v}}$ is somewhat slower than for $E_{\text{corr}}^{\text{c-cv}}$. At second-order the valence correlation energy contributions vary from the canonical reference outcome still by around 3 kcal/mol. These errors are drastically reduced to about 0.1 kcal/mol when going from second to third-order expansion. In the last three columns in Table 3.17 the comparison of the accuracy of the total correlation energies with respect to the canonical outcomes yield also small errors and hence the overall accuracy of Eq. 2.68 is also satisfactory. Note that the convergence of the core correlation contributions is faster than the one of the valence correlation contributions and this behavior is even more significant for MP2 level of theory. The number of individual calculations saved according to the comparison of the outcomes between Eq. 2.69 and Eq. 2.70 is equal to 21 at second-order and even to 161 at third-order of the incremental expansion, what corresponds to 23 and 43% savings, respectively.

The optimized structure of the binuclear titanium complex $\text{Ti}_2\text{Cl}_4\text{S}_2$ was obtained at the RI-BP86/SV(P) level of theory using the TURBOMOLE 5.10 [97, 98] program. The analysis of the Hessian identified the stationary point as a local minimum. The molecule was further investigated with Dunning's correlation consistent triple- ζ basis set (cc-pVTZ) and the corresponding core-valence analogon (cc-pwCVTZ) [143, 144] at incremental MP2 and CCSD(T) level of theory.

The $\text{Ti}_2\text{Cl}_4\text{S}_2$ complex is a challenging system to examine the core-valence treatment since the core is composed of a total of 48 occupied orbitals. From the overall 72 occupied orbitals available, the 16 energetically lowest ones corresponding to the [He]-shells of sulfur and chlorine atoms and the [Ne]-shell contributed by the titanium atom are excluded from the correlation calculation. The 32 orbitals treated as correlated core yield 8 core domains arising from the [Ne]-shells of sulfur and chlorine atoms and the [Ar]-shell of titanium. The remaining 24 valence orbitals are divided into 6 valence domains. The results for the incremental CCSD(T) and MP2 energies and the canonical reference with the cc-pVTZ and cc-pwCVTZ basis sets applied are summarized in Tables 3.18-3.19.

Table 3.18: Convergence of the incremental series for core (c), core-valence (cv) and valence (v) and total CCSD(T) and MP2 correlation energies of titanium complex $\text{Ti}_2\text{Cl}_4\text{S}_2$ obtained with the cc-pVTZ basis set. (16 frozen orbitals, 32 correlated core orbitals, 8 core domains, 24 valence orbitals, 6 valence domains, $t_{\text{con}}=3$ Bohr, $\text{dsp}=3$, $E_{\text{thres}}=1\text{e-}6$ au, Order-dependent distance screening according to: $t_{\text{dist}}=80/i^2$ in Bohr)

\mathcal{O}	E_{corr}^c au	$E_{\text{corr}}^{\text{cv}}$ au	$E_{\text{corr}}^{\text{c,cv}}(i)$ au	error kcal/mol	$E_{\text{corr}}^{\text{c,cv}}/E_{\text{can.}}$ %	$E_{\text{corr}}^v(i)$ au	error kcal/mol	$E_{\text{corr}}^v/E_{\text{can.}}$ %	$E_{\text{corr}}^{\text{total}}(i)$ au	error kcal/mol	$E_{\text{corr}}^{\text{total}}/E_{\text{can.}}$ %
1	-0.511892	–	-0.511892	176.12	64.59	-1.421812	101.91	89.75	-1.933704	278.02	81.36
2	0	-0.312134	-0.824026	-19.75	103.97	-1.614508	-19.01	101.91	-2.438533	-38.76	102.60
3	0	0.036512	-0.787513	3.16	99.36	-1.582135	1.31	99.87	-2.369649	4.47	99.70
4	0	-0.004404	-0.791917	0.40	99.92	-1.583746	0.29	99.97	-2.375664	0.69	99.95
canonical CCSD(T)			-0.792549			-1.584215			-2.376765		
1	-0.460855	–	-0.460855	193.20	59.95	-1.265479	117.55	87.11	-1.726334	310.75	77.71
2	0	-0.307639	-0.768494	0.16	99.97	-1.453275	-0.29	100.03	-2.221770	-0.14	100.01
3	0	0.000103	-0.768391	0.22	99.9	-1.452821	-0.01	100.005	-2.221211	0.21	99.98
canonical MP2			-0.768742			-1.452808			-2.221550		

Table 3.19: Convergence of the incremental series for core (c), core-valence (cv) and valence (v) and total CCSD(T) and MP2 correlation energies of titanium complex $\text{Ti}_2\text{Cl}_4\text{S}_2$ obtained with the cc-pwCVTZ basis set. (16 frozen orbitals, 32 correlated core orbitals, 8 core domains, 24 valence orbitals, 6 valence domains, $t_{\text{con}}=3$ Bohr, $\text{dsp}=3$, $E_{\text{thres}}=1\text{e-}6$ au, Order-dependent distance screening according to: $t_{\text{dist}}=80/i^2$ in Bohr)

\mathcal{O}	E_{corr}^c au	$E_{\text{corr}}^{\text{cv}}$ au	$E_{\text{corr}}^{\text{c,cv}}(i)$ au	error kcal/mol	$E_{\text{corr}}^{\text{c,cv}}/E_{\text{can.}}$ %	$E_{\text{corr}}^v(i)$ au	error kcal/mol	$E_{\text{corr}}^v/E_{\text{can.}}$ %	$E_{\text{corr}}^{\text{total}}(i)$ au	error kcal/mol	$E_{\text{corr}}^{\text{total}}/E_{\text{can.}}$ %
1	-1.816912		-1.816912			-1.449396	103.52	89.78	-3.266308		
2	0	-0.587040	-2.403952			-1.644325	-18.80	101.86	-4.048277		
3	0	0.039782	-2.364170			-1.612392	1.24	99.88	-3.976562		
4	0	-0.004309	-2.368479			-1.613901	0.29	99.97	-3.982380		
canonical CCSD(T)						-1.614366					
1	-1.742465	–	-1.742465	378.51	74.28	-1.290960	119.07	87.19	-3.033425	497.58	79.28
2	0	-0.602513	-2.344977	0.43	99.97	-1.481162	-0.29	100.03	-3.826139	0.14	99.99
3	0	0.000087	-2.344890	0.48	99.97	-1.480715	-0.01	100.00	-3.825605	0.48	99.98
canonical MP2			-2.345659			-1.480704			-3.826363		

The convergence of the incremental expansions for MP2 as well as CCSD(T) theory has already been proven not to be sensitive to the variation of different basis sets [55, 57] what is also evidenced in the study on the cystein molecule as shown in the Appendix C. Therefore one may expect the accuracy of the CCSD(T)/cc-pwCVTZ incremental results to be comparable to the CCSD(T)/cc-pVTZ ones. The CCSD(T)/cc-pVTZ results in Table 3.18 show that a fourth-order expansion of the core-valence incremental series covers 99.92 % of the correlation energy corresponding to an error of 0.4 kcal/mol. For the incremental valence-only contributions a fourth-order expansion is needed to gain chemically accurate results with an error of 0.3 kcal/mol. The overall accuracy is satisfactory with 99.95 % of the correlation energy and 0.7 kcal/mol aberration from the canonical result.

The extension of the basis sets from cc-pVTZ to cc-pwCVTZ quality has significant effects on the size of the core correlation contribution (see Table 3.19), which is about three times larger than for the cc-pVTZ results. The valence-only incremental expansion calculated with the cc-pwCVTZ basis set is as accurate as the cc-pVTZ results. As already mentioned in the introduction the reference calculation correlating 56 occupied orbitals at once was no longer feasible and therefore no comparison for the core-valence correlation energies is quoted in Table 3.19.

The convergence of the MP2 incremental series with both cc-pwCVTZ and cc-pVTZ basis sets is satisfactory, highly accurate results are obtained already at second-order with the largest error of only 0.4 kcal/mol. Since for both basis sets the overall correlation energy is covered by 99.98 % one can conclude that convergence behavior is not affected by the use of different basis sets.

The employment of the core-valence treatment allows the reduction of 28 individual calculations (31 %) at second-order and even 224 (62 %) and 826 (83 %) at third- and fourth-order respectively. This yields a significant improvement in terms of computational time. Additionally the usage of the order-dependent distance threshold of 80 further reduces the amount of individual third and fourth calculations by 24 and 81 respectively.

The tetrahedral Hg₄ cluster has been optimized with the RI-BP86/SVP gradient-corrected density functional theory method [145, 146] as implemented in the TURBOMOLE 5.6 quantum chemistry package [147].

For the incremental CCSD(T) and MP2 calculations a two-component relativistic multi-configuration Dirac-Hartree-Fock adjusted pseudopotential of Figgen et al. [148] with the correlation consistent core-valence basis set of double- ζ quality cc-pwCVDZ [141] designed for the outer-core electron correlation has been applied. The pseudopotential replaces the Hg 1s-4f shells, i.e. it removes 60 electrons, leaving 20 valence electrons per mercury atom. For the calculations presented here the remaining 5s5p shell was considered to be the outer-core shell whereas the 5d6s

shell comprises the valence electrons. Thus the 16 correlated outer-core orbitals were partitioned into 4 core domains and the 24 valence orbitals were divided into 4 valence domains. The incremental calculations are presented in Table 3.20 for CCSD(T) and MP2 correlation energies, respectively.

Table 3.20: Convergence of the incremental series for core (c), core-valence (cv), valence (v) and total CCSD(T) correlation energies of the Hg₄ cluster obtained with the MCDHF-PP/cc-pwCVDZ basis set. (0 frozen orbitals, 16 correlated core orbitals, 4 core domains, 24 valence orbitals, 4 valence domains, $t_{\text{con}}=3$ Bohr, $\text{dsp}=3$, $E_{\text{thres}}=1e-5$ au)

\mathcal{O}	E_{corr}^c au	E_{corr}^{cv} au	$E_{\text{corr}}^{c,cv}$ (i) au	error kcal/mol	$E_{\text{corr}}^{c,cv}/E_{\text{can.}}$ %	E_{corr}^v (i) au	error kcal/mol	$E_{\text{corr}}^v/E_{\text{can.}}$ %	$E_{\text{corr}}^{\text{total}}$ (i) au	error kcal/mol	$E_{\text{corr}}^{\text{total}}/E_{\text{can.}}$ %
1	-0.263467	–	-0.263467	325.79	33.66	-1.433329	14.20	98.45	-1.696796	339.99	75.80
2	0	-0.519521	-0.782988	-0.21	100.0	-1.456495	-0.34	100.44	-2.239483	-0.55	100.04
3	0	0.000335	-0.782653	0.00	100.0	-1.455955	0.00	100.00	-2.238608	0.00	100.00
canonical CCSD(T)			-0.782646			-1.455961			-2.238607		
1	-0.243710	–	-0.243710	355.64	30.07	-1.503336	20.11	97.91	-1.747046	375.75	74.47
2	0	-0.566718	-0.810428	0.02	100.00	-1.535337	0.03	100.00	-2.345765	0.05	100.00
3	0	-0.000034	-0.810461	0.00	100.00	-1.535387	0.00	100.00	-2.345849	0.00	100.00
canonical MP2			0.810464			-1.535385			-2.345849		

The convergence behavior for CCSD(T) $E_{\text{corr}}^{\text{c-cv}}(i)$ contributions is very good, the incremental series is in satisfactory agreement with the canonical result already at second-order and even a perfect agreement at third-order is observed. For the MP2 core and core-valence increments as already observed in case of the titanium complex a faster convergence behavior is found than for the CCSD(T) series, the incremental second-order results are almost equal to the canonical one.

The reduction of the number of individual calculations when going from Eq. 2.69 to Eq. 2.70 yield 17 % savings at second-order and 30 % at third-order with respect to a calculation according to Eq. 2.67.

From these considerations together with the discussion on the diallylmagnesium and the titanium complex the conclusion may be drawn that the approximations introduced in Eq. 2.68 are well chosen and allow an efficient and at the same time accurate description of the core-valence effects for MP2 and CCSD(T) correlation energies within the incremental framework. A third-order truncation of the incremental core-valence expansion assured highly accurate results with errors of less than 0.1 kcal/mol for all cases besides the titanium complex, where a fourth-order expansion was necessary to gain chemically accurate results. The convergence behavior of the proposed scheme is found to be quite insensitive to the basis set quality and compared to a standard incremental expansion the omission of inter-core correlation contributions allows the a priori elimination of a significant number of correlation calculations.

3.4.2.1 Further approximations

It is clearly seen so far that the correlation of the core region is a highly demanding task because the additional core orbitals which cause an enormously growth of the applied basis set if particularly core-valence basis sets are considered. The increase in the number of basis functions when comparing the cc-pVZX and cc-pCVXZ basis sets for the following systems: $\text{Mg}(\text{C}_3\text{H}_5)_2$ (X=D), Hg_4 (X=D) and $\text{Ti}_2\text{Cl}_4\text{S}_2$ (X=T) equals to a factor of 1.2, 1.4, and 1.6, respectively. The omission of the additional functions which occur in the CV-type basis sets would lead to savings with respect to the overall calculation time especially if higher order increments could be described with a smaller basis size since these consume most of the CPU time. From a chemical point of view such an approach would be appropriate for the valence only contributions but for the core-valence treatment one needs to examine such an approximation very carefully with respect to its accuracy. Therefore in Fig. 3.36 the CCSD(T) and MP2 third- and fourth-order incremental energy differences between the cc-pVXZ and the corresponding core-valence basis sets cc-pCVXZ of the discussed test systems are presented for the valence-only and core-valence contributions separately.

The fourth-order incremental valence-only contributions are almost identically in size

in both cc-pVXZ and cc-pCVXZ basis sets for the systems of this study. Interestingly the same finding holds also for the difference of fourth-order core-valence correlation contributions leading to the conclusion that a forth-order treatment with the cc-pVXZ basis set in comparison to its core-valence analogous basis set would only cause little loss in accuracy. The valence-only increments could also be described at third-order with the omission of core-valence functions since the greatest aberration is just 0.3 kcal/mol for the titanium complex at CCSD(T) theory level. This does not hold in general for the core-valence incremental energy differences with respect to the basis set change when the third-order expansion is considered. Even though very good agreement is found for all cases besides the titanium complex (as indicated by the position of all points close to the line which marks zero), an approximate third-order CCSD(T)/cc-pVTZ description in case of $\text{Ti}_2\text{Cl}_4\text{S}_2$ ($X=\text{T}$) would lead to an unacceptable high error of about 2 kcal/mol.

The outcomes of these considerations are very encouraging but a general advise regarding the question at which order one should start to use a smaller basis set can not be formulated at this stage because not enough systems have been considered so far.

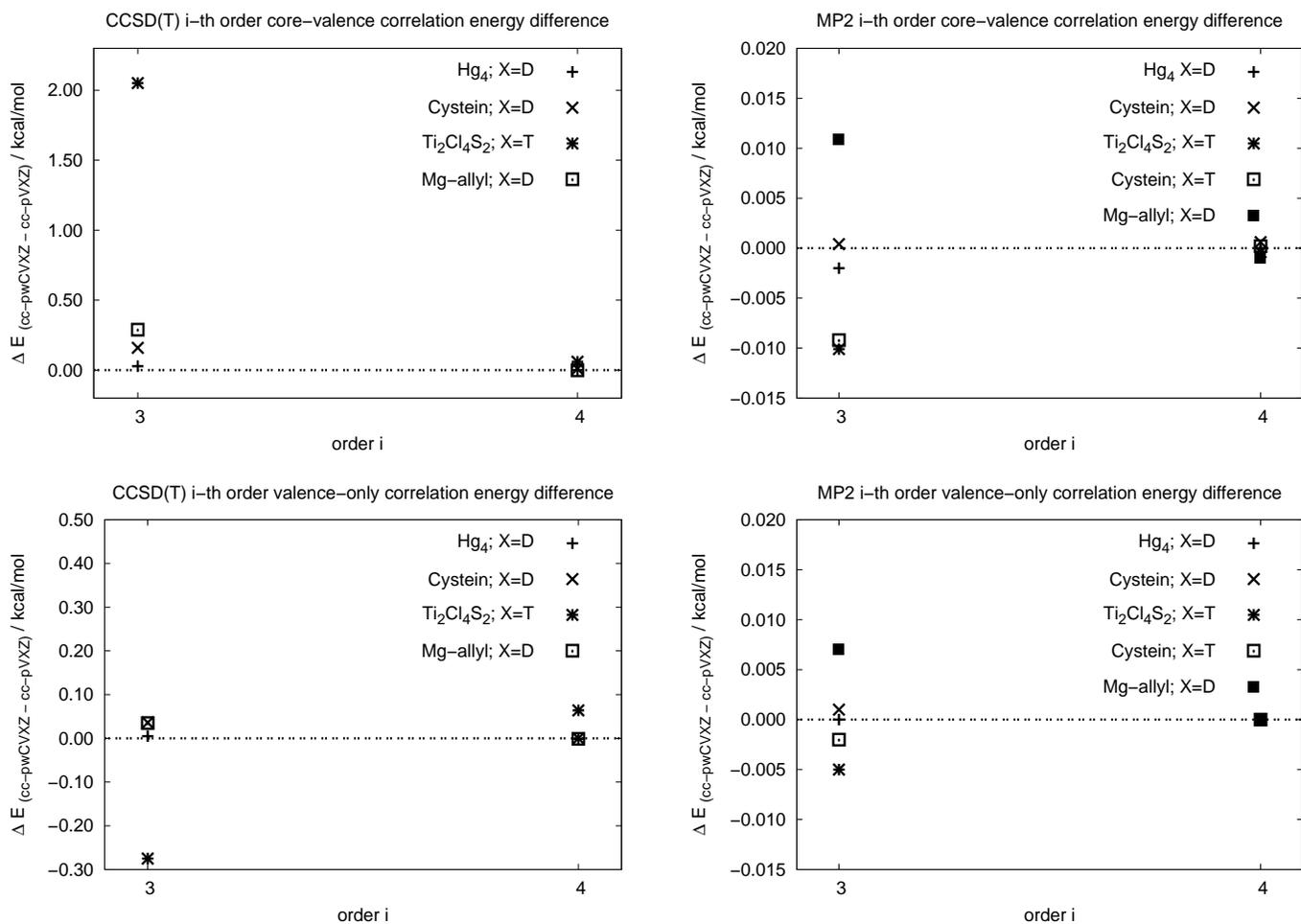


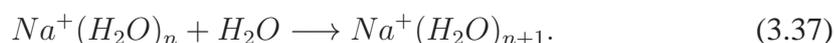
Figure 3.36: Incremental third- and fourth-order CCSD(T) (left hand side) and MP2 (right hand side) correlation energies differences in kcal/mol between cc-pwCVXZ and cc-pVXZ basis sets. For detailed analysis of data regarding cystein molecule see the Appendix C.

3.4.2.2 Core-correlation influence on interaction energies

The core-valence correlation is especially important for systems containing atoms with easily polarizable cores. Alkali and alkaline-earth atoms exhibit sizable core-correlation effects which have to be considered for an accurate geometrical and energetic description. Therefore Eq. 2.68 is applied to investigate the magnitude of the core-correlation for the hydration of the singly charged sodium cation according to:



The interaction energies ΔE_{int} were evaluated for $n=2,3$ and 4. The so-called sequential interaction energies ΔE_{seq} with $n=2$ and 3 have also been considered according to the reaction scheme:



The sodium cation water clusters $Na^+(H_2O)_n$ ($n=2,3$ and 4) were optimized using the RI-BP86/TZVPP gradient-corrected density functional theory method [145, 146] in TURBOMOLE 5.10 [97, 98], with D_{2d} , D_3 and S_4 symmetry constraints applied for $n=2,3$ and 4, respectively, the optimized geometries are shown in Fig 3.37.

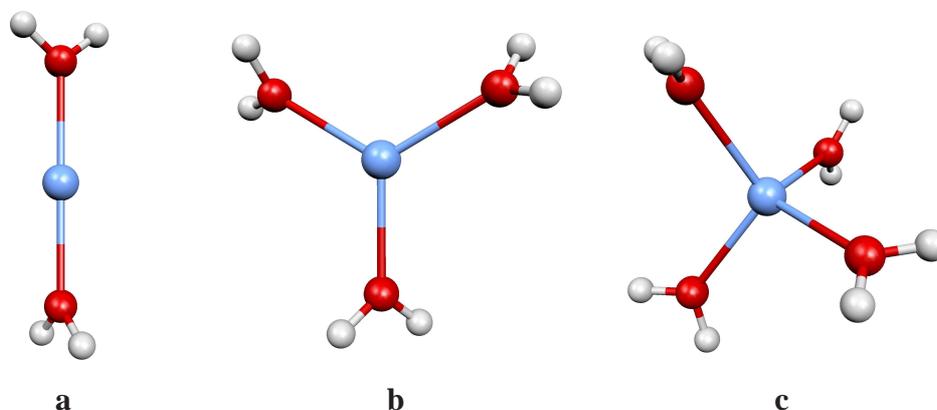


Figure 3.37: Optimized structures of di- to tetra-hydrate sodium complexes $Na^+(H_2O)_n$ ($n=2,3$ and 4) (a-c).

Stationary points were characterized by analyzing the Hessian matrix in order to confirm that the obtained structures correspond to minimum energy conformations on the respective potential energy surfaces.

The n valence domains consist of the 4 valence orbitals of each of the n water molecules. The [Ne]-shell of the sodium cation is correlated at all times and contributes one core domain, whereas each of the n water molecules contributes one core

domain consisting of the oxygen 1s shell. The incremental expansion for the n valence and $n + 1$ core domains were truncated at second-order. The calculations of a single water molecule and the isolated sodium cation needed to determine ΔE_{int} and ΔE_{seq} were performed with standard MP2 and CCSD(T) schemes.

Correlation-consistent core-valence basis sets of double- and triple- ζ quality have been employed for the sodium cation and on the oxygen atom, diffuse functions were also incorporated on the hydrogen and oxygen atoms to account for the water polarizability effects.

The incremental MP2 and CCSD(T) interaction and sequential interaction energies obtained with different basis sets are presented in Tables 3.21 and 3.22, respectively. The comparison of the incremental results with the canonical ones show remarkable good agreement when the expansion is truncated already at second-order. The largest error with respect to the reference calculations is just 0.07 kcal/mol. This is for the smallest water cluster with $n = 2$ not surprising since the valence-only incremental correlation energy in this case equals the canonical outcome since the expansion reaches the highest possible order. However, for the core-valence correlation treatments as well as for $n > 2$ in the valence-only calculations, this is not the case. The related increase of errors with respect to the canonical data set can be nevertheless regarded as negligible, since as mentioned above the highest aberration equals to just -0.07 kcal/mol. Therefore one can conclude that the application of Eq. 2.68 leads to highly accurate results not only for absolute energies, but also when relative energies are considered. This finding holds for both the CCSD(T) and the MP2 incremental treatment.

The difference between the results of the valence-only and the total energy calculations illustrates the magnitude of errors due to the frozen-core approximation and shows at least for these simple test reactions the importance of the core correlation treatment. Omitting the core correlation effects at CCSD(T) level leads to an underestimation of $\Delta E_{\text{int}}^{\text{V}}$ as well as $\Delta E_{\text{seq}}^{\text{V}}$ of about 0.5 up to even 5 kcal/mol, corresponding to a relative error of 3 to 12%. Very similar outcomes are obtained at the MP2 level, where the failure to consider the $E_{\text{int}}^{\text{c,cv}}$ and $E_{\text{seq}}^{\text{c,cv}}$ contributions to interaction energies causes 3 to 12% deviation from the total interaction energies. The basis set improvement from double to triple- ζ quality does not affect the convergence of the incremental results. However a somewhat significant influence is observed on the total interaction energies, including an increase of the core correlation contributions by about 1 to 3%.

This study clearly demonstrated that the core-valence correlation has a significant contribution to the total interaction energies as well as to the sequential interaction energies for the hydration of the singly charged sodium cation. It could also be shown that a computationally attractive second-order expansion is sufficient to obtain chemically accurate interaction and sequential energies.

Table 3.21: Comparison of the incremental valence-only (v), core-valence (cv) and total interaction energies (in kcal/mol) with the canonical calculations for $\text{Na}^+(\text{H}_2\text{O})_n$ complexes. Correlation energies obtained at MP2 and CCSD(T) level of theory with the cc-pCVXZ (X=D,T) basis set series at Na^+ , aug-cc-pCVXZ (X=D,T) basis sets at O and aug-cc-pVXZ (X=D,T) basis sets at H. Incremental expansion truncated at second-order ($t_{\text{con}}=3$ Bohr, dsp=3, dsp=4 for n=4, $E_{\text{thres}}=1\text{e-}5$ au).

$$E' = \Delta E_{\text{int}}^{\text{c,cv}} / \Delta E_{\text{int}}^{\text{total}}$$

	MP2				CCSD(T)			
	$\Delta E_{\text{int}}^{\text{v}}$	$\Delta E_{\text{int}}^{\text{total}}$	$\Delta E_{\text{int}}^{\text{c,cv}}$	E'	$\Delta E_{\text{int}}^{\text{v}}$	$\Delta E_{\text{int}}^{\text{total}}$	$\Delta E_{\text{int}}^{\text{c,cv}}$	E'
	kcal/mol			%	kcal/mol			%
X=D, n=2								
inc.	42.540	44.437	1.897	4	42.370	44.392	2.022	5
can.	42.540	44.453	1.913		42.370	44.410	2.041	
error	0	-0.016	-0.016		0	-0.019	-0.019	
X=T, n=2								
inc.	42.573	45.129	2.556	6	42.441	45.202	2.761	6
can.	42.573	45.152	2.579		42.441	45.231	2.790	
error	0	-0.024	-0.023		0	-0.029	-0.029	
X=D, n=3								
inc.	57.633	61.113	3.480	6	57.416	61.168	3.753	6
can.	57.634	61.159	3.526		57.413	61.206	3.793	
error	0	-0.046	-0.046		0.002	-0.038	-0.040	
X=T, n=3								
inc.	57.158	61.680	4.523	7	56.984	61.938	4.955	8
can.	57.158	61.748	4.590		56.981	61.999	5.017	
error	0	-0.068	-0.067		0.002	-0.060	-0.063	
X=D, n=4								
inc.	73.961	77.908	3.947	5	73.800	78.047	4.247	5
can.	73.961	77.950	3.990		73.791	78.067	4.276	
error	0	-0.043	-0.043		0.009	-0.020	-0.029	

Table 3.22: Comparison of the incremental valence-only (v), core-valence (cv) and total interaction energies (in kcal/mol) with the canonical calculations for $\text{Na}^+(\text{H}_2\text{O})_n$ complexes. Correlation energies obtained at MP2 and CCSD(T) level of theory with the cc-pCVXZ (X=D,T) basis set series at Na^+ , aug-cc-pCVXZ (X=D,T) basis sets at O and aug-cc-pVXZ (X=D,T) basis sets at H. Incremental expansion truncated at second-order ($t_{\text{con}}=3$ Bohr, dsp=3, dsp=4 for n=4, $E_{\text{thres}}=1\text{e-}5$ au).

$$E' = \Delta E_{\text{seq}}^{\text{c,cv}} / \Delta E_{\text{seq}}^{\text{total}}$$

	MP2				CCSD(T)			
	$\Delta E_{\text{seq}}^{\text{v}}$	$\Delta E_{\text{seq}}^{\text{total}}$	$\Delta E_{\text{seq}}^{\text{c,cv}}$	E'	$\Delta E_{\text{seq}}^{\text{v}}$	$\Delta E_{\text{seq}}^{\text{total}}$	$\Delta E_{\text{seq}}^{\text{c,cv}}$	E'
	kcal/mol			%	kcal/mol			%
X=D, n=2								
inc.	15.094	16.676	1.582	9	15.046	16.777	1.731	10
can.	15.094	16.707	1.613		15.044	16.796	1.752	
error	0	-0.031	-0.030		0.002	-0.019	-0.021	
X=T, n=2								
inc.	14.585	16.552	1.967	12	14.543	16.737	2.194	13
can.	14.585	16.596	2.011		14.540	16.768	2.227	
error	0	-0.044	-0.044		0.003	-0.031	-0.034	
X=D, n=3								
inc.	16.327	16.795	0.467	3	16.385	16.879	0.494	3
inc.	16.327	16.791	0.464		16.378	16.861	0.483	
error	0	0.004	0.003		0.007	0.018	0.011	

3.4.2.3 Domain-specific basis sets

Although the emphasis of this work is the investigation of the accuracy of equation 2.68 for CCSD(T) and MP2 energies, a study regarding the efficiency of the incremental domain-specific basis set approach [52, 55] has also been carried out. Within this approximation a second smaller basis set is used to describe the environment of the domains. For the studied $\text{Na}^+(\text{H}_2\text{O})_4$ complex, the smaller basis set is the STO-3G minimal basis at the hydrogen atoms and the 6-31G basis at the oxygen atoms. The environment basis is set on three water molecules at first- and two water molecules at second-order. In Table 3.23 the efficiency of the domain-specific basis set is demonstrated for the valence-only correlation for the $\text{Na}^+(\text{H}_2\text{O})_4$ complex using different basis sets of the Dunning aug-cc-pVXZ series. The CPU times are relative to the standard CCSD(T) calculation with MOLPRO using C_2 symmetry. The computational saving is about a factor of 4 for the triple and quadruple ζ -basis sets if the full S_4 symmetry of the complex is exploited. The wall time could be even further reduced by exploiting the parallelity of the incremental scheme [52, 55]. Considering the accuracy of the domain-specific basis set approach errors below 0.04 kcal/mol for all applied basis sets are found, which is a very satisfactory accuracy.

Table 3.23: Comparison of computational savings for the CCSD(T)/aug-cc-pVXZ valence only correlation energy for the $\text{Na}^+(\text{H}_2\text{O})_4$ molecule with respect to canonical calculation. The errors with respect to the canonical calculation are given in kcal/mol.

X	rel. time	rel. time	error
	inc. (C_1) / can. (C_2)	inc. (S_4) / can. (C_2)	
	%	%	kcal/mol
D	129	43	0.03
T	72	25	-0.04
Q	80	26	-0.03

3.4.2.4 Variation of the domain size

The analysis of the influence of different domain sizes on computational time for the $\text{Na}^+(\text{H}_2\text{O})_4$ complex is presented in Tab. 3.24. The relative computation times among different domain size parameters (dsp) are compared and also the percentage of calculation time relative to a given expansion order is shown. The accuracy of the CCSD(T)/6-31G* correlation energies and the amount of the incremental energy calculations $\mathcal{N}(i)$ with respect to the expansion order are also given in Tab. 3.24. A distance screening reducing the number of individual calculations was not applied be-

cause only the influence of dsp parameter on computational time will be considered herein.

Table 3.24: Comparison of the relative computational time for different domain sizes for the calculation of the CCSD(T)/6-31G* correlation energy of $\text{Na}^+(\text{H}_2\text{O})_4$. The errors are with respect to the canonical calculation (in kcal/mol). Number of individual calculations $\mathcal{N}(i)$ at a given expansion order i and percentage of calculation time relative to a given order are also given.

dsp	$\mathbb{D} \times \text{Orb.}$	i	error kcal/mol	$\mathcal{N}(i)$	time at i %	rel. time %
1	16x1	1	264.392	16	0.5	
		2	-28.114	120	3.8	
		3	1.315	560	20.3	
		4	-0.008	1820	75.4	14600 (C_1)
1	16x1	1	264.392	4	0.6	
		2	-28.114	32	4.5	
		3	1.315	140	20.8	
		4	-0.008	464	74.1	3412 (S_4)
3	2x2	1	42.106	5	26	
	3x4	2	0.009	10	74	100 (C_1)
4	4x4	1	0.740	4	29	
		2	-0.007	6	71	79 (C_1)
4	4x4	1	0.740	1	27	
		2	-0.007	2	73	26 (S_4)
5	2x4	1	0.624	3	32	
	1x8	2	-0.003	3	68	67 (C_1)

The investigated different domain size parameters are equal 1, 3, 4 and 5 and divide the system into 16, 5, 4 and 3 valence domains \mathbb{D} with a different amount of localized orbitals grouped into the domains. Exemplary the dsp value of 3 lead to a partitioning of two domains with two occupied localized orbitals and three domains with 4 orbitals, as indicated in Table 3.24 within the short notation 2x2 and 3x4 respectively. The re-

sults show that the optimum with respect to a given target accuracy and with respect to computational time is achieved for $\text{dsp}=5$. But the calculation with $\text{dsp}=4$ is only little slower and it has the great advantage that the partitioning preserves the symmetry properties of the molecule. The calculation in S_4 symmetry when dsp equals 4 is therefore the cheapest calculation at all. The worst scenario appears if a partitioning into single orbital domains is considered. In order to obtain similar accuracy as for $\text{dsp}=3,4$ and 5 with an aberration of just one hundredth of a kcal/mol the incremental expansion has to be evaluated up to the fourth-order. In these cases the huge number of individual calculations which have to be performed to obtain the desired accuracy cause the calculation to become the most expensive one. Even when the S_4 symmetry is exploited for $\text{dsp}=1$ the calculation time is by a factor of 3000 longer than the computation with $\text{dsp}=3$.

Chapter 4

Conclusion

In this contribution we proposed among others two approximations to the site-site function counterpoise (SSFC) method and one to the second-order Valiron-Mayer function counterpoise (VMFC) scheme, based on truncated ghost basis sets used for the monomer calculations.

Within one procedure we employ as a criterion for a systematic choice of the truncated orbital space, a distance threshold R acting on the geometry of a given structure. Within the second approach, the reduced basis set is obtained via a systematically analysis of the magnitude of the overlap integrals.

The calculations carried out for a water and methanol cluster series evidence convincingly that with only little loss in accuracy great savings with respect to the computational requirements can be achieved with both proposals - the SSFC(R) and the SSFC(S) - scheme. To a certain extent also the corresponding VMFC(2)(R) scheme provides advantages over the full approach yielding at the same time highly accurate results.

Furthermore the approximate approaches enable BSSE correction calculations for larger cluster sizes where the standard procedure fails to proceed. In addition we employed the domain-specific basis set approach within the framework of the incremental scheme to evaluate the total correlation energy of the clusters and demonstrated that this method yields very accurate results with respect to the canonical calculations.

We present accurate BSSE corrected stabilization energies for water clusters up to 20 subunits with the aug-cc-VXZ ($X=D,T$) basis sets at the MP2, CCSD and even CCSD(T) (with n up to 18) levels of theory.

The approximate SSFC(R) and SSFC(S) schemes are robust methods to account for BSSE corrections. With the variable use of the distance threshold or the overlap threshold one is able to control the desired accuracy as well as the computational resources needed to carry out the calculation.

Appendix A

MP2 and CCSD incremental interaction energies

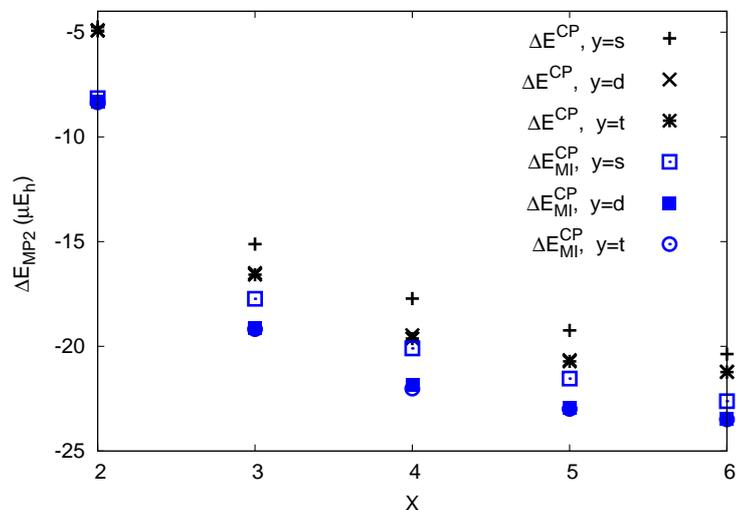


Figure A.1: Convergence behavior of the MP2 interaction energy of He_2 at equilibrium distance for the y-aug-cc-pVXZ (y=s,d,t and X=D,T,Q,5,6) basis sets.

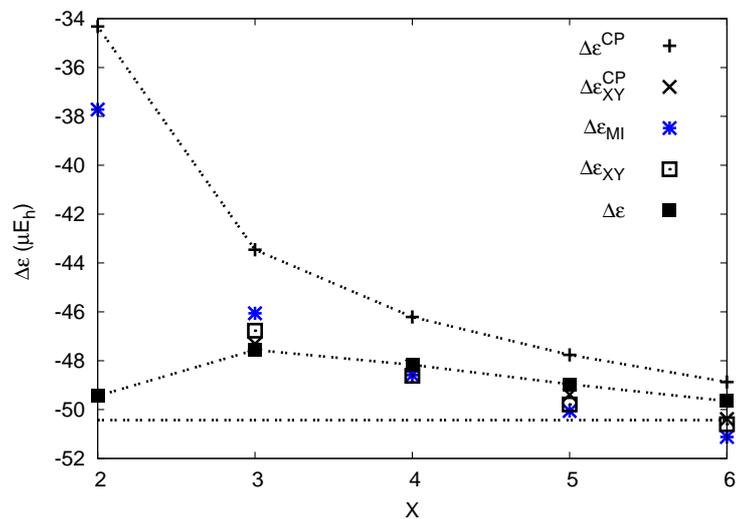


Figure A.2: MP2 correlation energy contribution to the interaction energy of He_2 as a function of the cardinal number X in the aug-cc-pVXZ basis set series. The horizontal dashed line $\Delta E_{56}^{\text{CP}}$ marks the estimated basis set limit.

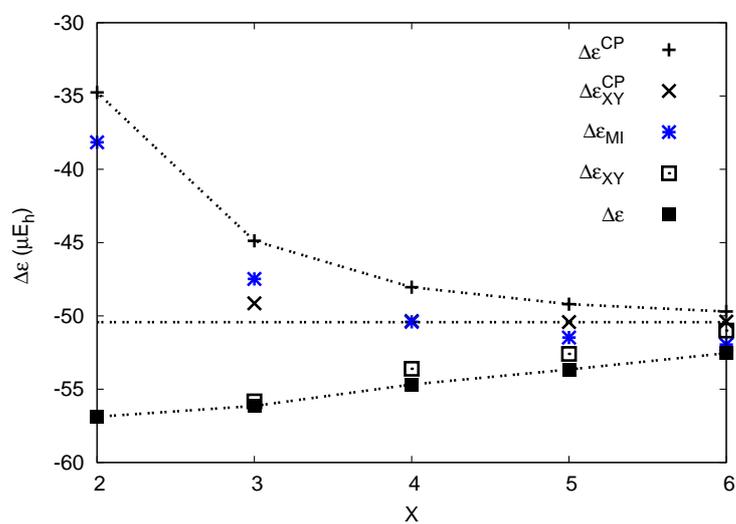


Figure A.3: As Fig. A.2, but for the d-aug-cc-pVXZ basis set series.

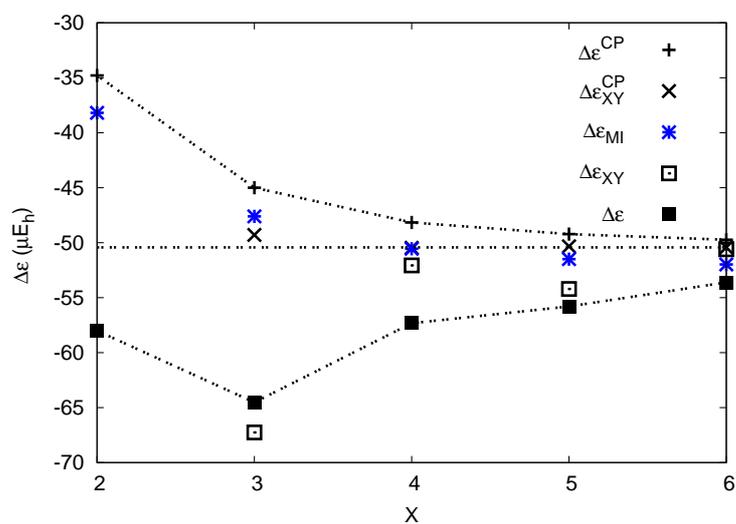


Figure A.4: As Fig. A.2, but for the t-aug-cc-pVXZ basis set series.

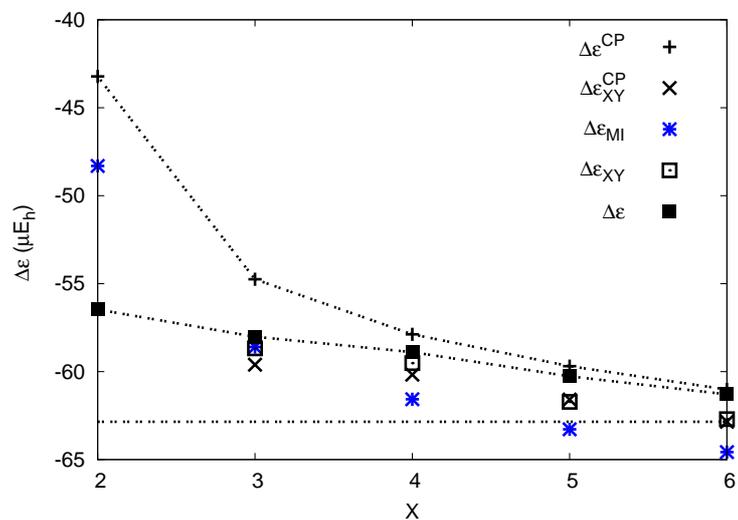


Figure A.5: CCSD(T) correlation energy contribution to the interaction energy of He_2 as a function of the cardinal number X in the aug-cc-pVXZ basis set series. The horizontal dashed line $\Delta E_{56}^{\text{CP}}$ marks the estimated basis set limit.

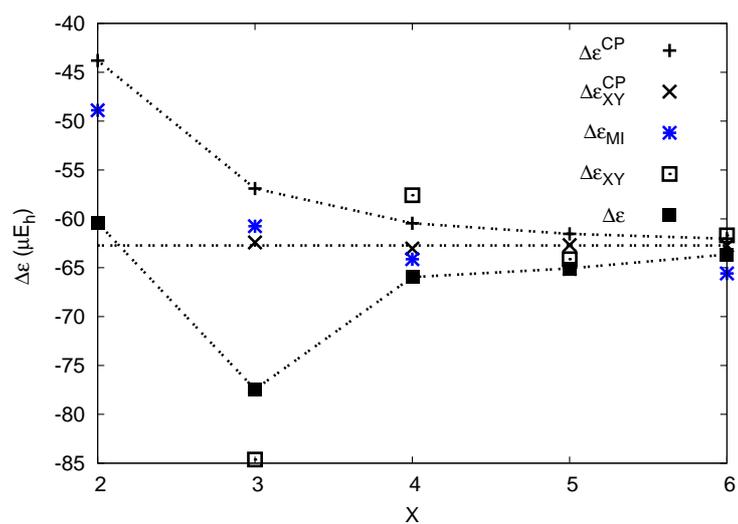


Figure A.6: As Fig. A.5, but for the t-aug-cc-pVXZ basis set series.

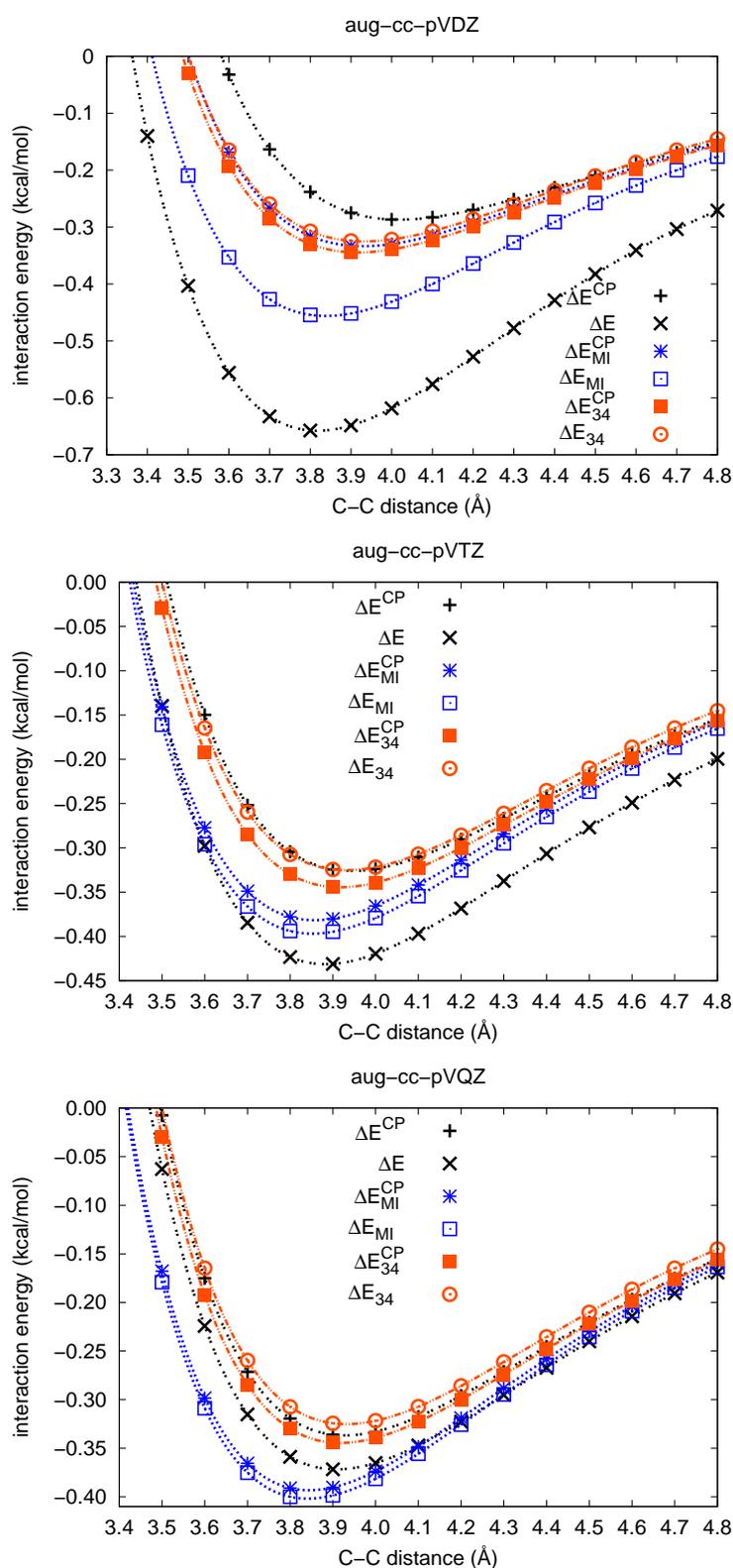
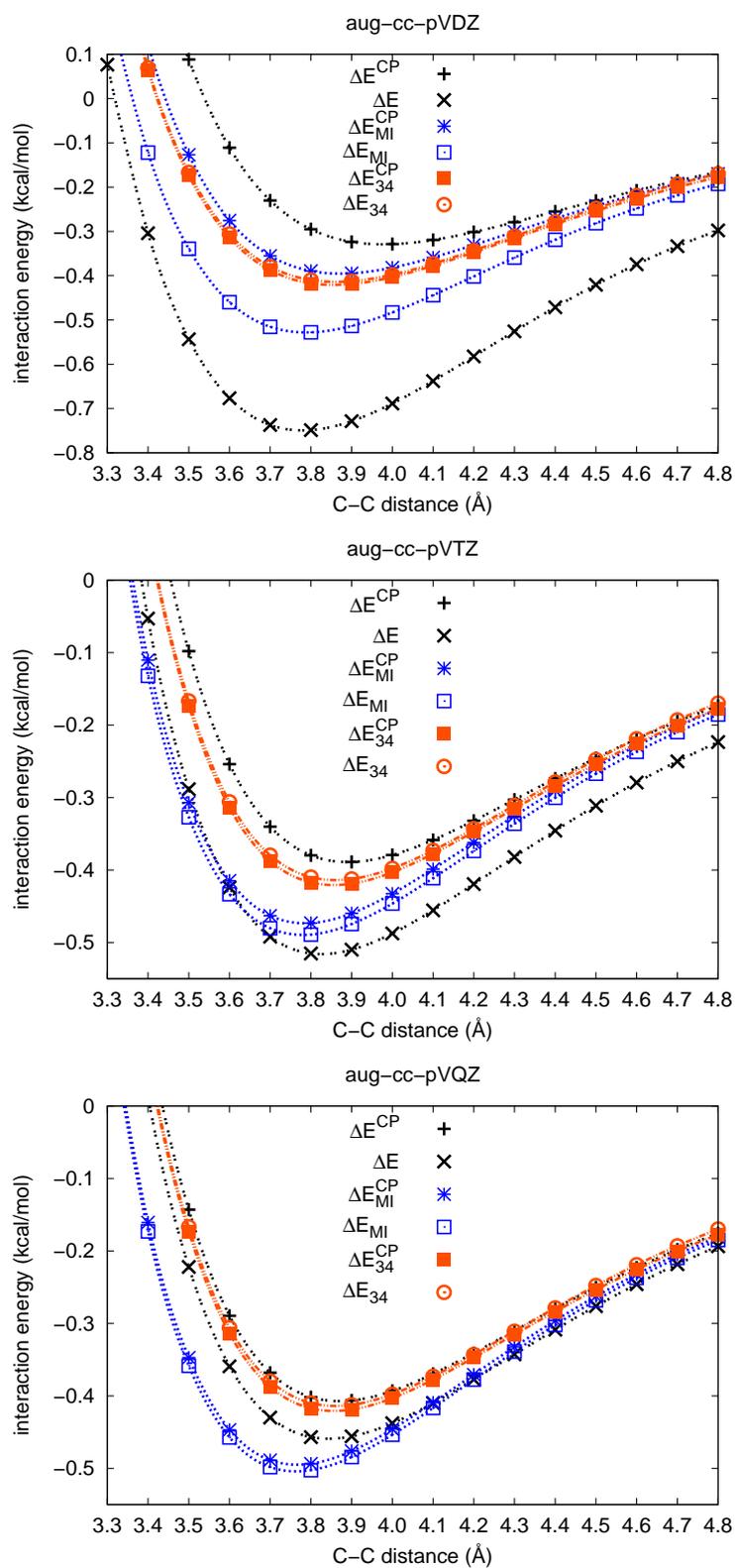


Figure A.7: Potential energy curves of the interaction energy for the CH₄ dimer, calculated with aug-cc-pVXZ (X=D, T and Q) basis sets at the CCSD level of theory as a function of the intermolecular C-C distance. The 3-4 extrapolated curves ΔE_{34}^{CP} and ΔE_{34} are based on CP-corrected and uncorrected correlation energies, respectively.

**Figure A.8:** As Fig. A.7, but for the MP2 level of theory.

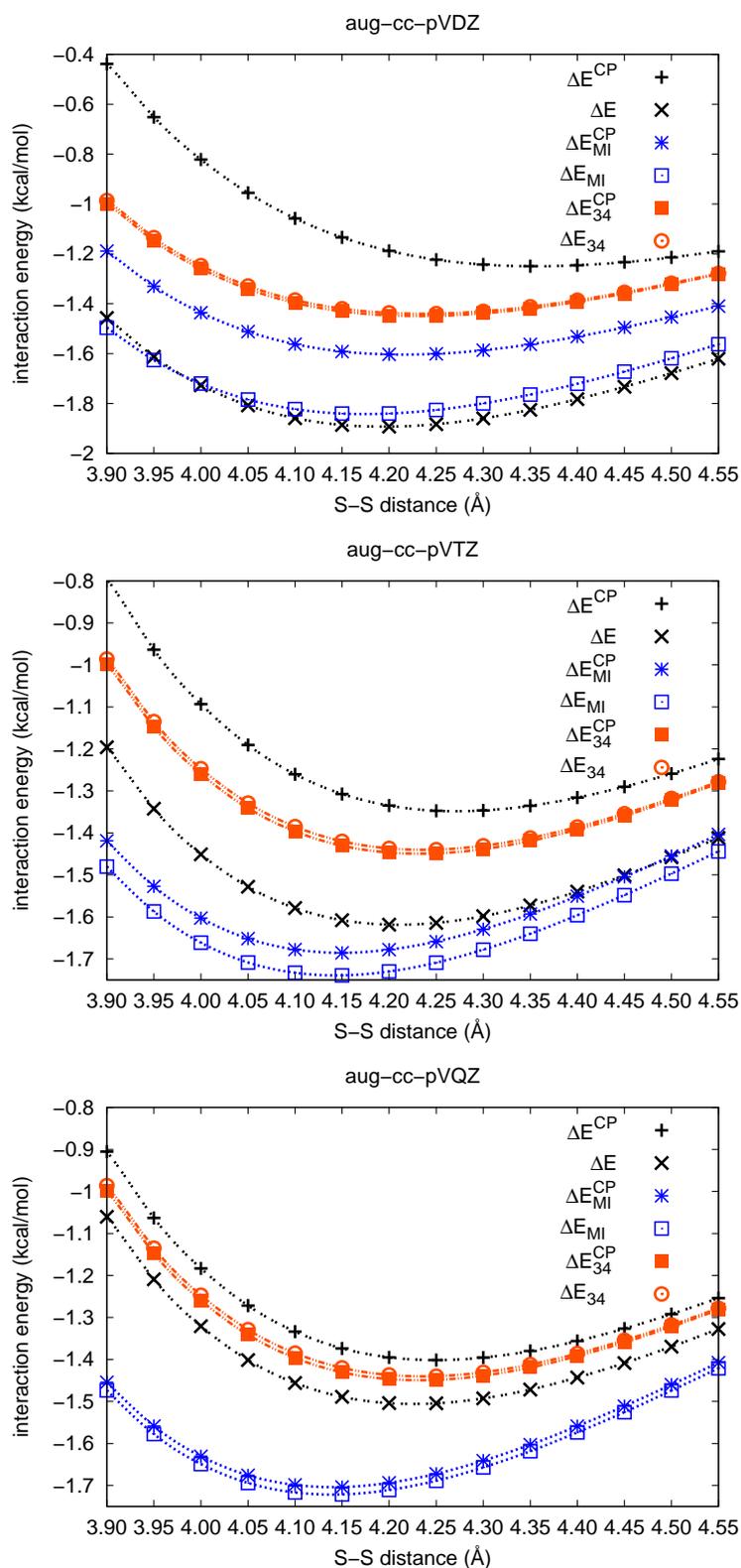
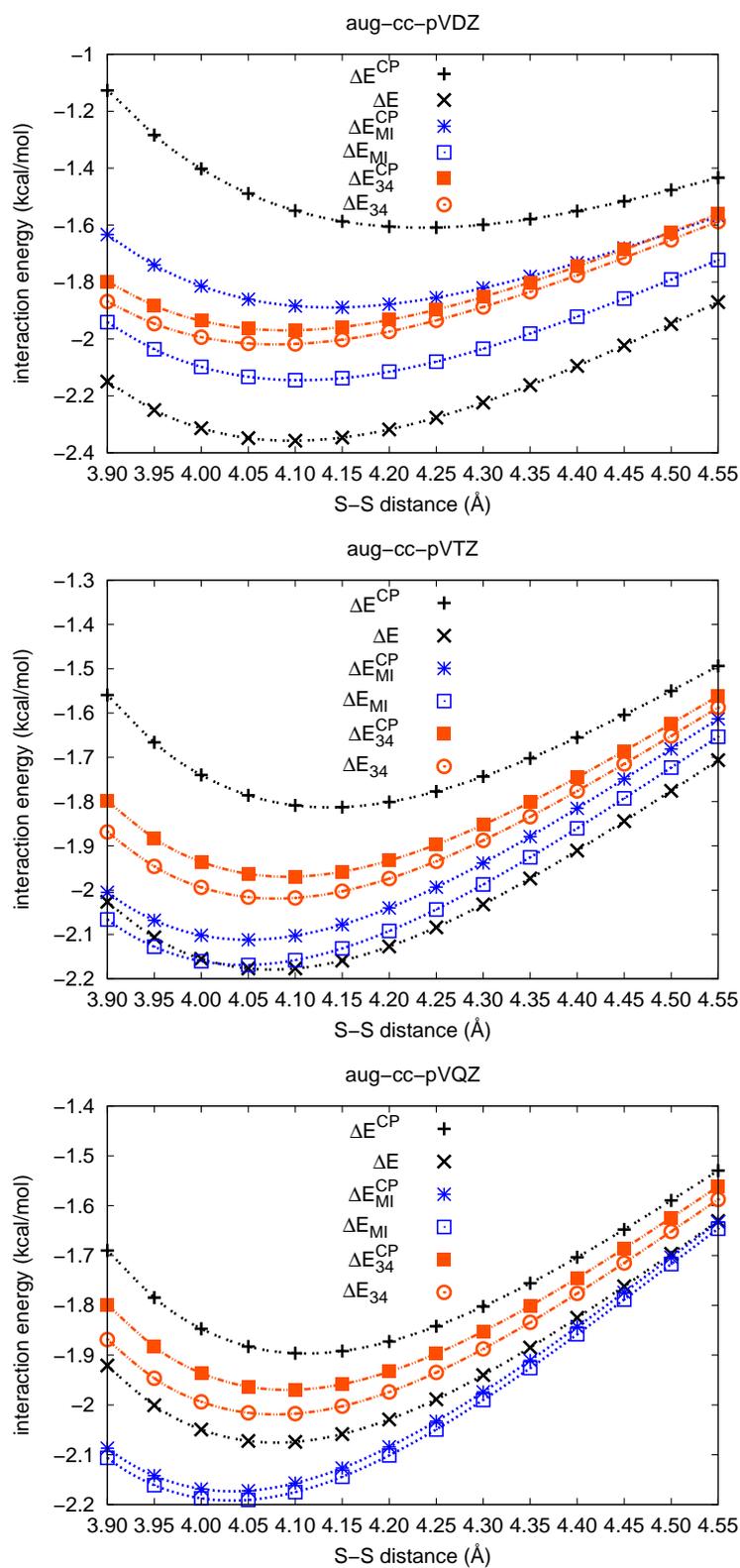


Figure A.9: Potential energy curves of the interaction energy for the H_2S dimer, calculated with aug-cc-pVXZ ($X=D, T$ and Q) basis sets at the CCSD level of theory as a function of the intermolecular C-C distance. The 3-4 extrapolated curves ΔE_{34}^{CP} and ΔE_{34} are based on CP-corrected and uncorrected correlation energies, respectively.

**Figure A.10:** As Fig. A.9, but for the MP2 level of theory.

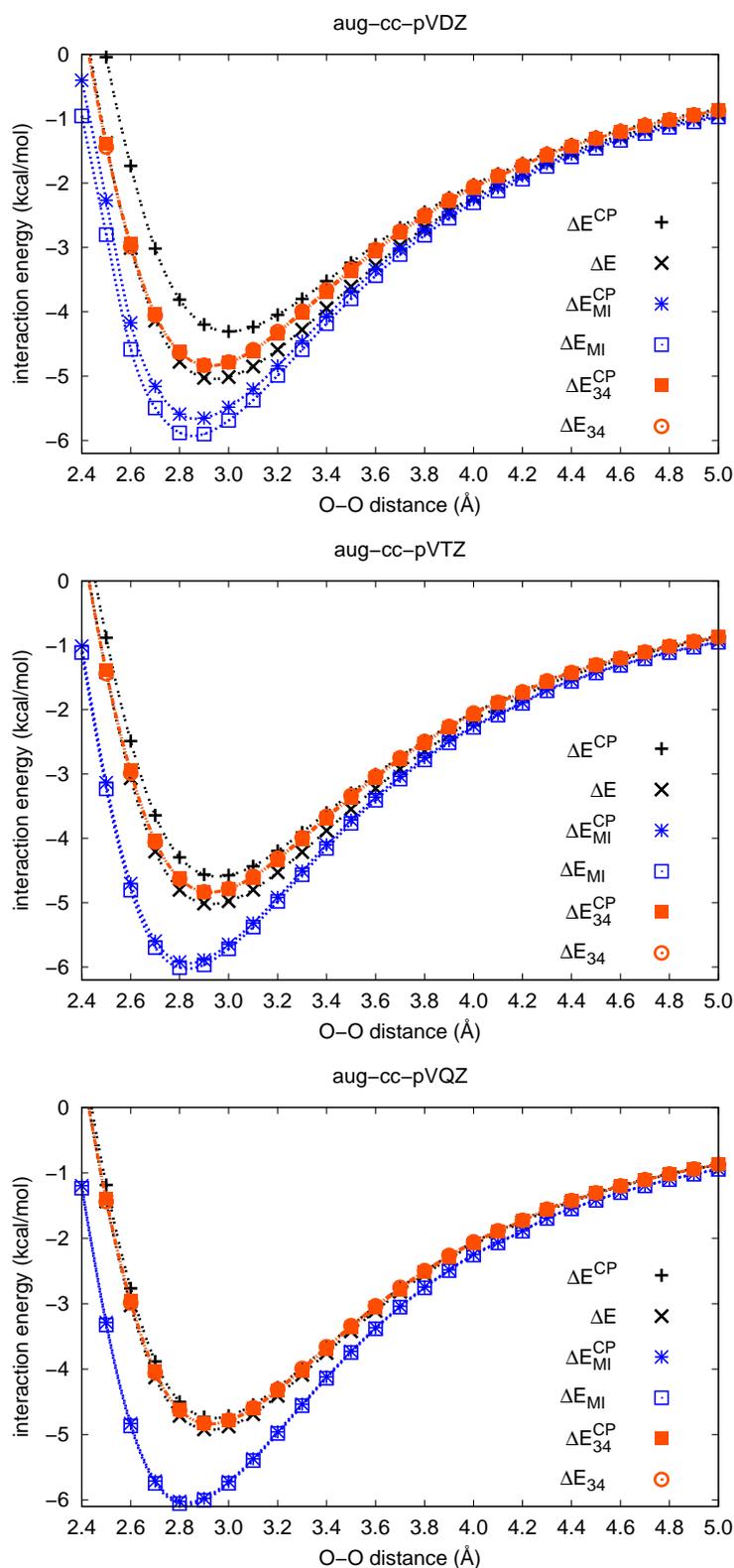


Figure A.11: Potential energy curves of the interaction energy for the H₂O dimer, calculated with aug-cc-pVXZ (X=D,T and Q) basis sets at the CCSD level of theory as a function of the intermolecular C-C distance. The 3-4 extrapolated curves ΔE_{34}^{CP} and ΔE_{34} are based on CP-corrected and uncorrected correlation energies, respectively.

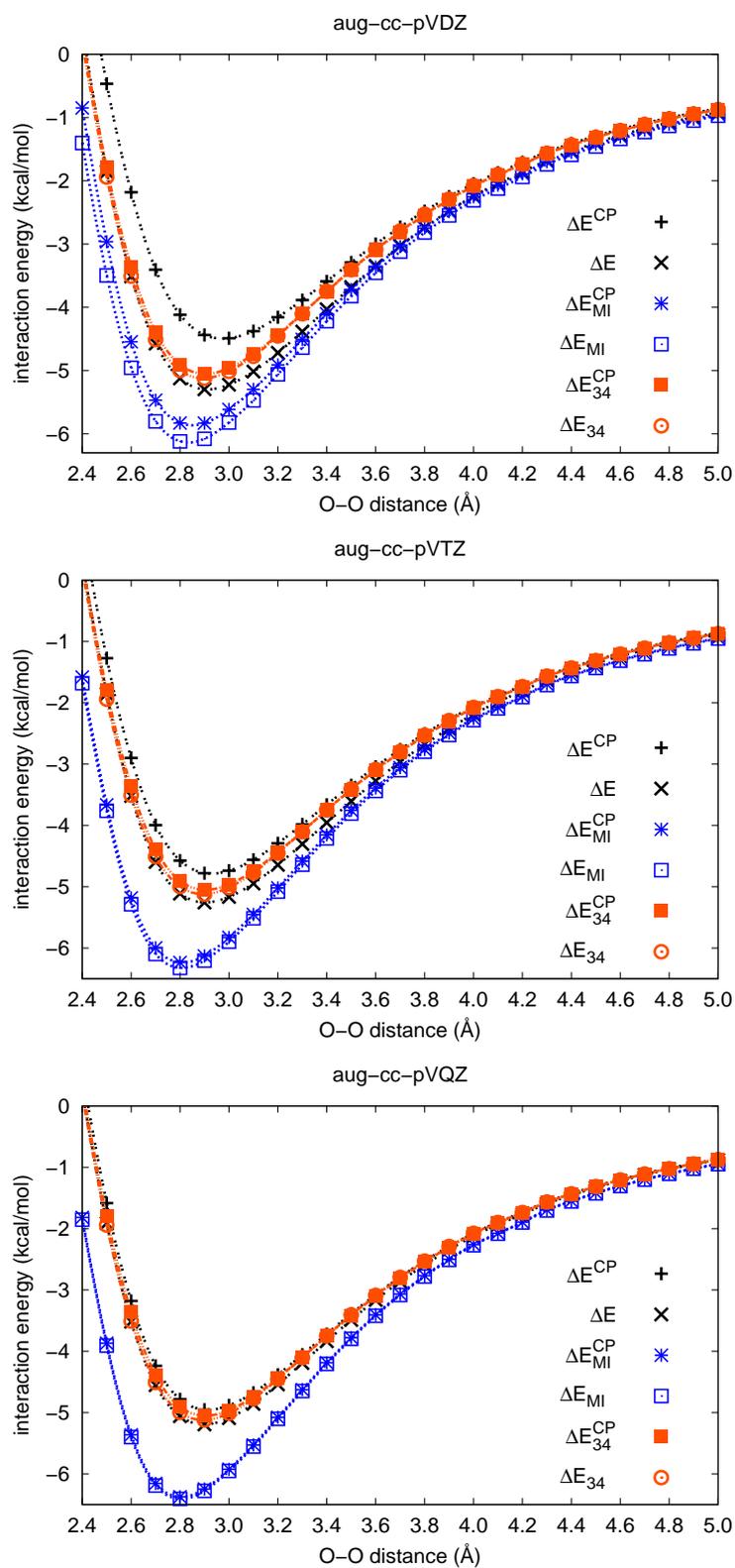


Figure A.12: As Fig. A.11, but for the MP2 level of theory.

Appendix B

**SSFC(R), VMFC(2)(R) and SSFC(S)
results at the MP2 and CCSD levels of
theory**

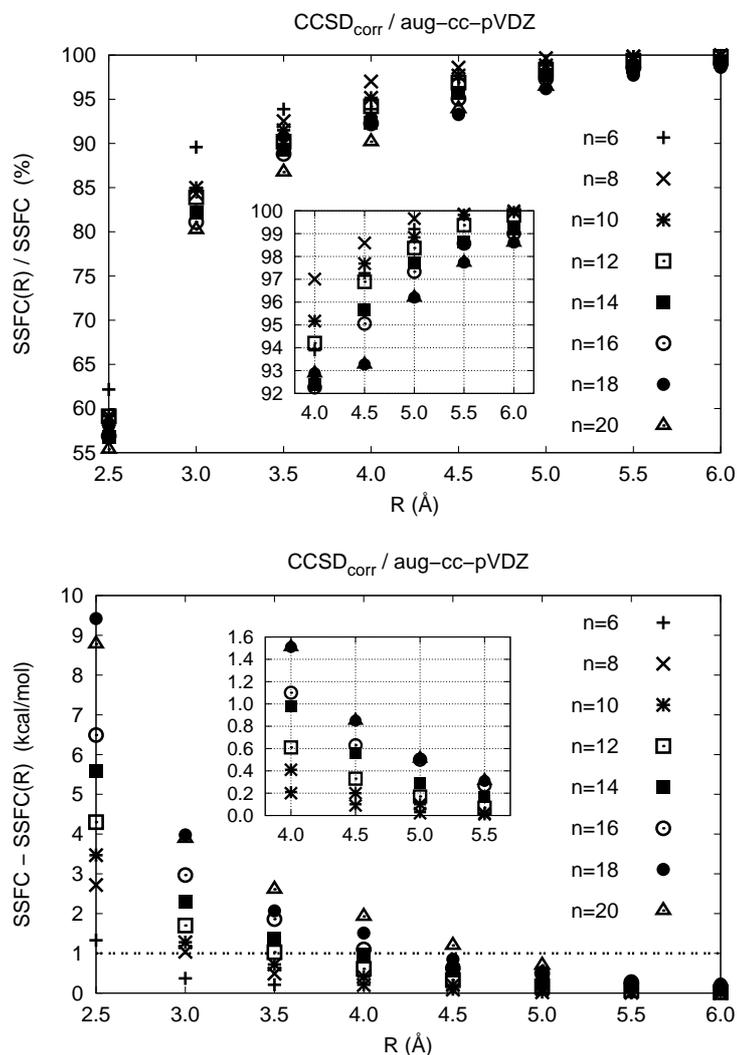


Figure B.1: BSSE correction (in percentage) depending on the distance threshold R (Å) with respect to the SSFC method and the corresponding absolute deviation from SSFC calculated for the water cluster series $(\text{H}_2\text{O})_n$ $n=6,8,\dots,20$ ($n=6$ boat structure) at the $\text{CCSD}_{\text{corr}}$ level of theory.

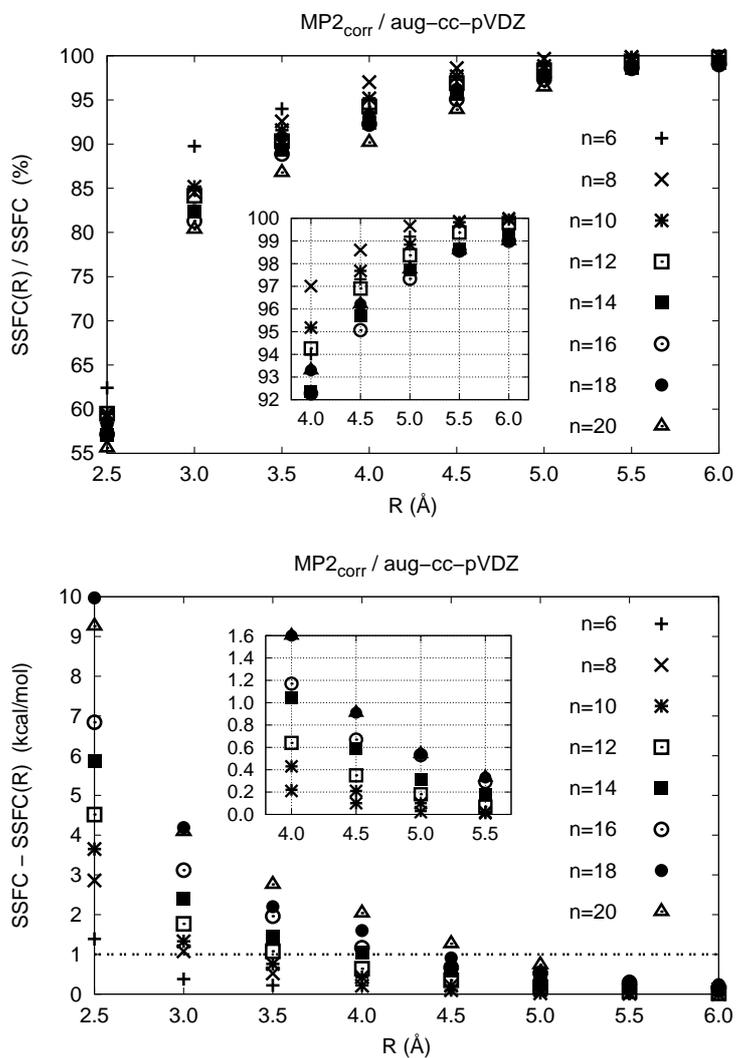


Figure B.2: As Fig. B.1 but for the MP2_{corr} level of theory.

Table B.1: Approximate BSSE second-order VMFC correction with respect to different distance thresholds at the aug-cc-pVDZ/CCSD level ($\Delta E = \text{VMFC}(2) - \text{VMFC}(2)(R)$).

R Å	VMFC(2)(R) n=8			VMFC(2)(R) n=10			VMFC(2)(R) n=12		
	kcal/mol	%	ΔE	kcal/mol	%	ΔE	kcal/mol	%	ΔE
2	0.18	24.61	0.54	0.32	35.02	0.59	0.40	36.50	0.70
2.5	0.53	73.21	0.19	0.65	71.21	0.26	0.77	69.88	0.33
3	0.63	87.12	0.09	0.79	87.15	0.12	0.93	85.18	0.16
3.5	0.68	94.48	0.04	0.83	91.97	0.07	0.98	89.79	0.11
4	0.71	98.51	0.01	0.87	96.20	0.03	1.03	94.49	0.06
4.5	0.72	99.57	0.003	0.89	98.02	0.02	1.06	97.02	0.03
5	0.72	99.78	0.002	0.90	99.47	0.005	1.08	98.54	0.02
5.5	0.72	100	0	0.91	99.92	0.001	1.09	99.52	0.01
6				0.91	100	0	1.09	99.86	0.002
∞	0.72	100	0	0.91	100	0	1.10	100	0

Table B.7: SSFC(R) BSSE correction (in kcal/mol) with respect to R for three different methanol clusters $(\text{CH}_3\text{OH})_n$ and the deviation (in kcal/mol and %) from the full SSFC method for n=8.

R Å	SSFC(R) kcal/mol	SSFC(R) %	error kcal/mol	SSFC(R) kcal/mol	SSFC(R) %	error kcal/mol	
		CCSD/aug-cc-pVDZ			MP2/aug-cc-pVDZ		
n=8							
4.0	12.12	85.50	2.05	12.76	85.54	2.16	
4.5	13.04	92.05	1.13	13.73	92.05	1.19	
5.0	13.58	95.84	0.59	14.30	95.83	0.62	
5.5	13.92	98.22	0.25	14.65	98.23	0.26	
∞	14.17	100	0.00	14.92	100	0.00	
n=12							
4.0	21.17			22.30			
4.5	22.89			17.19			
5.0	23.99			17.95			
n=16							
4.0	34.44			36.33			

Table B.2: CCSD and MP2 BSSE corrected and uncorrected stabilization energies of the water cluster (H₂O)₆ (bag structure) and BSSE corrections obtained with different schemes for the basis set series aug'-cc-pVXZ (X=D,T,Q and 5). ($\Delta E^{\text{BSIE}} = D^{\text{CBS(Q,5)}} - D^{\text{aug'-cc-pVXZ}}$, all energies in kcal/mol, values in parentheses in %).

X	D	T	Q	5	CBS(Q,5)
CCSD					
D	-36.52	-38.29	-39.24	-39.33	-39.43
ΔE^{BSIE}	2.91	1.15	0.19	0.11	0.00
SSFC (R=2.5)	3.01 (61.22)	1.58 (69.45)	0.62 (68.34)	0.25	0
SSFC (R=3.5)	4.48 (91.15)	2.12 (93.45)	0.84 (93.14)	0.34	0
SSFC (R=4.5)	4.89 (99.54)	2.26 (99.56)	0.90 (99.55)		
SSFC	4.92 (100)	2.27 (100)	0.91 (100)		
D (SSFC (R=2.5))	-33.52	-36.71	-38.62	-39.07	-39.46
D (SSFC (R=3.5))	-32.04	-36.16	-38.40	-38.99	-39.48
D (SSFC (R=4.5))	-31.63	-36.03	-38.34		
D (SSFC)	-31.61	-36.02	-38.34		
MP2					
D	-37.84	-38.90	-39.75	-39.77	-39.85
ΔE^{BSIE}	2.01	0.95	0.10	0.08	0.00
SSFC (R=2.5)	3.15 (61.26)	1.73 (69.53)	0.79 (68.75)	0.39	0.1
SSFC (R=3.5)	4.69 (91.08)	2.33 (93.50)	1.08 (93.50)	0.52	0.1
SSFC (R=4.5)	5.13 (99.54)	2.48 (99.56)	1.15 (99.58)		
SSFC	5.15 (100)	2.49 (100)	1.15 (100)		
D (SSFC (R=2.5))	-34.68	-37.17	-38.95	-39.38	-39.77
D (SSFC (R=3.5))	-33.15	-36.57	-38.67	-39.25	-39.76
D (SSFC (R=4.5))	-32.71	-36.42	-38.60		
D (SSFC)	-32.69	-36.41	-38.59		

Table B.3: SSFC, SSFC(R=4,4.5) BSSE corrections and BSSE corrected and uncorrected stabilization energies for the water cluster series (H₂O)_n at the CCSD/aug-cc-pVXZ (X=D,T and CBS(D,T)) level of theory, all energies in kcal/mol, values in parentheses in %.

n	6	8	10	12
aug-cc-pVDZ				
D	-36.68	-63.36	-81.71	-100.65
SSFC(R=4)	4.84 (92.98)	10.10 (96.58)	12.63 (94.32)	15.41 (93.06)
SSFC(R=4.5)	5.03 (96.68)	10.29 (98.40)	13.02 (97.22)	15.91 (96.12)
SSFC	5.20 (100)	10.45 (100)	13.39 (100)	16.55 (100)
D (SSFC (R=4))	-31.84 (15.20)	-53.07 (19.02)	-69.08 (18.28)	-85.25 (18.07)
D (SSFC (R=4.5))	-31.65 (15.90)	-52.95 (19.43)	-68.69 (18.95)	-84.74 (18.67)
D (SSFC)	-31.48 (16.53)	-51.77 (20.19)	-68.32 (19.60)	-84.10 (19.68)
aug-cc-pVTZ				
D	-38.65	-65.32	-84.08	-103.56
SSFC(R=4)	2.48	5.19	6.55	7.98
SSFC(R=4.5)	2.57	5.28	6.72	8.21
D (SSFC (R=4))	-36.17 (6.85)	-60.12 (8.64)	-77.53 (8.44)	-95.58 (8.35)
D (SSFC (R=4.5))	-36.08 (7.12)	-60.04 (8.80)	-77.36 (6.67)	-95.34 (8.59)
CBS(D,T)				
D	-39.39	-66.72	-85.78	-105.66
SSFC(R=4)	1.96	4.27	5.39	6.55
SSFC(R=4.5)	2.03	4.35	5.53	6.74
D (SSFC (R=4))	-37.43 (5.23)	-62.44 (6.84)	-80.39 (6.70)	-99.10 (6.61)
D (SSFC (R=4.5))	-37.36 (5.43)	-62.37 (6.97)	-80.25 (6.88)	-98.91 (6.81)

Table B.4: As Table B.3, but for the water cluster series $(\text{H}_2\text{O})_n$ $n=14,16,18,20$.

n	14	16	18	20
aug-cc-pVDZ				
D	-123.25	-143.13	-172.38	-185.35
SSFC(R=4)	18.89 (91.59)	21.98 (91.39)	32.81 (92.41)	28.13 (89.18)
SSFC(R=4.5)	19.64 (95.23)	22.74 (94.56)	33.99 (95.74)	29.41 (93.24)
SSFC	20.63 (100)	24.05 (100)	35.50 (100)	32.72 (100)
D (SSFC (R=4))	-104.36 (18.10)	-121.15 (18.14)	-139.57 (23.51)	-157.22 (17.89)
D (SSFC (R=4.5))	-103.61 (18.82)	-120.39 (18.89)	-138.39 (24.56)	-155.94 (18.86)
D (SSFC)	-102.62 (20.10)	-119.08 (20.19)	-137.77 (25.77)	-153.81 (20.51)
aug-cc-pVTZ				
D	-126.11	-146.83	-174.61	
SSFC(R=4)	9.74	11.31	16.31	
SSFC(R=4.5)				
D (SSFC (R=4))	-116.36 (8.37)	-135.52 (8.35)	-158.3 (10.30)	
D (SSFC (R=4.5))				
CBS(D,T)				
D	-128.65	-149.91	-178.44	
SSFC(R=4)	8.07	9.32	13.14	
SSFC(R=4.5)				
D (SSFC (R=4))	-120.58 (6.69)	-140.59 (6.63)	-165.3 (7.95)	
D (SSFC (R=4.5))				

Table B.5: SSFC, SSFC(R=4,4.5) BSSE corrections and BSSE corrected and uncorrected stabilization energies for the water cluster series (H₂O)_n at the MP2/aug-cc-pVXZ (X=D,T and CBS(D,T)) level, (all energies in kcal/mol, values in parentheses in %).

n	6	8	10	12
aug-cc-pVDZ				
D	-37.57	-65.40	-84.31	-104.00
SSFC(R=4)	5.03 (93.09)	10.47 (96.59)	13.10 (94.36)	16.00 (93.14)
SSFC(R=4.5)	5.23 (96.73)	10.67 (98.42)	13.50 (97.23)	16.52 (96.16)
SSFC	5.40 (100)	10.84 (100)	13.89 (100)	17.18 (100)
D (SSFC (R=4))	-32.54 (15.45)	-54.73 (19.13)	-71.21 (18.40)	-88.01 (18.18)
D (SSFC (R=4.5))	-32.34 (16.16)	-54.61 (19.54)	-70.81 (19.04)	-87.49 (18.88)
D (SSFC)	-32.17 (16.80)	-53.38 (20.31)	-70.43 (19.72)	-86.83 (19.78)
aug-cc-pVTZ				
D	-38.96	-66.67	-85.82	-105.84
SSFC(R=4)	2.68	5.55	7.01	8.55
SSFC(R=4.5)	2.77	5.64	7.19	8.80
D (SSFC (R=4))	-36.29 (7.38)	-61.12 (9.08)	-78.81 (8.90)	-97.28 (8.79)
D (SSFC (R=4.5))	-36.19 (7.67)	-61.03 (9.25)	-78.63 (9.13)	-97.03 (9.05)
CBS(D,T)				
D	-39.52	-67.78	-87.16	-107.48
SSFC(R=4)	2.16	4.62	5.85	7.12
SSFC(R=4.5)	2.24	4.70	6.00	7.33
D (SSFC (R=4))	-37.36 (5.78)	-63.17 (7.31)	-81.31 (9.17)	-100.37 (7.09)
D (SSFC (R=4.5))	-37.28 (6.00)	-63.09 (7.44)	-81.16 (9.38)	-100.16 (7.30)

Table B.6: As Table B.5, but for the water cluster series $(\text{H}_2\text{O})_n$ $n=14,16,18,20$.

n	14	16	18	20
aug-cc-pVDZ				
D	-127.21	-147.98	-183.29	-191.43
SSFC(R=4)	19.60 (91.63)	22.81 (91.43)	34.18 (92.47)	29.19 (89.22)
SSFC(R=4.5)	20.37 (95.25)	23.59 (94.58)	35.40 (95.76)	30.52 (93.25)
SSFC	21.39 (100)	24.95 (100)	36.96 (100)	32.72 (100)
D (SSFC (R=4))	-107.62 (18.21)	-125.17 (18.22)	-149.12 (22.92)	-162.23 (18.00)
D (SSFC (R=4.5))	-106.84 (18.93)	-124.38 (18.97)	-147.90 (23.93)	-160.91 (18.96)
D (SSFC)	-105.83 (20.21)	-123.03 (20.28)	-147.25 (25.10)	-158.70 (20.62)
aug-cc-pVTZ				
D	-128.89	-150.24	-184.44	
SSFC(R=4)	10.43	12.12	17.56	
SSFC(R=4.5)	10.82	12.52	18.12	
D (SSFC (R=4))	-118.46 (8.81)	-138.12 (8.77)	-166.88 (10.53)	
D (SSFC (R=4.5))	-118.08 (9.16)	-137.72 (9.09)	-166.32 (10.89)	
CBS(D,T)				
D	-130.93	-152.71	-187.82	
SSFC(R=4)	8.75	10.11	14.35	
SSFC(R=4.5)	9.07	10.46	14.80	
D (SSFC (R=4))	-122.18 (7.16)	-142.59 (7.09)	-173.47 (8.27)	
D (SSFC (R=4.5))	-121.86 (7.45)	-142.25 (7.35)	-173.02 (8.56)	

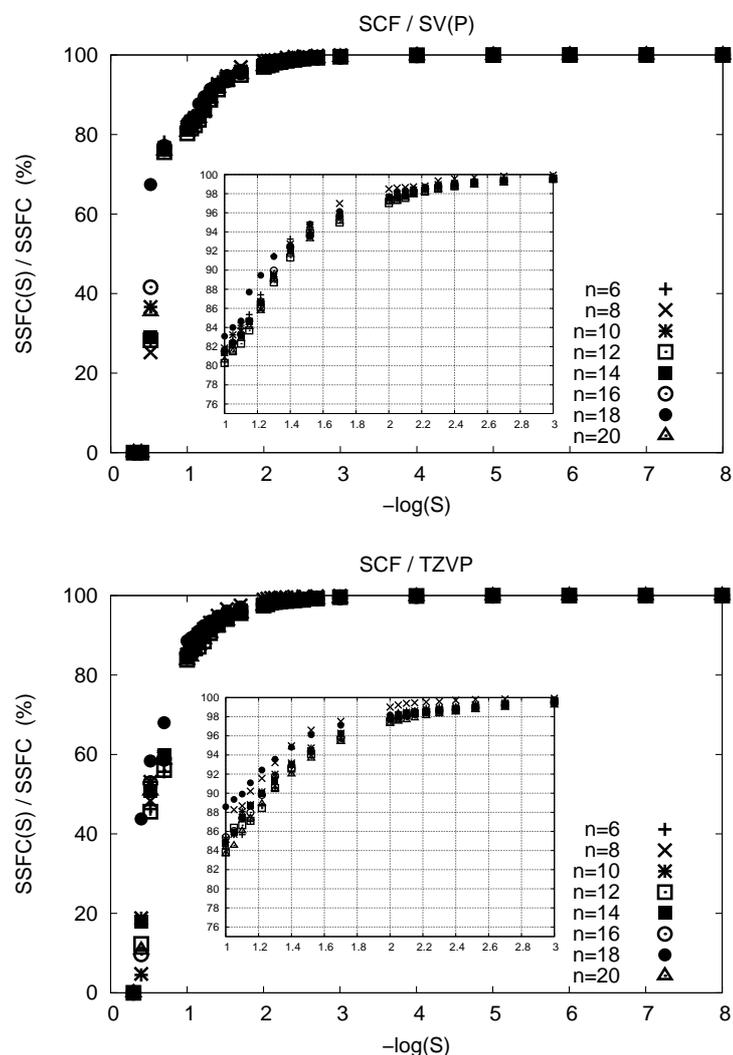


Figure B.3: Percentage coverage of the approximate SSFC(S) scheme with respect to the full SSFC approach as a function of $-\log(S)$ presented for various range of $-\log(S)$ at the SCF level of theory calculated with SV(P) and TZVP basis sets for the $(\text{H}_2\text{O})_n$ water cluster series.

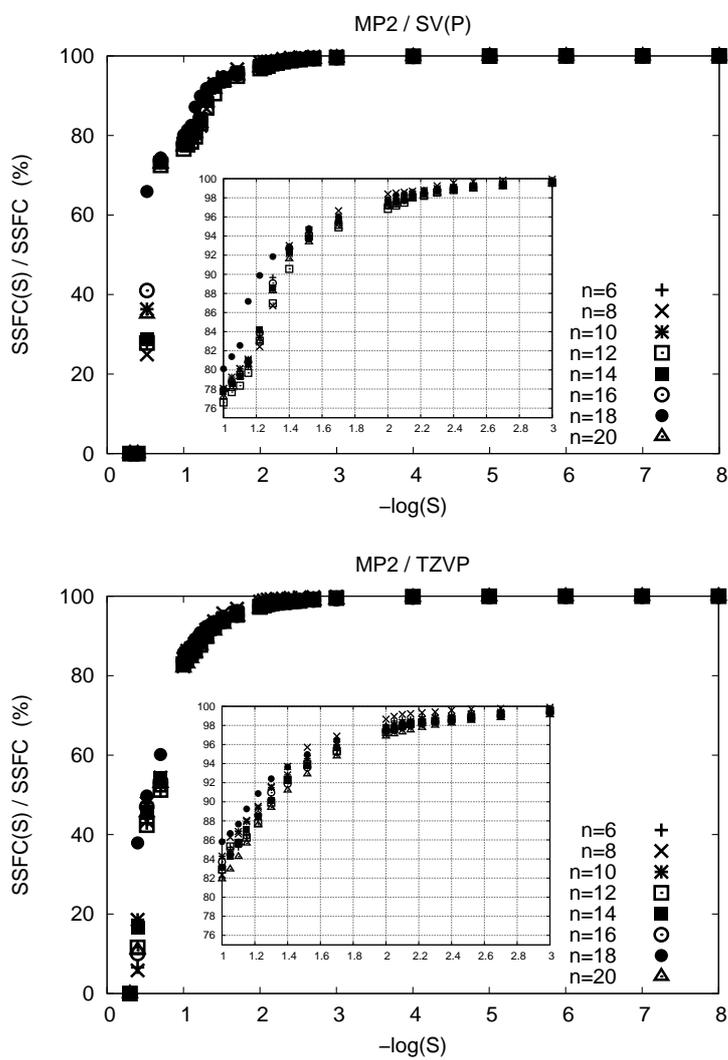


Figure B.4: As Table B.3, but for the MP2 level of theory.

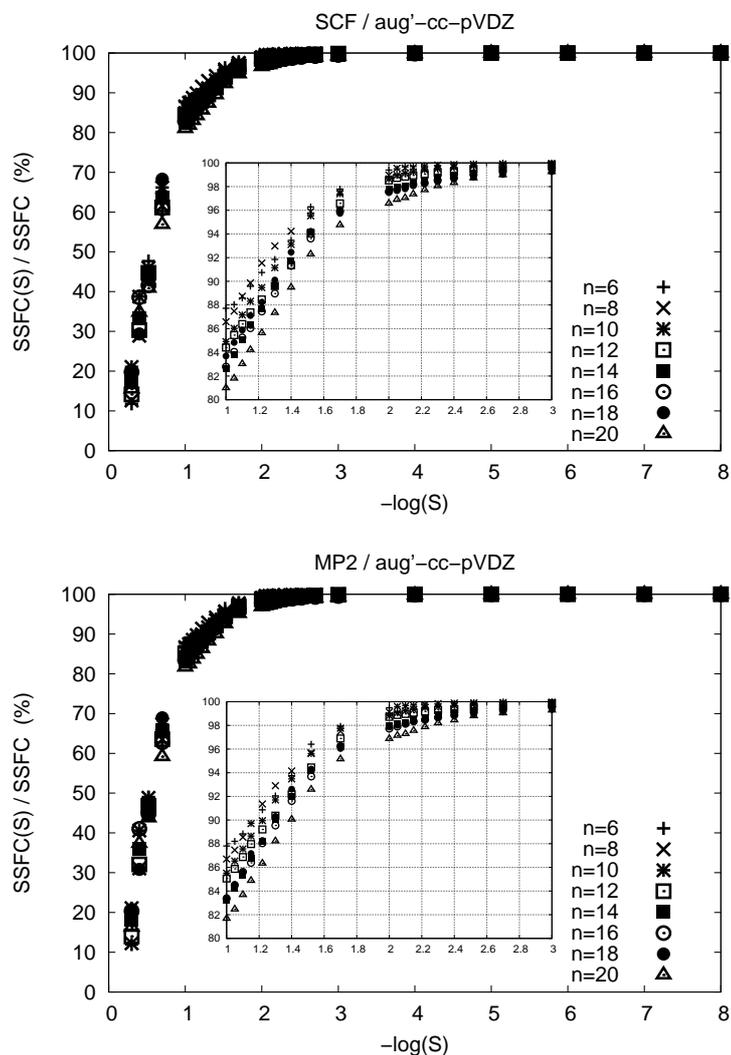


Figure B.5: Percentage coverage of the approximate SSFC(S) scheme with respect to the full SSFC approach as a function of $-\log(S)$ presented for various range of $-\log(S)$ at the MP2 and HF level of theory calculated with aug'-cc-pVDZ basis set for the $(\text{H}_2\text{O})_n$ water cluster series.

Appendix C

Core, core-valence and valence correlation energies

Table C.1: Convergence of the incremental series for core (c), core-valence (cv) and valence (v) CCSD(T) and MP2 correlation energy contribution of cysteine using standard basis sets (6-31G**, cc-pVDZ, cc-pVTZ) and the basis sets of Peterson et al. for the treatment of core correlation effects (cc-pCVDZ). (0 frozen orbitals, 11 correlated core orbitals, 7 core domains, 21 valence orbitals, 5 valence domains, $t_{\text{con}}=3$ bohr, $\text{dsp}=4$, $E_{\text{thres}}=1e-5$ au)

method	order	$E_{\text{corr}}^{\text{c,cv}}(\text{i})$ au	error kcal/mol	$E_{\text{corr}}^{\text{c,cv}}/E_{\text{can.}}$ %	$E_{\text{corr}}^{\text{v}}(\text{i})$ au	error kcal/mol	$E_{\text{corr}}^{\text{v}}/E_{\text{can.}}$ %
6-31G**							
CCSD(T)	2	-0.020023	-0.26	102.13	-1.152773	-2.79	100.39
	3	-0.019607	0.00	100.00	-1.148277	0.03	100.00
canonical		-0.019606			-1.148331		
MP2	2	-0.021226	0.01	99.93	-1.058695	1.24	99.81
	3	-0.021239	0.00	99.99	-1.060675	0.00	100.00
canonical		-0.021241			-1.060670		
cc-pVDZ							
CCSD(T)	2	-0.020425	-0.24	101.90	-1.181039	-3.01	100.41
	3	-0.020044	0.00	100.00	-1.176176	0.04	99.99
canonical		-0.020044			-1.176241		
MP2	2	-0.021662	0.01	99.95	-1.083987	1.30	99.81
	3	-0.021671	0.00	99.99	-1.086062	0.00	100.00
canonical		-0.021673			-1.086058		
cc-pVTZ							
CCSD(T)	2	-0.125832	-0.55	100.70	-1.467314	-3.03	100.33
	3	-0.124978	-0.01	100.01	-1.462440	0.03	100.00
canonical		-0.124961			-1.462486		
MP2	2	-0.123656	0.02	99.98	-1.359700	1.43	99.83
	3	-0.123682	0.00	100.00	-1.361980	0.00	100.00
canonical		-0.123687			-1.361975		
cc-pCVDZ							
CCSD(T)	2	-0.438931	-0.41	100.15	-1.194277	-3.05	100.41
	3	-0.438296	-0.01	100.00	-1.189358	0.04	99.99
canonical		-0.438284			-1.189423		
MP2	2	-0.427408	0.01	100.00	-1.098076	1.30	99.81
	3	-0.427417	0.00	100.00	-1.100150	0.00	100.00
canonical		-0.427421			-1.100145		

Bibliography

- [1] I. N. Levine, *Quantum Chemistry*, Prentice-Hall (2000).
- [2] F. Jensen, *Introduction to Computational Chemistry*, Wiley, Chichester (1999).
- [3] T. Helgaker, P. Jørgensen, J. Olsen, *Molecular Electronic-Struktur Theory*, John Wiley and Sons, LTD, West Sussex PO19 8SQ, England (2000).
- [4] B. O. Roos, P.-O. Widmark, *European Summerschool in Quantum Chemistry 2003 Book I*, University of Lund, Department of Theoretical Chemistry, S-221 00 Lund Sweden (2003).
- [5] A. Szabo, N. S. Ostlund, *Modern Quantum Chemistry*, Dover Publications, New York (1996).
- [6] P. R. von Rague Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. S. III, R. R. Schreiner (Eds.), *Encyclopaedia of Computational Chemistry*, Vol. 3, Wiley, Weinheim (1998).
- [7] C. J. Cramer, *Essentials of Computational Chemistry*, Wiley, New York (2002).
- [8] P. Pulay, S. Saebø, *Theor. Chem. Acc.* **69**, 357 (1986).
- [9] S. Saebø, P. Pulay, *J. Chem. Phys.* **86** (2), 914 (1987).
- [10] S. Saebø, P. Pulay, *Annu. Rev. Phys. Chem.* **44**, 213 (1993).
- [11] M. S. Gordon, J. M. Mullin, S. R. Pruitt, L. B. Roskop, L. V. Slipchenko, J. A. Boatz, *J. Phys. Chem. B* **113**, 9646 (2009).
- [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] J. E. Subotnik, M. Head-Gordon, *J. Chem. Phys.* **123**, 064108 (2005).

- [15] C. Pisani, L. Maschio, S. Casassa, M. Halo, M. Schütz, D. Usvyat, *J. Comput. Chem.* **29**, 2113 (2008).
- [16] C. Hampel, H.-J. Werner, *J. Chem. Phys.* **104**, 6286 (1996).
- [17] M. Schütz, H.-J. Werner, *J. Chem. Phys.* **114**, 661 (2001).
- [18] T. B. Adler, H.-J. Werner, *J. Chem. Phys.* **130**, 241101 (2009).
- [19] N. Flocke, R. J. Bartlett, *J. Chem. Phys.* **121**, 10935 (2004).
- [20] A. Auer, M. Nooijen, *J. Chem. Phys.* **125**, 024104 (2006).
- [21] V. Weijjo, P. Manninen, P. Jørgensen, O. Christiansen, J. Olsen, *J. Chem. Phys.* **127**, 074106 (2007).
- [22] K. Rosciszewski, B. Paulus, P. Fulde, H. Stoll, *Phys. Rev. B* **60**, 7905 (1999).
- [23] B. Doser, D. S. Lambrecht, C. Ochsenfeld, *Phys. Chem. Chem. Phys.* **10**, 3335 (2008).
- [24] E. E. Dahlke, H. R. Leverentz, D. G. Truhlar, *J. Chem. Theory Comput.* **4**, 33 (2008).
- [25] W. Li, P. Piecuch, J. R. Gour, S. Li, *J. Chem. Phys.* **131**, 114109 (2009).
- [26] V. Deev, M. A. Collins, *J. Chem. Phys.* **122**, 154102 (2005).
- [27] S. Li, W. Li, T. Fang, *J. Am. Chem. Soc.* **127**, 7215 (2004).
- [28] D. G. Fedorov, K. Kitaura, *J. Chem. Phys.* **121**, 2483 (2004).
- [29] H. M. Netzloff, M. A. Collins, *J. Chem. Phys.* **127**, 134113 (2007).
- [30] I. Roeggen, *J. Chem. Phys.* **79**, 5520 (1983).
- [31] M. Kobayashi, H. Nakai, *J. Chem. Phys.* **129**, 044103 (2008).
- [32] M. Kobayashi, H. Nakai, *J. Chem. Phys.* **131**, 114108 (2009).
- [33] I. Roeggen, *J. Chem. Phys.* **124**, 184502 (2006).
- [34] D. W. Zhang, J. Z. H. Zhang, *J. Chem. Phys.* **119**, 3599 (2003).
- [35] W. Yang, T.-S. Lee, *J. Chem. Phys.* **103**, 5674 (1995).

- [36] A. van der Vaart, V. Gogonea, S. L. Dixon, K. M. Merz, *J. Comp. Chem.* **21**, 1494 (2000).
- [37] H. Stoll, B. Paulus, P. Fulde, *Chem. Phys. Lett* **469**, 90 (2009).
- [38] R. A. Mata, H. Stoll, *J. Chem. Phys.* **134**, 034122 (2011).
- [39] H. Stoll, *Phys. Rev. B* **46**, 6700 (1992).
- [40] H. Stoll, *Chem. Phys. Lett.* **191**, 548 (1992).
- [41] H. Stoll, *Phys. Rev. B: Condens. Matter Mater. Phys.* **46**, 6700 (1992).
- [42] R. K. Nesbet, *Adv. Chem. Phys.* **14**, 1 (1969).
- [43] A. Shukla, M. Dolg, P. Fulde, H. Stoll, *Phys. Rev. B* **60**, 5211 (1999).
- [44] A. Abdurahman, A. Shukla, M. Dolg, *J. Chem. Phys.* **112**, 4801 (2000).
- [45] S. Kalvoda, B. Paulus, M. Dolg, H. Stoll, H.-J. Werner, *Phys. Chem. Chem. Phys.* **3**, 514 (2001).
- [46] I. Schmitt, K. Fink, V. Staemmler, *Phys. Chem. Chem. Phys.* **11**, 11196 (2009).
- [47] C. Müller, K. Hermansson, B. Paulus, *Chem. Phys.* **362**, 91 (2009).
- [48] J. Friedrich, M. Hanrath, M. Dolg, *Chem. Phys.* **346**, 266 (2008).
- [49] J. Friedrich, *Implementation of the Incremental Scheme for Highly Efficient Correlation Method*, Universität zu Köln, Doktorarbeit (2007).
- [50] J. Friedrich, M. Hanrath, M. Dolg, *J. Chem. Phys.* **126**, 154110 (2007).
- [51] G. Karypis, V. Kumar, *SIAM J. Sci. Compt.* **20** (1), 359 (1998).
- [52] J. Friedrich, M. Dolg, *J. Chem. Phys.* **129**, 244105 (2008).
- [53] J. Friedrich, *J. Chem. Theory Comput.* **6**, 1834 (2010).
- [54] A. G. Reza, I. Roeggen, *Theor. Chem. Acc.* **97**, 41 (1997).
- [55] J. Friedrich, M. Dolg, *J. Chem. Theory Comput.* **5**, 287 (2009).
- [56] J. Friedrich, K. Walczak, M. Dolg, *Chem. Phys.* **356**, 47 (2009).
- [57] K. Walczak, *Implementierung und Anwendung der Rumpf-Valenz-Korrelation im Rahmen der Inkrementenmethode*, Diplomarbeit, Universität zu Köln (2008).

- [58] H.-J. Werner, P. J. Knowles, R. Lindh, F. R. Manby, M. Schütz, *et al.*, *MOLPRO, a package of ab initio programs designed by H.-J. Werner and P. J. Knowles, Version 2002 and Version 2006*, University of Birmingham, Birmingham, UK (2006).
- [59] J. M. Foster, S. F. Boys, *Rev. Mod. Phys.* **32**, 300 (1960).
- [60] C. Edmiston, K. Ruedenberg, *Rev. Mod. Phys.* **35**, 457 (1963).
- [61] J. Friedrich, M. Hanrath, M. Dolg, *Chem. Phys.* **338**, 33 (2007).
- [62] N. R. Kestner, *J. Chem. Phys.* **48**, 252 (1968).
- [63] H. Jansen, P. Ros, *Chem. Phys. Lett.* **3**, 140 (1969).
- [64] B. Liu, A. D. McLean, *J. Chem. Phys.* **59**, 4557 (1973).
- [65] A. Galano, J. R. Alvarez-Idaboy, *J. Comput. Chem.* **27**, 1204 (2006).
- [66] N. R. Kestner, J. E. Combarize, *Reviews in Computational Chemistry*, Vol. 13, Wiley-VCH, New York (1999).
- [67] T. H. Dunning Jr., *J. Chem. Phys.* **90**, 1007 (1989).
- [68] R. A. Kendall, T. H. Dunning Jr., R. J. Harrison, *J. Chem. Phys.* **96**, 6769 (1992).
- [69] D. E. Woon, T. H. Dunning Jr., *J. Chem. Phys.* **100**, 2975 (1994).
- [70] R. M. Balabin, *J. Chem. Phys.* **129**, 164101 (2008).
- [71] H. Valdes, V. Klusak, M. Pitonak, O. Exner, I. Stary, P. Hobzda, L. Rulisek, *J. Comp. Chem.* **29**, 861 (2007).
- [72] A. Asturiol, M. Duran, P. Salvador, *J. Chem. Theory Comput.* **5**, 2574 (2009).
- [73] R. M. Balabin, *J. Chem. Phys.* **132**, 211103 (2010).
- [74] R. M. Balabin, *J. Chem. Phys.* **132**, 231101 (2010).
- [75] L. F. Holroyd, T. van Mourik, *Chem. Phys. Lett.* **442**, 42 (2007).
- [76] D. Moran, A. C. Simmonett, F. E. Leach, W. D. Allen, P. v. R. Schleyer, H. F. Schaefer, *J. Am. Chem. Soc.* **128**, 9342 (2006).
- [77] A. Asturiol, M. Duran, P. Salvador, *J. Chem. Phys.* **128**, 144108 (2008).
- [78] S. Saebo, W. Tong, P. Pulay, *J. Chem. Phys.* **98**, 2170 (1993).

- [79] M. Schütz, G. Rauhut, H.-J. Werner, *J. Phys. Chem. A* **102**, 5997 (1998).
- [80] I. Mayer, *Int. J. Quantum Chem.* **23**, 341 (1983).
- [81] I. Mayer, P. Valiron, *J. Chem. Phys.* **109**, 3360 (1998).
- [82] A. Hamza, Á. Vibók, G. J. Halász, I. Mayer, *J. Mol. Struct.: THEOCHEM* **501**, 427 (2000).
- [83] T. Nagata, S. Iwata, *J. Chem. Phys.* **120**, 3555 (2004).
- [84] S. Iwata, T. Nagata, *Theor. Chem. Acc.* **117**, 137 (2007).
- [85] R. Z. Khaliullin, M. Head-Gordon, A. T. Bell, *J. Chem. Phys.* **124**, 204105 (2006).
- [86] E. Gianinetti, M. Raimondi, E. Tornaghi, *Int. J. Quant. Chem.* **60**, 157 (1996).
- [87] T. Nagata, O. Takahashi, K. Saito, S. Iwata, *J. Chem. Phys.* **115**, 3553 (2001).
- [88] P. Salvador, I. Mayer, *J. Chem. Phys.* **120**, 5882 (2004).
- [89] B. Jeziorski, R. Moszynski, K. Szalewicz, *Chem. Rev.* **94**, 1887 (1994).
- [90] V. F. Lotrich, K. Szalewicz, *J. Chem. Phys.* **106**, 9688 (1997).
- [91] R. Moszynski, P. E. S. Wormer, B. Jeziorski, A. van der Avoird, *J. Chem. Phys.* **103**, 8058 (1995).
- [92] R. Moszynski, P. E. S. Wormer, B. Jeziorski, A. van der Avoird, *J. Chem. Phys.* **107**, 672 (1997).
- [93] S. F. Boys, F. Bernardi, *Mol. Phys.* **19**, 553 (1970).
- [94] T. Reuters (Ed.), *Web of Science*, <http://apps.webofknowledge.com>, as reported in the database on 26 January 2012.
- [95] Y. Wang, B. Paulus, *Chem. Phys. Lett.* **441**, 187 (2007).
- [96] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen, A. K. Wilson, *Chem. Phys. Lett.* **286**, 243 (1998).
- [97] F. Weigend, M. Häser, *Theo. Chem. Acc.* **97**, 331 (1997).
- [98] F. Weigend, M. Häser, H. Patzelt, R. Ahlrichs, *Chem. Phys. Lett.* **294**, 143 (1998).

- [99] 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 5.10*, Institut für Physikal. Chemie, Universität Karlsruhe, Kaiserstr. 12, D-76128 Karlsruhe.
- [100] C. Hampel, K. A. Peterson, H.-J. Werner, *Chem. Phys. Lett.* **190**, 1 (1992).
- [101] M. J. O. Deegan, P. J. Knowles, *Chem. Phys. Lett.* **227**, 321 (1994).
- [102] H.-J. Werner, P. J. Knowles, R. Lindh, F. R. Manby, M. Schütz, *et al.*, *MOLPRO, a package of ab initio programs designed by H.-J. Werner and P. J. Knowles version 2006*, University of Birmingham (2006).
- [103] D. Woon, T. H. Dunning Jr., *J. Chem. Phys.* **98**, 1358 (1993).
- [104] J. Friedrich, M. Hanrath, M. Dolg, *J. Chem. Phys.* **126**, 154110 (2007).
- [105] R. A. Aziz, M. J. Slaman, *J. Chem. Phys.* **94**, 8047 (1991).
- [106] J. P. Daudey, P. Claverie, J. P. Malieu, *Int. J. Quantum Chem.* **8**, 1 (1974).
- [107] C. S. Nash, *J. Chem. Theory Comput.* **1**, 261 (2005).
- [108] F. B. van Duijneveldt, J. G. C. M. van Duijneveldt-van de Rijdt, J. H. van Lenthe, *Chem. Rev.* **94**, 1873 (1994).
- [109] J. R. Collins, G. A. Gallup, *Chem. Phys. Lett.* **123**, 56 (1986).
- [110] J. R. Collins, G. A. Gallup, *Chem. Phys. Lett.* **129**, 329 (1986).
- [111] M. Gutowski, F. B. van Duijneveldt, G. Chalasinski, L. Piela, *Chem. Phys. Lett.* **129**, 325 (1986).
- [112] B. Liu, A. D. McLean, *J. Chem. Phys.* **91**, 2348 (1989).
- [113] M. Gutowski, J. H. van Lenthe, J. Verbeek, F. B. van Duijneveldt, G. Chalasinski, *Chem. Phys. Lett.* **124**, 370 (1986).
- [114] H. Stoll, B. Paulus, P. Fulde, *J. Chem. Phys.* **123**, 144108 (2005).
- [115] J. Friedrich, M. Hanrath, M. Dolg, *J. Phys. Chem. A* **111**, 9830 (2007).
- [116] S. S. Xantheas, *J. Chem. Phys.* **104**, 8821 (1996).

- [117] B. Kvamme, C. F. Wander, A. E. Clark, *Int. J. Quantum Chem.* **109**, 2474 (2009).
- [118] B. H. Wells, S. Wilson, *Chem. Phys. Lett.* **101**, 429 (1983).
- [119] P. Valiron, I. Mayer, *Chem. Phys. Lett.* **275**, 46 (1997).
- [120] J. C. White, E. R. Davidson, *J. Chem. Phys.* **93**, 8029 (1990).
- [121] P. Salvador, M. M. Szczesniak, *J. Chem. Phys.* **118**, 537 (2002).
- [122] S. Kalvoda, M. Dolg, H.-J. Flad, P. Fulde, H. Stoll, *Phys. Rev. B* **57** (1998).
- [123] P. N. Day, R. Pachter, M. S. Gordon, G. N. Merrill, *J. Chem. Phys.* **112**, 2063 (2000).
- [124] M. M. Pires, V. F. DeTuri, *J. Chem. Theory Comput.* **3**, 1073 (2007).
- [125] *TURBOMOLE V6.3 2011, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from <http://www.turbomole.com>.*
- [126] W. J. Hehre, R. Ditchfield, J. A. Pople, *J. Chem. Phys.* **56**, 2257 (1972).
- [127] W. J. Hehre, R. F. Stewart, J. A. Pople, *J. Chem. Phys.* **51**, 2657 (1969).
- [128] D. M. Bates, G. S. Tschumper, *J. Phys. Chem. A* **113**, 3555 (2009).
- [129] A. D. Boese, G. Jansen, M. Torheyden, S. Hoefener, W. Klopper, *Phys. Chem. Chem. Phys.* **13**, 1230 (2011).
- [130] P. C. Hariharan, J. A. Pople, *Theor. Chim. Acta* **28**, 213 (1973).
- [131] A. F. Holleman, E. Wiberg, N. Wiberg, *Lehrbuch der anorganischen Chemie*, 101. Ed., Walter de Gruyter (1995).
- [132] M. Atoji, W. N. Lipscomb, *Acta Crystallogr.* **7**, 173 (1954).
- [133] A. Schäfer, H. Horn, R. Ahlrichs, *J. Chem. Phys.* **97**, 2571 (1992).
- [134] A. Schäfer, C. Huber, R. Ahlrichs, *J. Chem. Phys.* **100**, 5829 (1994).
- [135] P. R. von Rogue Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. S. III, R. R. Schreiner (Eds.), *Encyclopaedia of Computational Chemistry*, Vol. 1, Wiley, Weinheim (1998).

- [136] C. Estarellas, X. Lucas, A. Frontera, D. Quinonero, P. M. Deya, *Chem. Phys. Lett.* **489**, 254 (2010).
- [137] D. Tzeli, A. A. Tsekouras, *Chem. Phys. Lett.* **496**, 42 (2010).
- [138] T. H. Dunning Jr., *J. Chem. Phys.* **90**, 1007 (1989).
- [139] D. E. Woon, T. H. Dunning Jr., *J. Chem. Phys.* **103** (11), 4572 (1995).
- [140] K. A. Peterson, T. H. Dunning Jr., *J. Chem. Phys.* **117** (23), 10548 (2002).
- [141] K. A. Peterson, C. Puzzarini, *Theor. Chem. Acc.* **114**, 283 (2005).
- [142] S. C. Chmely, C. N. Carlson, T. P. Hanusa, A. L. Rheingold, *J. Am. Chem. Soc.* **131**, 6344 (2009).
- [143] D. Feller, *J. Comp. Chem.* **17**, 1571 (1996).
- [144] K. Schuchardt, B. Didier, T. Elsethagen, L. Sun, V. Gurumoorthi, J. Chase, J. Li, T. Windus, *J. Chem. Inf. Model.* **47**, 1045 (2007).
- [145] A. D. Becke, *Phys. Rev. A* **38** (6), 3098 (1988).
- [146] J. P. Perdew, *Phys. Rev. B* **33** (12), 8822 (1986).
- [147] 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 5.6*, Institut für Physikal. Chemie, Universität Karlsruhe, Kaiserstr. 12, D-76128 Karlsruhe (2002).
- [148] D. Figgen, G. Rauhut, M. Dolg, H. Stoll, *Chem. Phys.* **311**, 227 (2005).

Danksagung

Meinen ganz herzlichen Dank richte ich an meinen Doktorvater Herrn Prof. Dr. Michael Dolg, der mich fortwährend in der Studien-, Diplom- und Promotionszeit unterstützt und gefördert hat. Ich bedanke mich für die hervorragende Betreuung dieser Arbeit, für die interessante Aufgabenstellung und die vielen Ideen und Anregungen, die zu dieser Arbeit geführt haben. Ich bedanke mich auch sehr für die Möglichkeit der Teilnahme an zahlreichen Konferenzen sowie für die Förderung des Besuches einer Sommerschule.

Für die Zusammenarbeit und Unterstützung bei der Umsetzung der in dieser Arbeit neu vorgestellten Methoden bedanke ich mich sehr beim Herrn Jun.-Prof. Dr. Joachim Friedrich.

Herrn Priv.-Doz. Dr. Michael Hanrath danke ich für die freundlich Übernahme des Zweitgutachtens dieser Arbeit.

Für die administrativen Arbeiten an den PC-Clustern des Institutes bedanken ich mich bei Birgitt Börsch-Pulm, PD Dr. M. Hanrath und Joseph Held.

Für die Vorlesungen, Seminare und Übungen, die ich während meiner Ausbildung besucht habe, bedanke ich mich bei Prof. Dr. M. Dolg, PD Dr. M. Hanrath, JP Dr. J. Friedrich, Dr. X. Cao-Dolg, Dr. M. Burkatzki und Dr. A. Engels-Putzka.

Für das sehr angenehme Arbeitsklima, die zahlreichen Ratschläge sowie die vielen anregenden Diskussionen danke ich allen Mitarbeitern des Institutes der Theoretischen Chemie in Köln: Martin Böhler, Birgitt Börsch-Pulm, Dr. Mark Burkatzki, Dr. Xiaoyan Cao-Dolg, Jan Ciupka, Dr. Anna Engels-Putzka, Tim Hangele, Dr. M. Hanrath, Norah Heinz, Joseph Held, Michael Hülsen, Daniel Pape, Dr. Anja Pfletschinger, Dr. Jonas Wiebke, Dr. Anna Weigand und Daniel Weißmann. Ich bedanke mich bei Joseph Held, der mir bei der Beantragung und Einrichtung des SuGi-Accounts am Rechenzentrum in Köln hilfsbereit zu Seite stand.

Ein sehr herzlicher Dank gilt schließlich meinem Vater, meiner Schwester und meiner Freundin Friederike Rathmann, die alle auf ihre Weise zum Gelingen dieser Arbeit beigetragen haben.

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. M. Dolg betreut worden.

Teilpublikationen:

- J. Friedrich, K. Walczak, M. Dolg, F. Piestert, T. Lauterbach, D. Worgull, A. Gansauer, J. Am. Chem. Soc. **130**, 5 (2008).
- J. Friedrich, K. Walczak, M. Dolg, Chem. Phys. **356**, 47 (2009).
- K. Walczak, J. Friedrich, M. Dolg, Chem. Phys. **365**, 38 (2009).
- K. Walczak, J. Friedrich, M. Dolg, Chem. Phys. **376**, 36 (2010).
- K. Walczak, J. Friedrich, M. Dolg, J. Chem. Phys. **135**, 134118 (2011).

Ich versichere, dass ich alle Angaben wahrheitsgemäß nach bestem Wissen und Gewissen gemacht habe und ich verpflichte mich, jedmögliche, die den obigen Angaben betreffende Veränderung, dem Dekanat unverzüglich mitzuteilen.

Datum

Unterschrift

