Study of Molecular Clusters with Explicitly Correlated Wave Function Methods

Von der Fakultät Chemie der Universität Stuttgart
zur Erlangung der Würde eines Doktors der
Naturwissenschaften (Dr. rer. nat.) genehmigte Abhandlung

Vorgelegt von
Oliver Marchetti
aus Esslingen

Hauptberichter: Prof. Dr. H.-J. Werner, Universität Stuttgart
Mitberichter: Prof. Dr. G. Rauhut, Universität Stuttgart
Tag der mündlichen Prüfung: 17.07.2013

Institut für Theoretische Chemie
der Universität Stuttgart

2013

Acknowledgments

Thanks to ...

... my mother for personal support, food supply and precise professional haircuts.

... my brother Benjamin for lashing me on.

... Prof. Werner for taking me as his protégé.

Contents

General definitions

ab initio method[4]	first principles quantum mechanics
dimer	super-system from two identical monomers
oligomer	super-system of $\approx$ 10 to 30 identical (or similar) monomers
polymer	super-system of $\geq$ 31 identical (or similar) monomers
cluster	super-system of any (identical or different) N monomers with N $\geq$ 2

General abbreviations

cusp	singular point of a curve
PES	Potential Energy Surface
MAD	Mean Absolute Deviation
MAX	Maximum error
RMS	Root Mean Square (error)

Chemical

PhH	Benzene
CpH	1,3-Cyclopentadien
$\parallel$	parallel orientation
$\perp$	perpendicular orientation

Symbols

ϵ_0	dielectric constant
μ^N	dipole moment of system N
Θ^N	quadrupole-moment of system N
$\bar{\alpha}^N$	orientation-averaged polarizability of system N
U^N	ionization potentials of system N

Basis sets

VnZ	cc-pVnZ (correlation consistent-polarized Valence n-Tuple Zeta)
AVnZ	aug-cc-pVnZ (Augmented cc-pVnZ)
AVnZ'	C, N, O = aug-cc-pVnZ, H = cc-pVnZ
CBS[nm]	Complete Basis Set extrapolation, from VnZ and VmZ
BSSE	Basis Set Superposition Error
CP	Counterpoise Correction
AO	Atomic Orbital (χ_α)
GTO	Gaussian Type Orbital
MO	Molecular Orbital (ϕ_i, ϕ_i^{can})
LMO	Localized Molecular Orbital (ϕ_i^{loc})
PAO	Projected Atomic Orbital ($\phi_{a'}$)

Methods

LCAO	Linear Combination of Atomic Orbitals
HF	Hartree-Fock
SCF	Self Consistent Field
MP2	Møller Plesset second-order perturbation theory
LMP2	Local Møller Plesset second-order perturbation theory
CC	Coupled Cluster
FCI	Full Configuration Interaction
R12	linear correlation factor
F12	Slater type function correlation factor
RI	Resolution of the Identity
DF	Density Fitting
CABS	Complementary Auxiliary Basis Set

MP2-F12

3C	Ansatz **3**, approximation C
(D)	Diagonal-Ansatz
fix	Fixed Amplitudes Ansatz

1 Introduction

In the classical ideal gas there are no interparticular forces, it is described by:

$$PV = nRT \tag{1}$$

with P the pressure, V the volume, n the amount of substance of the gas (in moles), R the gas constant ($8.314\ J \cdot K^{-1}mol^{-1}$) and T the absolute temperature.
The Dutch scientist Johannes Diderik van der Waals derived a correction leading to the van der Waals equation:

$$\left(P + \frac{n^2 a}{V^2}\right)(V - nb) = nRT \tag{2}$$

where $a = N_A^2 a'$ is a measure of interparticular forces and $b = N_A b'$ is the volume excluded by a mole of particles. The interparticular forces in the van der Waals equation can not be explained by covalent bonds (characterized by the sharing of electron pairs) of the particles (atoms and molecules) in the gas. Those forces can only be explained via weak interactions and can be attractive or repulsive. The weak interactions between atoms and molecules have later been named after Johannes Diderik van der Waals[a]. Weak interactions are not restricted to intermolecular interactions but can also occur intramolecular. In both cases the weak interactions can be responsible for the shape of molecules, supersystems and clusters. A famous example for this is the protein folding in RiboNucleic Acid (RNA) and DeoxyriboNucleic Acid (DNA). The effect of medical compounds often is founded on weak interactions with *something* in the human body. Unfortunately, the calculation of weak interactions is a challenging task. Very accurate *ab initio* quantum mechanical methods and a good description of the molecular orbitals are needed. Both requirements lead to high computational demands (calculation time and storage usage for main memory and hard disc). In *ab initio* quantum chemistry one normally only calculates the (total) energy E of a system. This is done with an accuracy of approximately 99% (when compared to the exact result[b]) if the Hartree-Fock method (HF) is used. For many cases this accuracy is not sufficient to achieve the *chemical accuracy* of 1 kcal/mol. In contrast to thermochemistry where (reaction) enthalpies (ΔH_r^Θ) are typically calculated from standard enthalpies of formation (ΔH_f^Θ) (which are of the same order of magnitude as the reaction enthalpies and a certain accuracy of the formation enthalpies leads to a

[a] as van der Waals force
[b] this is taken as the Full Configuration Interaction energy

result of similar quality of the relative energies) in *ab initio* quantum chemistry calculations, the calculated total energies are larger by orders of magnitude than the energy of the desired property. For weak interactions the ratio of calculated values and the desired property is even larger and dynamic electron correlation, which is not covered at all by HF, plays a larger role or is sometimes responsible for the complete interaction energy. In particular the dispersion energy is a pure electron correlation effect. Methods that give a correction to HF to include dynamic electron correlation in the calculated energy need larger basis sets (see **Section 3.3**) for the description of the molecular orbitals than HF itself.

In the example of the interaction energy of the benzene dimer, the total electronic energy of the dimer is approximately 300000 kcal/mol, whereas the interaction energy is approximately 3 kcal/mol. This ratio of 10^5 illustrates why an accuracy of more than 99% and hence methods that cover dynamic electron correlation are necessary. The Coupled Cluster method with single and double excitations and triple excitations at perturbation theory level (CCSD(T)) (see **Section 3.10**) is considered to be highly accurate for most weak interactions.

The convergence of the total energy with the basis set size can be greatly improved by the introduction of explicit correlation. Unfortunately CCSD(T) is rather expensive and systems with more than 30 atoms are very hard to calculate. Many systems of interest with weak interactions are larger, and therefore CCSD(T) is not feasible for their calculation. The MP2 method is much cheaper[c] than the CCSD(T) method, but unfortunately not as accurate. Earlier attempts to improve MP2 like SCS-MP2[5] (see **Section 3.9.2**) have been successful to some extent but were never totally satisfying. Other attempts such as MP2.5[6] (see **Section 3.9.4**) are more expensive and therefore also restricted to smaller systems.

The goal of this work was to test the effects of explicit correlation in the CCSD(T) method and to develop a new method for the calculation of weak interactions that is able to improve on MP2 and SCS-MP2 towards CCSD(T). A new method - Dispersion-Weighted-MP2[d] (DW-MP2) has been developed in this work. It combines MP2 and SCS-MP2 and will be shown to achieve the accuracy needed for *chemical accuracy*.

[c] less demanding in calculation time, memory and hard disk usage

[d] DW-MP2 has only been used in combination with explicit correlation (DW-MP2-F12) in order to avoid basis set errors.

2 Fundamentals

2.1 Atomic units

Atomic units (au) are often used in quantum chemistry. They are defined by setting

$$e = m_e = \hbar = \frac{1}{4\pi\varepsilon_0} = 1. \tag{3}$$

Atomic units are used in this work throughout, unless otherwise noted. The atomic unit of energy and distance are one Hartree $E_h = 627.509469$ kcal/mol and one Bohr radius $a_B = 0.5217721$ Å, but for convenience interaction energies will usually be given in kcal/mol and distances in Angstrom [Å].

2.2 Weak interactions

Weak interactions in a chemical sense describe the forces between (two) molecules (or atoms) which are not linked by a covalent bond or mechanically-interlocked[a]. A covalent bond is characterized by the sharing of electron pairs between atoms of the molecules involved. Forces between molecules (or atoms) are summarized under the expression van-der-Waals interactions and include a subset of multipole $\cdots$ multipole interactions. The van-der-Waals forces are all those that cause the deviation from ideal gas law behavior and include:

- hydrogen-bonds
- (static) multipole $\cdots$ multipole interactions
- (static) multipole $\cdots$ induced multipole interactions [b]
- interactions of instantaneous multipoles (dispersion energy[7])

Ionic interactions are part of the multipole $\cdots$ multipole interactions but not of the van-der-Waals type. They are not considered as weak interactions and have not been studied in this work. Dipole $\cdots$ dipole interactions are a subset of multipole $\cdots$ multipole interactions and also cover hydrogen bonds. Sometimes hydrogen bonds however are considered separately from the multipole $\cdots$ multipole set of interactions since they show some covalent character.

[a] Catenane or Rotaxane are mechanically-interlocked systems, for example
[b] Debye forces

The *IUPAC. Compendium of Chemical Terminology, 2nd ed*[8] gives following definition: van der Waals forces:
"The attractive or repulsive forces between molecular entities (or between groups within the same molecular entity) other than those due to bond formation or to the electrostatic interaction of ions or of ionic groups with one another or with neutral molecules. The term includes: dipole $\cdots$ dipole, dipole $\cdots$ induced dipole and London (instantaneous induced dipole $\cdots$ induced dipole) forces. The term is sometimes used loosely for the totality of nonspecific attractive or repulsive intermolecular forces."

2.3 Strength of weak interactions

The different types of weak interactions can be sorted by their strength. There is of course an overlap between the strength of the weak interactions according to their type, but in general one can say that hydrogen-bonds (due to their covalent character) are the strongest. The multipole $\cdots$ multipole interactions, multipole $\cdots$ induced multipole interactions and interactions of instantaneous multipoles are weaker in that order. It is not possible to identify the character of the interaction just by looking at the interaction energy. For systems large enough it is possible to find interactions of instantaneous multipoles that are stronger than hydrogen-bonds. Interactions of instantaneous multipoles are also present in macroscopic systems. In those macroscopic systems they can become big as seen in graphene mono layers[9] and they are strong enough to enable a gecko to climb a wall[10] in order to escape the living part of Schrödinger's cat.[11]

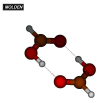
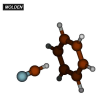
$\frac{1}{R^n}$	ΔE	example cluster	dominating interaction
← distance dependence	interaction strength →	formic acid dimer 	hydrogen-bond
		uracil dimer stack 	multipole ··· multipole
		benzene HCN 	multipole ··· induced multipole
		benzene dimer 	interactions of instantaneous multipoles

Table I: Cluster interactions

2.4 Potential energy of weak interactions

The different van-der-Waals interactions have different potentials. The most important contributions are summarized in this **Section (2.4)**. For a more complete and detailed description please refer to the books *The Theory of Intermolecular Forces*[12] and *Intermolecular Forces.*[13]

2.4.1 Dipole ··· dipole interactions

The PES of a dipole ··· dipole interaction between two interacting systems X and Y is described by

$$U_{dip}^{dip} = P_{Orient} \frac{\mu^X \mu^Y}{R^3} \tag{4}$$

Here R is distance between the centers of mass of the systems, and the prefactor P_{Orient} gives the sign and strength of the interaction according to the orientation.
μ^X and μ^Y are the magnitudes of the molecular dipole moments of X and Y, respectively.

System	Relative orientation				
X	$\uparrow$	$\downarrow$	$\rightarrow$	$\leftarrow$	$\nearrow_{55^\circ}$
Y	$\uparrow$	$\uparrow$	$\rightarrow$	$\rightarrow$	$\nearrow_{55^\circ}$
P_{orient}	-2	$+2$	$+1$	-1	± 0

Table II: Dipole $\cdots$ dipole interaction orientation prefactors for planar structures[12]

2.4.2 Dipole $\cdots$ quadrupole interactions

The potential of dipole $\cdots$ quadrupole interactions:

$$U_{quad}^{dip} = P_{orient}\frac{\mu^X \Theta^Y}{R^4}, \tag{5}$$

where the magnitude of the molecular dipole-moment is μ^X and Θ^Y is the magnitude of the molecular quadrupole-moment.

System	Relative orientation		
X	$\uparrow$	$\downarrow$	$\downarrow$
Y	$\updownarrow$	$\updownarrow$	$\leftrightarrow$
P_{orient}	-3	$+3$	$-1\frac{1}{2}$

Table III: Dipole $\cdots$ quadrupole interaction orientation prefactors for planar structures[12]

2.4.3 Quadrupole $\cdots$ quadrupole interactions

For quadrupole $\cdots$ quadrupole interactions it is described by:

$$U_{quad}^{quad} = P_{Orient}\frac{\Theta^X \Theta^Y}{R^5} \tag{6}$$

System	Relative orientation				
X	$\updownarrow$	$\leftrightarrow$	$\leftrightarrow$	$\bullet$	$\swarrow\!\!\nearrow$
Y	$\updownarrow$	$\updownarrow$	$\leftrightarrow$	$\leftrightarrow$	$\swarrow\!\!\nearrow$
P_{orient}	$+6$	-3	$+2\frac{1}{4}$	$+\frac{3}{4}$	$-2\frac{7}{16}$

Table IV: Quadrupole $\cdots$ quadrupole interaction orientation prefactors[12]

where Θ^X and Θ^Y are magnitudes of the molecular quadrupole-moments of the respective systems.

2.4.4 Dipole $\cdots$ induced-dipole interactions

Regarding the PES of dipole $\cdots$ induced-dipole interactions, they are given by:

$$U_{idip}^{dip} = P_{orient} \frac{\mu^X \mu_{ind}^Y}{R^6} \tag{7}$$

2.4.5 Interactions of instantaneous dipoles

An approximate[c] potential for interactions of instantaneous dipoles is given by the London formula:

$$U_{disp} \approx -\frac{3U^X U^Y}{2(U^X + U^Y)} \frac{\bar{\alpha}^X \bar{\alpha}^Y}{R^6} \tag{8}$$

where $\bar{\alpha}^X$ $\bar{\alpha}^Y$ are the orientation-averaged dipole polarizabilities of the systems. The ionization potentials are U^X and U^Y. This formulation does not take into account the interactions of instantaneous multipoles of higher order than dipoles.

[c] averaged over relative orientations of the molecules of the R^-6 term

3 *ab initio* Theory

3.1 Schrödinger equation

In this work only time-independent *ab initio* theory has been applied to investigate the problems, therefore it is only necessary to consider the time-independent (stationary) form of the electronic Schrödinger equation.[14] The usage of the electronic Schrödinger equation[14] implies that the movement of the electrons and therefore the electronic wave-function $\Psi_e \equiv \Psi(\mathbf{x})$ can be considered independent of the nuclear motion and therefore the nuclear wave function Ψ_N. That is justified by the fact that the ratio between nuclear and electronic masses is high and so the movement of the electrons can be considered separate of those of the nuclei. This separation is called Born-Oppenheimer approximation. The Born-Oppenheimer approximation allows the wavefunction of a molecule to be broken into its electronic and nuclear components.

In the *ab initio* quantum mechanics formulation, a system is represented completely by its electronic wave-function $\Psi(\mathbf{x})$ where $\mathbf{x}$ is a vector which holds all spatial and spin coordinates of the electrons and $\hat{H}$ is the electronic Hamiltonian, the operator that represents the total energy E of a system for fixed nuclei. The operator $\hat{H}$ can be written as a sum of the operator for potential energy $\hat{V}$ and the operator for kinetic energy of the electrons $\hat{T}_e$ (in atomic units):

$$\hat{H} = \underbrace{-\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2}_{=\hat{T}_e} \underbrace{-\sum_{m=1}^{M}\sum_{i=1}^{N}\frac{Z_m}{r_{im}} + \sum_{i<j}\frac{1}{r_{ij}} + \sum_{m<n}\frac{Z_m Z_n}{r_{mn}}}_{=\hat{V}} \tag{9}$$

With the potential energy operator $\hat{V}$:

$$\hat{V} = \underbrace{-\sum_{m=1}^{M}\sum_{i=1}^{N}\frac{Z_m}{r_{im}}}_{(electron\ nucleus\ potential)} \underbrace{+\sum_{i<j}\frac{1}{r_{ij}}}_{(electron\ electron\ potential)} \underbrace{+\sum_{m<n}\frac{Z_m Z_n}{r_{mn}}}_{(nucleus\ nucleus\ potential)} \tag{10}$$

The Hamiltonian $\hat{H}$ is thus known exactly within the scope of non-relativistic quantum mechanics. The only task left is to solve the electronic Schrödinger equation[14] to get the exact total energy E and the corresponding wave function Ψ. For systems with more that one electron an exact solution is not possible so that one has to use approximations which introduce an error in the total energy E.

3.2 Hartree-Fock

The foundation for most of the *ab initio* calculations is the Hartree-Fock theory (HF). HF uses a single Slater determinant to approximate the wave function and these calculations yield $\approx 99\%$ of the exact total energy.

$$\Psi^{\mathrm{HF}} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_1(\mathbf{x}_1) & \psi_2(\mathbf{x}_1) & \dots & \psi_N(\mathbf{x}_1) \\ \psi_1(\mathbf{x}_2) & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \psi_1(\mathbf{x}_N) & \dots & \dots & \psi_N(\mathbf{x}_N) \end{vmatrix} = |\psi_1(\mathbf{x}_1)\psi_2(\mathbf{x}_2)\dots\psi_N(\mathbf{x}_N)\rangle, \tag{11}$$

where $\psi_i(\mathbf{x}_j)$ is a molecular spin-orbital with the vector $\mathbf{x}_j$ that holds the spatial and spin coordinates of the electron j.

$\mathbf{x}_j = \{\mathbf{r}_j, \mathbf{s}_j\}$

Using a Slater determinant as an *Ansatz* for the wave function guarantees the compliance with the Pauli principle.

In this work, only closed-shell systems have been examined. Closed-shell systems have the same number of α and β spin electrons and the spatial orbitals $\{\phi_i\}$ are the same for both sets (α and β).

The energy expectation value for the closed shell case can thus be written in the form:

$$E_{HF} = \langle\Psi^{\mathrm{HF}}|\hat{H}|\Psi^{\mathrm{HF}}\rangle = \sum_{i}^{N/2} 2\langle i|\hat{h}|i\rangle + \sum_{ij}^{N/2} [2(ii|jj) - (ij|ji)] \tag{12}$$

with the one and two electron integrals given by

$$\langle i|\hat{h}|j\rangle = \int \phi_i^*(\mathbf{r}_1)\hat{h}(\mathbf{r}_1)\phi_j(\mathbf{r}_1)d\mathbf{r}_1 \tag{13}$$

$$(ij|kl) = \int \phi_i^*(\mathbf{r}_1)\phi_j(\mathbf{r}_1)r_{12}^{-1}\phi_k^*(\mathbf{r}_2)\phi_l(\mathbf{r}_2)d\mathbf{r}_1 d\mathbf{r}_2 \tag{14}$$

Since Hartree-Fock is by construction a variational method, the orbitals are obtained by minimizing the energy expression, under the side condition that the orbitals are orthonormal. This is done via a Lagrangian formulation. The Lagrangian is given by:

$$\mathcal{L} = \langle\Psi^{\mathrm{HF}}|\hat{H}|\Psi^{\mathrm{HF}}\rangle - 2\sum_{ij}^{N/2} \epsilon_{ji} \left[\langle i|j\rangle - \delta_{ij}\right], \tag{15}$$

with $\langle i|j\rangle$ an overlap integral and Lagrangian multipliers ϵ_{ji}.

The Lagrangian is invariant with respect to unitary transformations amongst the occupied orbitals. Thus one can define a canonical orbital basis in which $\epsilon_{ij} = \delta_{ij}\epsilon_i$. After several steps, the minimization of the Lagrangian then leads to the canonical HF equation

$$\hat{f}|i\rangle = \epsilon_i|i\rangle \tag{16}$$

where $\hat{f}$ is the Fock operator defined as

$$\hat{f}(i) = \hat{h}(i) + \sum_j \left[2\hat{J}_j(i) - \hat{K}_j(i)\right] = \hat{h}(i) + \hat{g}(i). \tag{17}$$

The Coulomb- $\hat{J}_j(i)$ and exchange- $\hat{K}_j(i)$ operators are defined as

$$\hat{J}_j(1)\phi_i(\mathbf{r}_1) = \int \phi_j^*(\mathbf{r}_2)\frac{1}{r_{12}}\phi_j(\mathbf{r}_2)\phi_i(\mathbf{r}_1)d\mathbf{r}_2, \tag{18}$$

$$\hat{K}_j(1)\phi_i(\mathbf{r}_1) = \int \phi_j^*(\mathbf{r}_2)\frac{1}{r_{12}}\phi_j(\mathbf{r}_1)\phi_i(\mathbf{r}_2)d\mathbf{r}_2. \tag{19}$$

In the canonical basis, the Fock matrix is diagonal. The diagonal elements are the orbital energies ϵ_i. Note that the sum of the occupied orbital energies is different from the HF energy.

The Linear Combinations of Atomic Orbitals (LCAO), as the name already reveals, is a technique for forming molecular orbitals as a linear combination of atomic orbitals (AOs).

$$\phi_i = \sum_\mu C_{\mu i}\chi_\mu \tag{20}$$

With this, the HF energy expressed in the AO basis is given by

$$\begin{aligned} E_{HF} &= \sum_{\mu\nu} D_{\mu\nu}\{\ h_{\mu\nu} + \frac{1}{2}\sum_{\rho\sigma} D_{\rho\sigma}[(\mu\nu|\rho\sigma) - \frac{1}{2}(\mu\sigma|\rho\nu)]\} \\ &= \frac{1}{2}\sum_{\mu\nu} D_{\mu\nu}(\ h_{\nu\mu} + \ f_{\nu\mu}), \end{aligned} \tag{21}$$

where $h_{\nu\mu}$ and $f_{\nu\mu}$ represent $\hat{h}$ and $\hat{f}$ in the AO basis.

The density matrix D is defined as

$$D_{\mu\nu} = 2\sum_{i}^{\frac{N}{2}} C_{\mu i} C_{\nu i}^{*}. \tag{22}$$

with the coefficients $C_{\mu i}$ and $C_{\nu i}^{*}$ from **Equation (20)**.

If the canonical HF equation (16) is expressed in the AO basis, we obtain the Hartree-Fock-Roothan equation

$$\mathbf{FC} = \mathbf{SCE} \tag{23}$$

where **F** is the Fock matrix in AO basis, **C** the MO coefficient matrix, **S** is the overlap matrix and **E** the (diagonal) matrix with orbital energies as diagonal elements. It must be noted that both **Equations (16)** and **(23)** are implicit non-linear equations, because the Fock operator depends on the occupied orbitals via **(18)**. They are thus usually solved in an iterative fashion with the so-called self-consistent-field method. Starting with an initial guess for the orbitals, Fock operators are constructed and diagonalized to find new occupied orbitals, until the the occupied orbitals used to construct the Fock operator are also its eigenvectors. This is called self-consistency.

3.3 Basis sets

The molecular orbitals used in all calculations of this work are formed by linear combination of atomic orbitals, as described earlier in **Section 3.2**. In most cases of *ab initio* calculations Slater or Gaussian type functions are used.

While Slater type basis sets have the advantage of being able to describe molecular orbitals more efficiently, integrals over Gaussian type basis functions can be calculated more easily. Therefore so called contracted Gaussian functions are used to approximate Slater type functions. Those χ_α were exclusively used in this work. In the Cartesian form the Gaussian type functions take the form:

$$\chi_{\mu}^{klm}(x,y,z) = x^k y^l z^m \sum_{i} C_{i\mu} e^{[-\alpha_{\mu i}(x^2+y^2+z^2)]} \tag{24}$$

where α_i are the exponents and C_i the contraction coefficients. The sum $k+l+m = L$ is the specifier of the type, with $k+l+m = 0, 1, 2, ...$ for $s-, p-, d-, ...$ type functions and x, y, z are cartesian coordinates.

More frequently than the Cartesian form of the Gaussian basis functions, the spherical

form is used:

$$\chi_{\mu}^{Lm}(x,y,z) = S_m^L(x,y,z) \sum_i C_{i\mu} e^{[-\alpha_{\mu i}(x^2+y^2+z^2)]} \tag{25}$$

Here S_m^L are the so-called solid harmonics,[15] which are linear combinations of Cartesians $x^k y^l z^m$ with $k+l+m=L$ (if real basis functions are used).

In this work only Dunning-type basis sets[16] have been used. Those Dunning-type basis sets are correlation consistent-polarized Valence n-Tuple Zeta (cc-pVnZ) where n stands for the highest angular momentum (L) used. There are also Dunning-type bases with additional diffuse functions, they are named Augmented cc-pVnZ (aug-cc-pVnZ) or short AVnZ. All of those are exclusively atom centered basis sets. Other basis sets have additional "bond functions" that are beneficial for intermolecular energy calculations.[17],[18] However, "bond functions" have not been used in order to compare with the conventional methods by using the same basis set.

For explicitly correlated calculations (see **Section 4.3**) there are sets of especially optimized basis sets[19] that improve over the Dunning-type basis sets in terms of convergence to the basis set limit. Those have not been used for the same reasons as "bond functions" have been neglected.

3.4 Post Hartree-Fock methods

For many *ab initio* methods the wave function is approximated by a linear combination of the Hartree-Fock Slater determinant and excited configurations:

$$\Psi = \Psi^{HF} + \sum_{ia} c_a^i \Phi_i^a + \frac{1}{2} \sum_{ij} \sum_{ab} c_{ab}^{ij} \Phi_{ij}^{ab} + \dots \tag{26}$$

The excited configurations $|\Phi_i^a\rangle$, $|\Phi_{ij}^{ab}\rangle$, ... are constructed via excitations operators from the HF reference

$$|\Phi_i^a\rangle = \hat{E}_{ai} |\Psi^{\mathrm{HF}}\rangle \tag{27}$$

$$|\Phi_{ij}^{ab}\rangle = \hat{E}_{ai} \hat{E}_{bj} |\Psi^{\mathrm{HF}}\rangle \tag{28}$$

$$\dots \tag{29}$$

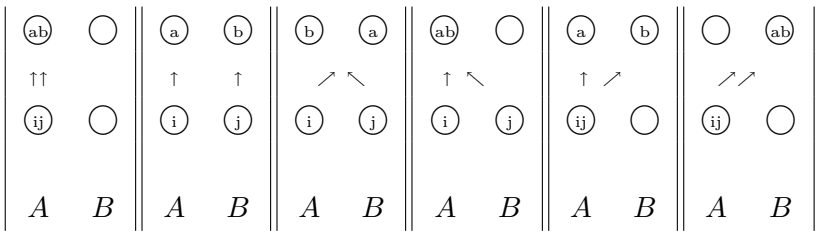
Here $\hat{E}_{ai}$ ($\hat{E}_{bj}$) are spin-adapted excitation operators which promote one electron from an occupied orbital i (j) to a virtual orbital a (b). The difference between the exact energy and the Hartree-Fock energy is the electron correlation energy[20] ($E_{corr.}$) (see **Figure 3**). For single reference cases the electron correlation is dynamic and accounts for short-range electron repulsion not described by HF.

$$E_{corr.} = E_{exact} - E_{HF} \tag{30}$$

The Full CI energy is often taken as E_{exact}.

3.5 Basis Set Superposition Error

The Basis Set Superposition Error[21] is the result of an incomplete basis causing qualitative and quantitative errors due to superposition of basis sets. In the case of weak interactions the BSSE results from the fact that in the calculation of the super-system AB both monomers A and B (equal or unequal) have also the basis functions of the other monomer χ_A and χ_B respectively for the description of the molecular orbitals. This is present in Hartree-Fock and post Hartree-Fock methods.
The BSSE of the correlation contribution in conventional *ab initio* post Hartree-Fock methods in a system of two monomers A and B forming a super-system AB is shown in **Figure 1**.



a,a' **b** **c** **d,d'** **e,e'** **f,f'**

The excitation(s)
a is intramolecular correlation
b describes London dispersion force
c describes exchange
d and **e** describe ionic interactions and
f are excitations that cause BSSE.
n' is **n** with exchanging A and B.

Figure 1: Types of excitations in conventional methods

The Basis Set Superposition Error in the interaction energy can be corrected with the Counterpoise Correction as shown in the following **Section 3.6**.

3.6 Counterpoise Correction

As mentioned previously in **Section 3.5**, the BSSE[21] can be corrected by using a method called *Counterpoise Correction* after Boys and Bernardi.[22]

In a calculation of a super-system AB the description of both monomers is improved by the basis functions of the other monomer respectively. In the calculations of the monomers A and B they only have there own basis functions and are therefore not described as well as in the super-system. That leads to an unbalanced description which results in an overestimation of the interaction energy. This effect gets smaller with increasing size of the basis set or the introduction of local methods.[23] The uncorrected interaction energy ΔE is:

$$\Delta E = E(AB_{\chi_A\chi_B}) - [E(A_{\chi_A}) + E(B_{\chi_B})] \tag{31}$$

In order to correct the BSSE in the interaction energy the basis set of the super-system (χ_A and χ_B) is used for both monomers. This is done by the introduction of the so called *dummy atoms* that hold the basis functions of the missing monomer but have no charge. In that way, the BSSE is also introduced in the monomer calculations leading to a balanced description of the interaction energy.

The Counterpoise corrected interaction energy ΔE^{CP} is thus:

$$\Delta E^{CP} = E(AB_{\chi_A\chi_B}) - [E(A_{\chi_A\chi_B}) + E(B_{\chi_A\chi_B})] \tag{32}$$

The difference between the Counterpoise corrected interaction energy ΔE^{CP} and the uncorrected interaction energy ΔE is the Counterpoise Correction E^{CPC}:

$$E^{CPC} = \Delta E - \Delta E^{CP} = [E(A_{\chi_A\chi_B}) + E(B_{\chi_A\chi_B})] - [E(A_{\chi_A}) + E(B_{\chi_B})] \tag{33}$$

At most, the calculation time is ~ 2 times as expensive (MP2[a])[b] for a super-system with monomers of equal size and drops down to approximately a factor of $\sim 1\frac{1}{4}$ for CCSD(T)[c] since the calculation time of the monomers without the Counterpoise Correction is negligible compared to the calculation of the super-system. The Counterpoise Correction

[a] canonical MP2

[b] here the monomers are half as expensive compared to the supersystem, based on a cost scaling estimate of mN^4 for the integral transformation (m: occupied orbitals, N: AO basis)

[c] here the monomers are a eighth as expensive compared to the supersystem, based on a cost scaling estimate of m^3V^4 for the triples (m: occupied orbitals, V: virtual orbitals)

does not correct for the BSSE introduced by explicit correlation (see **Section 4.3.8**), this is taken care of by the selection of an appropriate Ansatz for the calculation (see **Section 4.3.3**).

3.7 Hierarchy of *ab initio* methods

The hierarchy of the Hartree-Fock method and single reference post Hartree-Fock methods is presented according to the level of theory but holds also for the average quality as well as for the calculation time. In this limited selection of *ab initio* methods, each method is also a subset of the higher level method.

The hierarchy of *ab initio* methods used in this work towards FCI is:

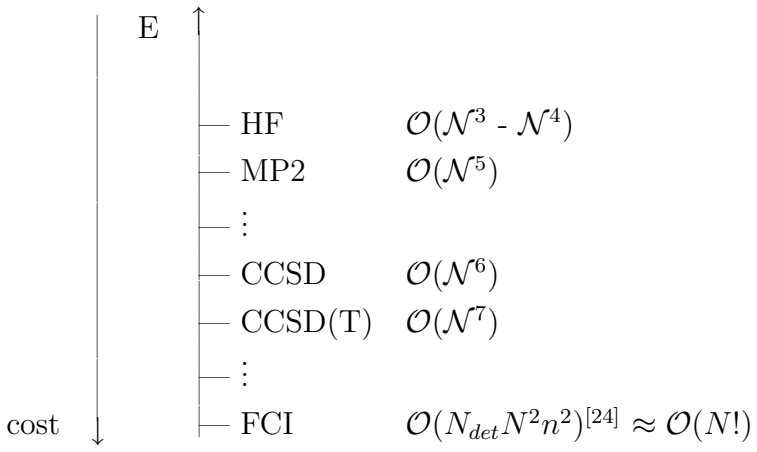


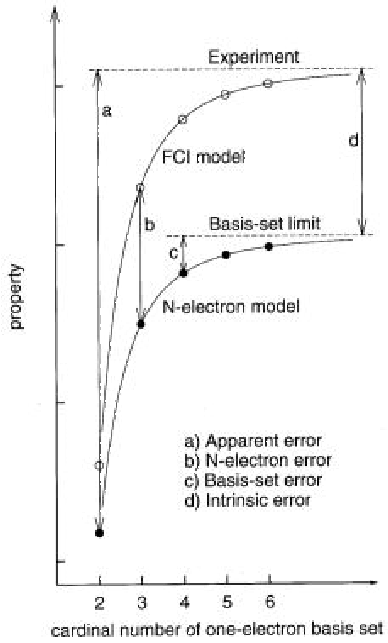
Figure 2: The hierarchy of single-reference *ab initio* methods

The smaller the deviation from the exact (FCI) solution, the more expensive[d] the method becomes. The CCSD(T) method (see **Section 3.10**) is often called the *gold standard* [e] and is used as the reference in all calculations in this work.

[d] demanding in calculation time memory and hard disk usage
[e] for its excellent compromise of accuracy and cost

3.8 Errors in QM calculations

The error that is made in quantum mechanical calculations can be considered as a two dimensional one with the two dimensions intrinsic error and basis set error. That is only correct if one assumes that neglecting relativistic effects and assuming that the Born-Oppenheimer approximation is not introducing a defect to the wave function.



a) Apparent error:
deviation at a certain basis and method.
b) N-electron-error:
deviation from FCI at a certain basis.
c) Basis set error:
deviation from basis set limit.
d) Intrinsic error:
deviation from FCI at the basis set limit.

Figure 3: Error in QM calculations.[25] Figure taken from [25].

The assumption that relativistic effects are a negligible source of error holds well for the calculation of weak interactions if only first and second row atoms are involved. The Born-Oppenheimer approximation is valid and can be used savely as long as only ground state single reference systems are investigated and (therefore) no conical intersection is involved.

3.9 Second-order Møller-Plesset perturbation theory

As mentioned in the Hartree-Fock **Section 3.2**, the HF approximation yields ≈ 99% of the exact energy. Therefore, the requirement of perturbation theory, that the perturbation is small, is generally given. The Hamiltonian in perturbation theory is separated in two parts, the reference $\hat{H}^{(0)}$ and the perturbation $\hat{H}^{(1)}$. To obtain the energy corrections of order up to $2n+1$, the wave-function only has to be known up to nth order (Wigner's $2n+1$ rule). Hence in second-order Møller-Plesset perturbation theory (MP2) one only has to calculate the first-order wave function to obtain the second-order perturbation energy. As a reference $\Psi^{(0)}$ the HF wave function will be used, which by construction is an eigenfunction of $\hat{H}^{(0)}$ in Møller-Plesset perturbation theory. Summed up, the energies of order $n=0$ and $n=1$ will give the HF energy:

$$E^{(0)} = \langle \Psi^{\mathrm{HF}} | \hat{H}^{(0)} | \Psi^{\mathrm{HF}} \rangle = \langle \Psi^{\mathrm{HF}} | \sum_{i}^{N} \hat{f}_i | \Psi^{\mathrm{HF}} \rangle = \sum_{i}^{N} \epsilon_i \tag{34}$$

$$E^{(1)} = \langle \Psi^{\mathrm{HF}} | \hat{H}^{(1)} | \Psi^{\mathrm{HF}} \rangle = -\frac{1}{2} \sum_{ij} [2(ii|jj) - (ij|ji)] \tag{35}$$

$$E_{HF} = \langle \Psi^{\mathrm{HF}} | \hat{H}^{(0)} + \hat{H}^{(1)} | \Psi^{\mathrm{HF}} \rangle = E^{(0)} + E^{(1)} \tag{36}$$

The MP2 correction is thus comprised in the term $E^{(2)}$. This energy term already involves the first-order wave function $\Psi^{(1)}$, which must be known. According to Brillouin's theorem, singly excited configurations will not interact with the $\Psi^{(0)}$ wave function[f] if the HF orbitals are optimized, i.e. $\langle \Phi_i^a | \hat{H} | \Psi^{\mathrm{HF}} \rangle = 0$. A linear combination of doubly excited configurations build the first-order wave function:

$$|\Psi^{(1)}\rangle = \frac{1}{2} \sum_{ij} \sum_{ab} T_{ab}^{ij} |\Phi_{ij}^{ab}\rangle \tag{37}$$

The doubly excited configurations $|\Phi_{ij}^{ab}\rangle$ are constructed via excitation operators from the HF reference as in **Equation (28)** where $\hat{E}_{ai}$ ($\hat{E}_{bj}$) excites spin-adapted from an occupied orbital i (j) to a virtual orbital a (b).

The configurations Φ_{ij}^{ab} are not orthonormal, therefore contra-variant configurations and

[f] as long as the basis set used is the same (see **Section 4.3.7**) and therefore have no contribution to the energy.

amplitudes are used:

$$\tilde{\Phi}_{ij}^{ab} = \frac{1}{6}(2\Phi_{ij}^{ab} + \Phi_{ji}^{ab}) \quad (38)$$

$$\tilde{T}_{ab}^{ij} = 2T_{ab}^{ij} - T_{ab}^{ji} \quad (39)$$

They have the properties:

$$\langle\tilde{\Phi}_{ij}^{ab}|\Phi_{kl}^{cd}\rangle = \delta_{ac}\delta_{bd}\delta_{ik}\delta_{jl} + \delta_{ad}\delta_{bc}\delta_{il}\delta_{jk} \quad (40)$$

$$\langle\tilde{\Phi}_{ij}^{ab}|\Psi^{(1)}\rangle = T_{ab}^{ij} \quad (41)$$

$$\langle\tilde{\Phi}_{ij}^{ab}|\hat{H}|\Psi^{\mathrm{HF}}\rangle = K_{ab}^{ij} \quad (42)$$

This step provides a simplification that leads to the formulation of the first correction to HF:

$$\Delta E_{\mathrm{MP2}} = E^{(2)} = \langle\Psi^{\mathrm{HF}}|\hat{H}|\Psi^{(1)}\rangle = \sum_{ij}\sum_{ab}\langle\Psi^{\mathrm{HF}}|\hat{H}|\tilde{\Phi}_{ij}^{ab}\rangle\tilde{T}_{ab}^{ij}$$
$$= \sum_{ij}\sum_{ab} K_{ab}^{ij}\tilde{T}_{ab}^{ij} \quad (43)$$

where the term $K_{ab}^{ij} = (ia|jb)$ is an exchange integral.

The amplitudes are obtained by solving the first-order perturbation equations

$$R_{ab}^{ij} = \langle\tilde{\Phi}_{ij}^{ab}|\hat{H}^{(0)} - E^{(0)}|\Psi^{(1)}\rangle + \langle\tilde{\Phi}_{ij}^{ab}|\hat{H}|\Psi^{(0)}\rangle = 0 \quad (44)$$

which leads to:

$$R_{ab}^{ij} = K_{ab}^{ij} + \sum_{c}(f_{ac}T_{cb}^{ij} + T_{ac}^{ij}f_{cb}) - \sum_{k}(f_{ik}T_{ab}^{kj} + T_{ab}^{ik}f_{kj}) \quad (45)$$

In the canonical case all off diagonal elements f_{rs} zero out and **Equation (45)** can be written in the form:

$$R_{ab}^{ij} = K_{ab}^{ij} + (\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j)T_{ab}^{ij} = 0 \quad (46)$$

leading to:

$$T_{ab}^{ij} = -\frac{K_{ab}^{ij}}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j} \quad (47)$$

From **Equation (43)** one can get to the most popular MP2 formulation:

$$E_{\mathrm{MP2}} = \sum_{ij} \sum_{ab} \frac{K_{ab}^{ij}(2K_{ab}^{ij} - K_{ab}^{ji})}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b}. \tag{48}$$

Second-order is the most often used level in perturbation theory but also higher orders have their applications. The MPn series converges for some cases in a zig-zag-oscillating manner, for other cases it diverges.

The MP2 method can be considered as a low cost correlation method and scales with $\mathcal{O}(\mathcal{N}^5)$ in the canonical case. Its energies are size-consistent and not variational. Unfortunately in some cases, MP2 leads to significant errors in the energy. For a small HOMO-LUMO gap the denominator of **Equation (48)** ($\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b$) may become very small and therefore the term $\frac{1}{\epsilon_i+\epsilon_j-\epsilon_a-\epsilon_b}$ can become very big, introducing a large error. Also, if the correlation correction to HF is very big perturbation theory is not the best choice.

3.9.1 Size consistency

In a size consistent method, two molecules separated by a large distance [g] both calculated simultaneously have exactly the sum of the energy of the isolated molecule calculated separately. Size consistency is an important property of most modern methods used on a regular basis.

3.9.2 Spin Component Scaled MP2

Any linear combination of configurations (Slater determinants) with the excitation level of two ($|\Phi_{ij}^{ab}\rangle$) can be separated into a parallel (both electrons in virtual orbitals have the same spin) and an antiparallel (both electrons in virtual orbitals have opposite spin) part. Those energy contributions can be calculated separately, leading in the MP2 case to correlation contributions in the form:

$$E_{\mathrm{MP2}}^{corr} = E_{\mathrm{MP2}}^{sing} + E_{\mathrm{MP2}}^{trip} \tag{49}$$

The idea of Spin Component Scaled MP2[5] (SCS-MP2) as described in the paper "*Improved second-order Møller-Plesset perturbation theory by separate scaling of parallel- and*

[g] large enough to be a non-interacting super system

antiparallel-spin pair correlation energies" is to scale parallel and antiparallel (alternatively singlet and triplet) components of the correlation contribution with empirical factors:

$$E_{\mathrm{SCS}}^{corr} = 1.2\, E_{\mathrm{MP2}}^{sing} + 0.6\overline{2}\, E_{\mathrm{MP2}}^{trip} \tag{50}$$

According to by Grimme *et al.*,[5] SCS-MP2 improves the Mean Absolute Deviation (MAD), Root Mean Square (RMS) and maximal errors (MAX) over MP2 when compared with QCISD(T)[h][26] for a set of 51 reaction energies.

error	MP2	SCS-MP2
MAD	3.3	1.8
RMS	4.6	2.3
MAX	13.3	5.1

Table V: Improvement of SCS-MP2 over MP2

3.9.3 Local correlation MP2

The idea of local correlation methods[27] is to exploit the fact that dynamic electron correlation is a short range effect that, and similarly to the dispersion energy, is decaying approximately with r^{-6}. Therefore to avoid the steep scaling of the computational cost with the system size, local approximations capitalize on this. The calculation of electron pairs with greater distance are completely neglected (or sometimes calculated less accurately) which can lead down to a linear scaling of the computational cost with the system size.

In local methods one uses localized orbitals[27] in order to be able to restrict excitations from occupied orbitals $|\phi_i^{\mathrm{loc}}\rangle$ to domains of virtual orbitals. The localized orbitals are formed by a transformation of the occupied space into a local orbital basis:[28]

$$|\phi_i^{\mathrm{loc}}\rangle = \sum_{\mu} |\chi_\mu\rangle L_{\mu i} = \sum_{k} |\phi_k^{\mathrm{can}}\rangle U_{ki} \tag{51}$$

There are different ways to undertake this transformation with different localization schemes. The two most common being the Boys[29] and the Pipek-Mezey[30] schemes. The Pipek-Mezey scheme is often preferred due to its *physical looking* orbitals.

The usage of localized orbitals is not an approximation as such if an orbital invariant

[h] Quadratic Configuration Interaction

method is used and no orbital pairs ij are neglected. Since MP2 and its explicitly correlated version can be formulated as orbital invariant methods, the MP2 energy within the accuracy of the selected numerical threshold is the same. The use of Projected Atomic Orbitals (PAOs) and the excitations from an occupied orbital $|\phi_i^{\text{loc}}\rangle$ to virtual orbitals will only introduce an approximation if the excitations are restricted to domains of virtual orbitals that are smaller than the total virtual orbital space.

The LMP2 correlation energy is:

$$E^{(2)} = \sum_{ij} \sum_{rs \in [ij]} K_{rs}^{ij} (2T_{rs}^{ij} - T_{sr}^{ij}), \tag{52}$$

where $[ij]$ denotes a domain for the orbital pair ij.

3.9.4 MP2.5

The very pragmatic Ansatz of MP2.5[6] by Grimme *et al.* is to mix MP2 and MP3 (energies) half and half:

$$E_{MP2.5} = \frac{1}{2} E_{MP2} + \frac{1}{2} E_{MP3}. \tag{53}$$

The fact that, while MP2 usually overshoots, MP3 typically undershoots brought Grimme *et al.* to this very simple way of mixing the two and naming it as pragmatic as possible.

3.10 Coupled Cluster

The Coupled Cluster (CC) theory is, like MP2, non-variational[i] and size consistent. Its fast convergence behavior towards the FCI limit together with size consistency has made it a very popular method since the available computational power has greatly increased in the last years.

The CC wave function can be written in the form:

$$|\Psi^{\mathrm{CC}}\rangle = e^{\hat{T}}|\Psi^{\mathrm{HF}}\rangle \tag{54}$$

The excited configurations are created by the cluster operator $\hat{T}$ acting on the reference function.

The cluster operator includes excitation operators up to a given order: $\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \cdots + \hat{T}_N$, with the definition of the excitation operators being:

$$\hat{T}_1 = \sum_i \sum_a \hat{E}_{ai} t_a^i \tag{55}$$

$$\hat{T}_2 = \frac{1}{2} \sum_{ij} \sum_{ab} \hat{E}_{ai} \hat{E}_{bj} T_{ab}^{ij} \tag{56}$$

$$\vdots$$

The exponential operator ($e^{\hat{T}}$) of **Equation (54)** is expanded in a Taylor series:

$$e^{\hat{T}} = 1 + \hat{T} + \frac{\hat{T}^2}{2!} + \frac{\hat{T}^3}{3!} + \dots \tag{57}$$

The Coupled Cluster method will give the FCI result if the order of the excitation operator matches the number of electrons.
It is common usage to truncate the cluster operator to double excitations. Since products of single and double operators create configurations which in principle can be seen as excitations of higher level, the Coupled Cluster Singles and Doubles method (CCSD) gives already a (*somehow rather good*) approximation towards FCI.

For CCSD the expansion takes the form:

$$e^{\hat{T}_1+\hat{T}_2} = 1 + \hat{T}_1 + \left(\hat{T}_2 + \frac{\hat{T}_1^2}{2}\right) + \left(\hat{T}_2\hat{T}_1 + \frac{\hat{T}_1^3}{6}\right) + \dots \tag{58}$$

[i] if the order of excitations is restricted to a number smaller that the number of electrons

where terms like $\hat{T}_2\hat{T}_1$ and $\hat{T}_1^3$ represent excitations of a level higher than 2. Applying the Hamilton operator and projecting from the left with the reference gives the CCSD correlation energy.

$$\begin{aligned}\Delta E^{\mathrm{CCSD}} =& \langle \Psi^{\mathrm{HF}} | \hat{H} - E_0 | \Psi^{\mathrm{CCSD}} \rangle \\ =& \sum_{ai} \langle \Psi^{\mathrm{HF}} | \hat{H} | \Phi_i^a \rangle t_a^i \\ &+ \sum_{ij} \sum_{ab} \langle \Psi^{\mathrm{HF}} | \hat{H} | \Phi_{ij}^{ab} \rangle \left(T_{ab}^{ij} + \frac{1}{2} t_a^i t_b^j \right).\end{aligned}$$

The explicit expressions in terms of integrals and amplitudes is:

$$\Delta E^{\mathrm{CCSD}} = \sum_{ai} 2 f_{ai} t_a^i + \sum_{abij} \left[2\mathbf{K}^{ij} - \mathbf{K}^{ji} \right]_{ab} \left(T_{ab}^{ij} + \frac{1}{2} t_a^i t_b^j \right) \tag{59}$$

As mentioned before, Coupled Cluster has a fast convergence behavior towards FCI if excitations of higher order are included. The CCSD energy can be improved by including triple excitations at a perturbation theory level.

In the CCSD(T) method[31],[32] the contributions of the connected triple excitations are taken into account in a perturbative, non-iterative way. This leads to a substantial improvement of the overall accuracy.

The energy of the CCSD(T) improves[33],[34] over the energy of CC with inclusion of full single, double and triple excitations (CCSDT)[35]–[37] (somehow unexpected) in the statistical average, due to an error compensation and is therefore often called the *gold standard* in *ab initio* quantum chemistry calculations. Furthermore CCSD(T) has the advantage of its scaling dependency of $\mathcal{N}^7$, *i.e.*, one exponentiation lower than CCSDT.

Because of its convergence behavior towards FCI and the fact that the (T) correction is done independently, it is one of the most accurate methods still suitable for medium size systems[j] hence becoming the standard method for high accuracy single reference calculations.

[j] up to $\approx$ 30 to 40 first and second row atoms using AVDZ basis

4 Incompleteness in the description of molecular orbitals

The quality (size) of the basis sets used, (see **Section 3.3**) is vital for the accurate description of the molecular orbitals and the building of the excited configurations to converge the total energy towards the basis set limit.

4.1 Convergence of the total energy

The convergence of the total energy to its limit with respect to the basis set used is very slow ($\approx n^{-3}$) while the increase in cost is at least $\rightarrow \mathcal{O}(\mathcal{N}^4)$[a] (MP2[b]) with the basis set size.

The main reason for this lies in the difficulty to describe the electron-electron cusp by expansions of the wave function in terms of orbital products.

Figure 4 shows the convergence of the total energy of MP2 with the increase of the basis set size for the system H_2O. This behavior is alike for all systems with electron-electron correlation using any conventional *ab initio* post Hartree-Fock method. The red curve shows the convergence properties for the VnZ basis and the green curve for the AVnZ basis.

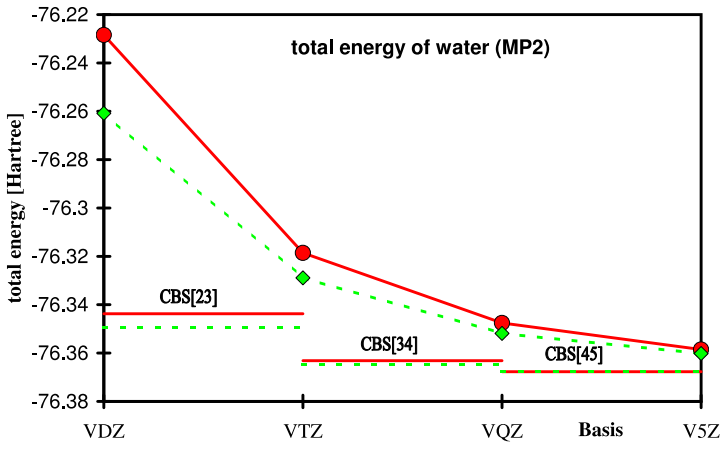


Figure 4: Convergence of the total energy

[a] system size dependent
[b] using density fitting (DF-MP2) $\rightarrow \mathcal{O}(\mathcal{N}^3)$

In **Table VI** this convergence behavior for a conventional *ab initio* post Hartree-Fock method is presented. The values are taken from "*Introduction to computational chemistry*[38]".

Basis set quality	Correlation energy relative to the CBS limit
cc-pVDZ	$\sim 65\%$
cc-pVTZ	$\sim 85\%$
cc-pVQZ	$\sim 93\%$
cc-pV5Z	$\sim 96\%$
cc-pV6Z	$\sim 98\%$

Table VI: Convergence of the correlation energy to the CBS limit with basis set size

In order to reduce the error caused by the basis set incompleteness in the total energy, extrapolation methods are a possible solution. The most common basis set extrapolation method is presented in the next **Section 4.1.1**.

4.1.1 Basis set extrapolation

As mentioned in **Section 4.1** and shown in **Figure 4** the total energy of a system slowly converges with the basis set size. This convergence is quite systematic and from the behavior in atoms one can derive an analytic formula for the extrapolation of the correlation energy that also holds for molecules:

$$E_n^{corr} = E_\infty^{corr} + an^{-3} \tag{60}$$

where n is the highest angular momentum in the AO basis.

It is applied by using two[c] (or more) different basis sets to do an extrapolation to the basis set limit. The extrapolation scheme is used to estimate the limit of the correlation contribution[39] only. Often the Hartree-Fock energy is not extrapolated but the energy of the largest basis set is used as the limit. That is usually a good approximation since the HF energy converges faster to the basis set limit with the basis set size than the correlation contribution. There are some approaches to do an extrapolation of the HF[40],[41] energy, but those have not been used in this work. The extrapolation of the correlation

[c] due to the fact that there are two unknown in the **Equation (60)** two different basis sets are needed

contribution starts being reasonably accurate using VTZ and VQZ or better VQZ and V5Z correlation consistent basis sets with or without diffuse functions[42] (cc-pVnZ or aug-cc-pVnZ). Estimating the Complete Basis Set limit (CBS) via extrapolation from n=3,4 is denoted CBS[34] or n=4,5 is denoted CBS[45] and so on. Extrapolation methods are not completely unproblematic when using augmented basis sets due to a possible BSSE.

4.2 Motivation for F12 methods

4.2.1 Electron-electron cusp

As mentioned in the previous **Section** (**4.1**), the electron-electron cusp is the main reason for the slow convergence to the basis set limit. **Figure 5**[d] below shows the wave function Ψ of the system according to the distance (r_{12}) (represented by the angle[e] ϑ) of *red* electron to the *blue* electron. Low absolute values of Ψ also indicate a low probability density Ψ^2 for one electron being found close to an other. The cusp itself originates from the cancellation of the singularity of the electron-electron Coulomb potential at $r_{12} = 0$ by a compensating singularity of the kinetic energy operator. This allows for $H\Psi$ to be finite even at electron coalescence.

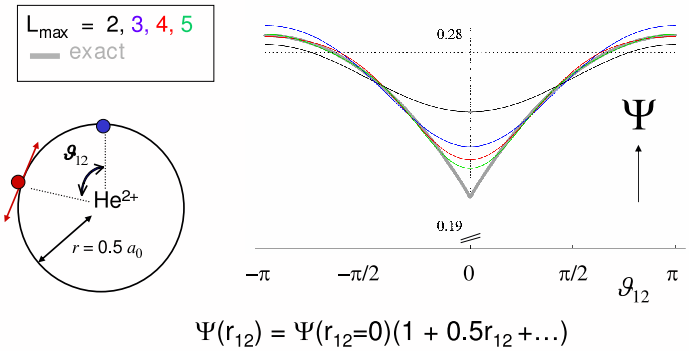


Figure 5: Electron-electron cusp

[d] (thanks to Dr. David Tew)
[e] $r_{12} = 2r|sin(\frac{\vartheta_{12}}{2})|$

One can see that the wave function Ψ can be approximated quite well for larger angles of ϑ_{12} but the area around the cusp is never represented accurately enough by expansions of the wave function in terms of orbital products, even if large basis sets are used. To solve that problem and speed up the convergence to the basis set limit, R12/F12 methods (see **Section 4.3**) are employed.

4.3 R12/F12 concept

The general idea of R12/F12 methods is to improve the convergence of the correlation energy with the basis set size. In F12 methods this is done by including terms in the wavefunction that depend explicitly on the interelectronic distance r_{12}.

This concept goes back as early as 1929 when Hylleraas[43] used explicitly electron-distance dependent functions to describe the Helium atom. But since in many-electron systems the explicitly correlated terms lead to numerous 3-electron and 4-electron integrals that are extremely costly the whole idea of F12 was unpractical for larger molecules and did not get used for long time. Only when the technique called resolution of the identity (RI)[44]–[46] was introduced (see **Section 4.3.6**) the F12 methods were made feasible.

4.3.1 F12 wave function

This **Section** is intended to give a very rough overview on the general idea of the F12 wave function. In the upcoming two **Sections** (**4.3.3** and **4.3.4**) a more detailed description about the used F12 methods is given.

The general expression of the wavefunctions written in a form that holds for MP2 (CCSD) and MP2-F12 (CCSD-F12) is:

$$\Psi_{\mathrm{MP2(-F12)}} = (1 + \widehat{T}_2)\Psi_{\mathrm{HF}} \tag{61}$$

$$\Psi_{\mathrm{CCSD(-F12)}} = e^{\widehat{T}_1 + \widehat{T}_2}\Psi_{\mathrm{HF}} \tag{62}$$

For explicitly correlated methods the double excitation operator $\widehat{T}_2$ is extended by an additional term with the explicitly correlated amplitudes $\mathcal{T}_{\alpha\beta}^{ij}$. The single excitations, only present in **Equation (62)**, are not expanded because the explicit correlation acts

on electron pairs:

$$\widehat{T}_1 = \sum_{ia} t_a^i \widehat{E}_i^a \tag{63}$$

$$\widehat{T}_2 = \frac{1}{2}\sum_{ijab} T_{ab}^{ij}\widehat{E}_{ij}^{ab} \underbrace{+\frac{1}{2}\sum_{ij\alpha\beta} \mathcal{T}_{\alpha\beta}^{ij}\widehat{E}_{ij}^{\alpha\beta}}_{\text{explicit correlation}} \tag{64}$$

$$= \frac{1}{2}\left(\sum_{ijab} T_{ab}^{ij}\hat{E}_{ai}\hat{E}_{bj} + \sum_{ij\alpha\beta}\mathcal{T}_{\alpha\beta}^{ij}\hat{E}_{\alpha i}\hat{E}_{\beta j}\right)$$

where the explicitly correlated amplitudes are build by the projector $\widehat{Q}_{12}$ and the correlation factor F_{12} as

$$\mathcal{T}_{\alpha\beta}^{ij} = \sum_{mn} T_{mn}^{ij}\langle mn|F_{12}\widehat{Q}_{12}|\alpha\beta\rangle \tag{65}$$

$$\tag{66}$$

	$i, j, k, l, \ldots$		occupied orbitals
with the indices	a, b	representing the	virtual orbitals
	α, β		complete virtual orbital basis

and F_{12} is defined in **Section 4.3.2**, (**Equation (67)**). There are essentially three Ansätze for $\widehat{Q}_{12}$ and they will be discussed later. At this point it shall only be mentioned that it ensures strong orthogonality of the explicitly correlated part to the reference part.

4.3.2 F12 correlation factor

As a further development of early R12 methods, where the excitation factor was linear ($F(r_{12}) = r_{12}$), modern methods employ a Slater type function of the inter-electronic distance.

$$F(r_{12}) = -\frac{1}{\gamma}\exp(-\gamma r_{12}) \tag{67}$$

This was first proposed by Ten-no[47] and subsequently used by several other authors.[47]–[55] The exponent γ is a chosen input parameter. Various tests have shown that an optimal value for γ is especially dependent on the basis set size and varies for the most cases from

0.8 to 1.4. The value 1 is a good choice for most cases.

The effect of the variation of γ on the results is explored and discussed in the paper: "*Accurate calculations of intermolecular interaction energies using explicitly correlated wave functions*"[56] and is also is presented in my diploma thesis and will therefore not be discussed in detail here but in general it holds that larger values of γ are better for bigger basis sets. The reason for this is, that the description of the cusp is fairly reasonable *deeper into* the cusp the bigger the basis set becomes. For small basis sets the correction also accounts for the outer part of the cusp and that is facilitated by smaller values of γ.

The Slater type function in the correlation factor F_{12} (see **Equation (67)**) is approximated by Gaussians[49],[54] what is done due to the convenience of Gaussian type functions in computation of the necessary integrals.

$$F_{12} = -\frac{1}{\gamma} e^{-\gamma r_{12}} \approx \sum_{i=1}^{6} c_i e^{-\alpha_i (r_{12})^2} \tag{68}$$

The number of Gaussian functions used to build a linear combination to approximate the Slater type function from has been tested, and 6 has been found as being sufficient.

4.3.3 MP2-F12(FIX)

The MP2-F12 method has a large number of Ansätze and approximations[57] which have been tested for many systems for various kinds of properties.[1],[58]–[60] Some of the approximations are not orbital invariant and therefore localized orbitals must be used. Other approximations introduce geminal basis set superposition errors (GBSSE). All *diagonal-type* Ansätze like the Diagonal (D), Diagonal Fixed (FIX) or Block[61] (B) ones, avoid this GBSSE (see **Section 4.3.8**).[54],[62] For a detailed overview on most approximations, the papers [54], [55] are recommended.

In this work only Ansatz **3** has been used where the strong orthogonality projector $\widehat{Q}_{12}$ is:

$$\widehat{Q}_{12} \;=\; (1-\hat{o}_1)(1-\hat{o}_2)(1-\hat{v}_1\hat{v}_2) \tag{69}$$

with $\hat{o}_n = \sum_i |\phi_i(\mathbf{r}_n)\rangle\langle\phi_i(\mathbf{r}_n)|$ projecting onto the occupied space and
$\hat{v}_n = \sum_a |\phi_a(\mathbf{r}_n)\rangle\langle\phi_a(\mathbf{r}_n)|$ projecting onto the virtual space. $\hat{Q}_{12}$ keeps the explicitly correlated terms orthogonal to the conventional part of the wave function. In some of the

resulting terms the complete orthonormal basis is approximated using the union of the orbital basis and a complementary auxiliary (CA) basis set (CABS). Then summations over the complete virtual space are approximated as

$$|\alpha\rangle\langle\alpha| = 1 \approx \underbrace{\sum_{x} |x\rangle\langle x| + \sum_{a} |a\rangle\langle a|}_{\text{CABS approach}} \tag{70}$$

with CA orbitals $|x\rangle$ and normal orbitals $|a\rangle$. The CABS consists of basis functions constructed by orthogonalization of the RI basis to the MO basis. It has the same number of functions as the RI basis unless functions have to be deleted due to linear dependency.

By the construction of the projector $\widehat{Q}_{12}$, the amplitudes $\mathcal{T}_{rs}^{ij}$, $\mathcal{T}_{kx}^{ij}$, $\mathcal{T}_{xk}^{ij}$ are zero for any r, s, k, x, with r or s denoting any orbitals in the MO basis.

The inclusion of the $\mathcal{T}_{ab}^{ij}$ in explicitly correlated amplitudes would not change the numerical results since they are already present in the conventional double excitations. The contribution of both $\mathcal{T}_{ab}^{ij}$ would just be shared and this would lead to Ansatz **2**.[45] The first Ansatz for F12, Ansatz **1**[45] excludes excitations of the type Φ_{ij}^{ax} ($\mathcal{T}_{ax}^{ij} = 0$) and does not give satisfactory results[a].

The usage of Ansatz **3** ensures that the explicitly correlated terms are orthogonal to both Hartree-Fock and conventional MP2 configurations.[45], [46], [54], [63]–[65]

Diagonal-amplitudes (`D`) So far the, the geminal basis set superposition error (GBSSE) (see **Section 4.3.8**) is still present but can be removed by the introduction of Diagonal amplitudes:

$$T_{mn}^{ij} := 0 \text{ if } mn \neq ij \text{ or } mn \neq ji \tag{71}$$

Diagonal Fixed-amplitudes (`FIX`) As a further restriction to the Diagonal-amplitudes and therefore also to the GBSSE free method the Diagonal Fixed-amplitudes, according

[a] in the CABS approach

to Kato's cusp condition[47],[66] were introduced:

$$\begin{aligned} T_{ii}^{ii} &= \frac{1}{2} \\ T_{ij}^{ij} &= \frac{3}{8} \\ T_{ji}^{ij} &= \frac{1}{8} \end{aligned} \tag{72}$$

The remaining amplitudes are set to 0

With the use of Diagonal Fixed amplitudes the method becomes size consistent and orbital invariant and the use of localized orbitals (LMOs) is no longer a necessity. The amplitudes of the explicitly correlated part do not have to be calculated. The error introduced by this approximation depends on the system and the property calculated but is in general rather small. The author of this work recommends MP2-F12(`FIX`) for all explicitly correlated MP2 and CCSD(T) level calculations because canonical orbitals can be used and the error introduced by the Diagonal Fixed amplitudes is small. It is used for all MP2 level calculations in this work.

4.3.4 CCSD(T*)-F12a/b

As for MP2-F12, there is for CCSD(T)-F12 a set of approximations[54],[55],[58] which are not be discussed in this work.

In CCSD-F12a only linear terms in the amplitudes T_{kl}^{ij} are kept as well as those are treated at second-order perturbation theory level. As for CCSD-F12b more coupling terms between the conventional and explicitly correlated amplitudes are included. Those rather strong simplifications give are very good approximation to the CCSD basis set limit even if a double zeta basis is used. While CCSD-F12a shows the trend to overestimate the correlation energy, CCSD-F12b is systematically below the basis set limit.

4.3.5 Scaling for the (T) to match -F12 basis convergence

Explicit correlation has not been implemented in the perturbative triples correction (T) yet. Therefore a trick has to be used to speed-up the convergence in the (T) energy contribution with respect to the basis set size. The ratio of the MP2-F12 and MP2 correlation energies is used to estimate[58] a scaling factor ($\frac{\mathrm{E_{MP2-F12}}}{E_{\mathrm{MP2}}}$) for the perturbative

triples correction.

$$\Delta E_{(\mathrm{T}^*)} = \Delta E_{(\mathrm{T})} \cdot \frac{\mathrm{E}^{\mathrm{corr}}_{\mathrm{MP2-F12}}}{E^{corr}_{\mathrm{MP2}}} \tag{73}$$

This relies on the assumption that the triples energy would be affected by explicit correlation with the same ratio than the MP2 correlation contribution. Taking into account that the same assumption is made in the basis set extrapolation, where the whole correlation contribution is extrapolated in the same way, this appears to be valid. This is underlined by the good results of the CCSD(T*)-F12a being very close to the CCSD(T) basis set limit.

Since the scaling factor can be different for different molecules it is not ensured that calculated energy differences are size consistent. To ensure this property all triples contributions of all systems involved in the calculated energy difference have to be scaled with the same scaling factor. This is done the simplest way possible by taking the scaling factor of the biggest system. Initially this size consistent scaling has been denoted as (T*′). It gives better results than the original (T*) and after **Section 5.4.1** where the results are shown only the size consistent scaling will be used and is then renamed to (T*).

4.3.6 Resolution of the identities (RIs)

As already mentioned in **Section 4.3**, the basic idea of explicit correlation has been developed quite some time ago, it goes back as far as 1929 to Hylleraas.[43] However, routine applications only came within reach when resolution of the identity approximations were introduced which simplified the required many electron integrals:

$$\langle ij|r_{12}^{-1}\hat{o}_1\widehat{F}_{12}|kl\rangle \;=\; \sum_m \langle ijm|r_{12}^{-1}\hat{F}_{32}|mlk\rangle. \tag{74}$$

Kutzelnigg and Klopper[44], [45] introduced the RI by expansion of $\hat{o}_1$ to $\hat{o}_1\hat{p}'_2$:

$$\hat{o}_1 \rightarrow \hat{o}_1\hat{p}'_2 \tag{75}$$

$$\hat{p}'_2 \;= \sum_\alpha |\alpha(\mathbf{r}_2)\rangle\langle\alpha(\mathbf{r}_2)| \simeq 1 \tag{76}$$

In principle the RI is not an approximation but only as long as $|\alpha\rangle\langle\alpha|$ is a complete basis. That is of course impossible but a good choice of the RI basis assures the this is a very good approximation with negligible error.

This brings everything down to products of 2-electron integrals which are generally computed faster, speeding up the whole calculation.

$$\begin{aligned}\langle ij|r_{12}^{-1}\hat{o}_1\widehat{F}_{12}|kl\rangle &= \langle ij|r_{12}^{-1}\hat{o}_1\hat{p}_2'\widehat{F}_{12}|kl\rangle \\ &= \sum_{m\alpha}\langle ij|r_{12}^{-1}|m\alpha\rangle\langle m\alpha|\widehat{F}_{12}|kl\rangle. \end{aligned} \tag{77}$$

4.3.7 CABS singles

For conventional methods the basis set convergence of the correlation contribution is much slower than that of Hartree-Fock, and therefore the deviation of the HF energy to the basis limit is not the most significant contribution in the basis set incompleteness error of a calculation. For explicitly correlated methods the basis set convergence behavior is reversed. The main basis set error is now introduced by HF. This error can be reduced by the introduction of perturbative single excitations into the complementary auxiliary (CA) orbital space, which are determined by the following MP2-like equations:[55]

$$f_\alpha^i = \sum_k t_\alpha^k f_k^i - \sum_\beta f_\alpha^\beta t_\beta^i \tag{78}$$

$$\Delta E_s = 2\sum_{i\alpha} f_\alpha^i t_\alpha^i. \tag{79}$$

f_α^β are the matrix elements of the closed-shell Fock operator, t_α^β are the singles amplitudes, and the indices α, β run over both the virtual and the CA orbitals.

The excitations into the complementary auxiliary orbital space and the lowering of the energy can not be considered a correlation effect. The CABS singles take care of the basis set incompleteness of Hartree-Fock. The CABS energy correction ΔE_s is added to the HF energy.

4.3.8 Geminal BSSE

As seen in the **Section 3.5** certain excitations can cause a Basis Set Superposition Error. This effect can also be seen in the explicitly correlated part and is called Geminal Basis Set Superposition Error. The Geminal BSSE in the interaction energy cannot be corrected with the Counterpoise Correction. It can be avoided with the use of all *diagonal-type* approximations like Diagonal-amplitudes or Diagonal Fixed-amplitudes as shown in **Section 4.3.3**.

5 Performance of -F12 methods

In this **Section** I present the results of the calculations done in this work.
There are three main parts:

- Calculations done to investigate the improvement of the basis set error due to the usage of explicit correlated methods
- Development of Dispersion-Weighted-MP2-F12 (DW-MP2-F12)
- Calculations done to show the lowering of the intrinsic error due to newly developed method (DW-MP2-F12)

Before investigating the explicitly correlated methods, we first discuss the used reference values and the impact of Hartree-Fock contributions. This is important because in explicitly correlated methods the basis set limit is closely approached and thus high quality reference values are required. We go on by investigating the performance of the CCSD(T*)-F12 results. Based on those the intrinsic error of the more approximate MP2 method is tested and improvements due to spin component scaling are examined.

All calculation have been done using the `Molpro` quantum chemistry package.[67] Since the program developed during this work, various versions have been used. Most of the calculations however have been done using version 2010.2.

In all results for post Hartree-Fock methods with explicit correlation the CABS singles correction is included.

5.1 Training- and test-sets

The clusters used in this work are from the S22-,[2] S66-,[3] JSCH-2005-set[68] and some additional structures. The S22- and S66- sets are sets of 22 and 66 small and medium size weak interacting clusters. JSCH-2005 is a benchmark set of over 100 complexes with experimental and optimized structure of biomacromolecules like DNA, RNA and proteins.

5.2 MP2 vs. MP2-F12

The improvement of the basis set error due to the usage of explicit correlation in MP2 has already been presented in my diploma thesis and will not be shown here but the improvements are closely comparable to those in CCSD(T).

5.3 Quality of the CCSD(T)/CBS[3'4'] reference

In the Complete Basis Set estimate CBS[3'4'] AVTZ and AVQZ basis is used for all atoms except for the hydrogen atom, where VTZ and VQZ where utilized (see **Section 3.3**). CCSD(T)/CBS[3'4'] reference values where only accessible for a subset of the S22[2] set due to computational cost. Estimating the Complete Basis Set limit (CBS) via extrapolation (see **Section 4.1**) from triple and quadruple zeta basis sets is known to be a good reference but extrapolation from quadruple and quintuple zeta basis sets is preferred if available. The CBS[3'4'] reference has been tested against CBS[4'5'] for an even smaller subset of only 5 systems. Again due to computational cost, those where the only ones available. The CBS[4'5'] reference is known to be a very good reference. For this small subset the quality of the CCSD(T)/CBS[3'4'] reference has been found to be nearly identical to the CCSD(T)/CBS[4'5'] ones.

The plot shows the deviation of CCSD(T)/CBS[3'4'] from the CCSD(T)/CBS[4'5']:

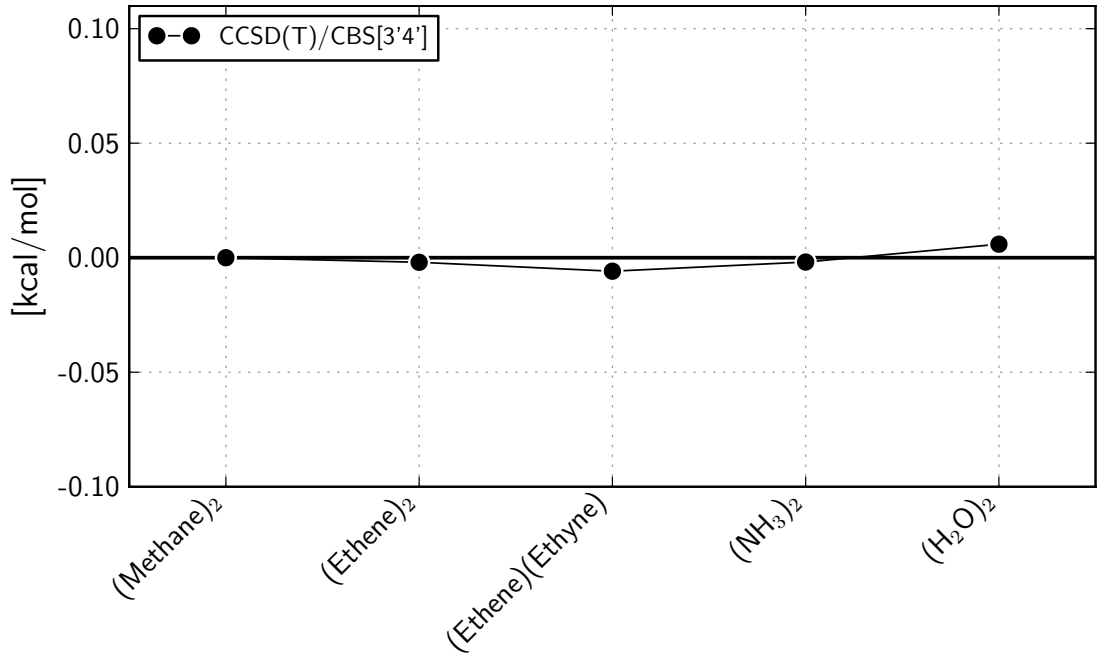


Figure 6: Deviation of the CBS[3'4'] reference to the CBS[4'5'] for a S22[2] subset

The values of the deviation of CCSD(T)/CBS[3'4'] from the CCSD(T)/CBS[4'5'] are summarized in this table:

Cluster	ΔE CCSD(T)/CBS[3'4']
$(Methane)_2$	0.000
$(Ethene)_2$	-0.002
(Ethene)(Ethyne)	-0.006
$(NH_3)_2$	-0.002
$(H_2O)_2$	0.006

Table VII: Quality of the CCSD(T)/CBS[3'4'] reference for the S22[2] subset.

The results confirm that the CBS[3'4'] is a reliable reference for the subset and it is hoped that this also holds for the other systems.

5.4 CABS singles correction

For conventional methods if a AVDZ basis is used the HF error is usually around 10% of the total error and therefore not of special importance. In -F12 calculations it amounts up to 90% of the total error (AVDZ basis) and therefore becomes important. Including the CABS singles correction (ΔE_s) (see **Section 4.3.7**) improves the HF results approximately to the next basis set cardinality.

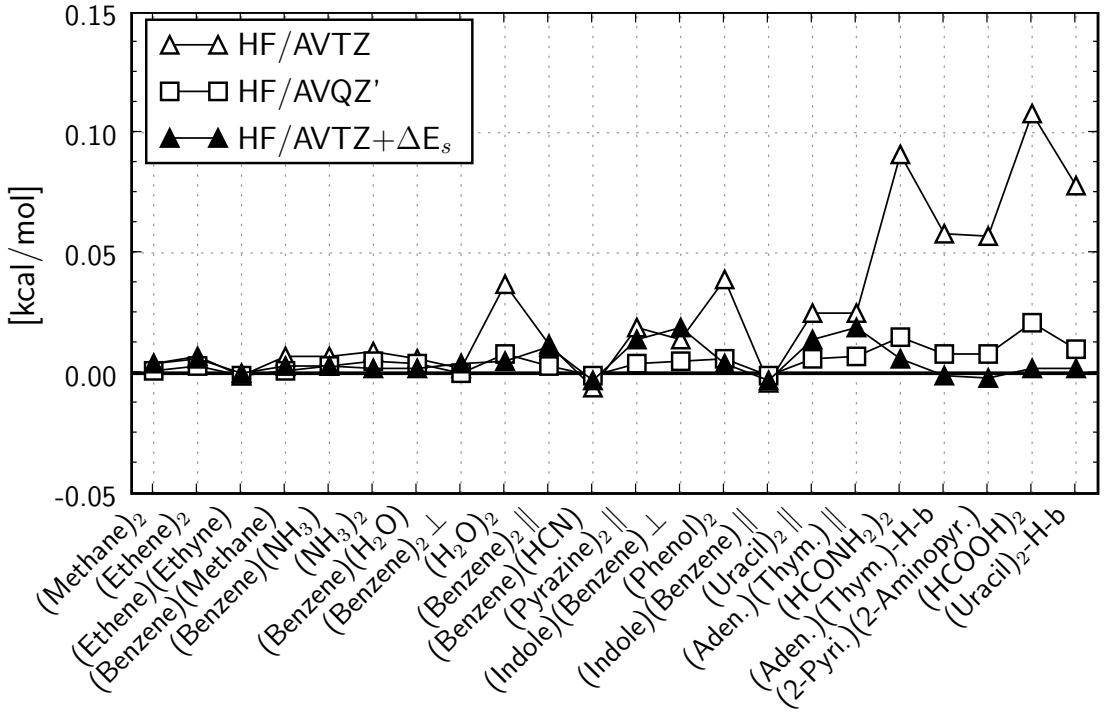


Figure 7: Basis set error for HF in the S22[2] set, reference HF/AV5Z

RMS, MAD and Max [kcal/mol]			
	AVTZ	AVQZ	AVTZ $+\Delta E_s$
MAD	0.028	0.006	0.006
RMS	0.041	0.007	0.008
Max	0.108	0.021	0.019

Table VIII: RMS MAD Max basis set error in HF (S22) (reference AV5Z)

5.4.1 Scaled (T)

Since explicit correlation has not been implemented in the perturbative triples correction (T) yet, the scaling described in **Section 4.3.5** is used to speed-up the convergence in the (T) energy with respect to the basis set size.

The graph shows only the (T) contribution of the CCSD(T) energy. The (T) denotes unscaled triples, in (T*) all systems a scaled independently and (T*′) denotes a common scaling factor for the dimer and the two monomers.

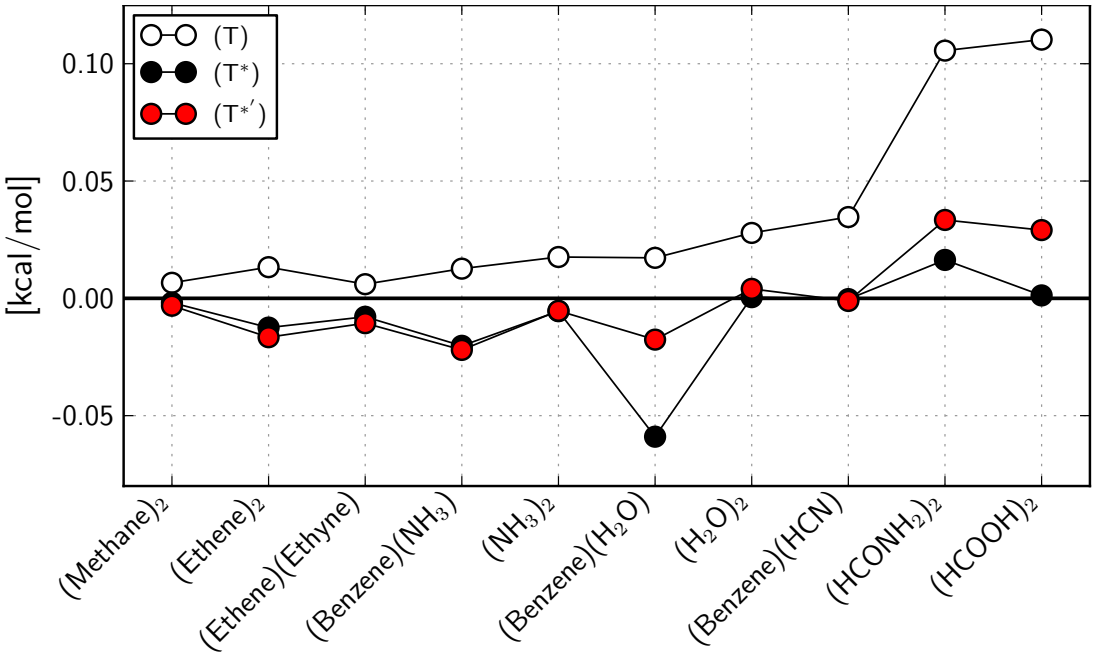


Figure 8: Basis set error in (T) for S22[2] subset, reference: (T)/CBS[3'4']

The following table summarizes the key statistics:

RMS, MAD and Max [kcal/mol]			
	(T)	(T*)	(T*′)
MAD	0.027	0.014	0.013
RMS	0.040	0.022	0.016
Max	0.106	-0.059	0.033

Table IX: RMS MAD Max basis set error in the triples (S22)

Since (T*′) gives better results than the original (T*) and is size consistent, (T*′) is now called (T*) and used exclusively.

5.5 CCSD(T) vs. CCSD(T*)-F12

The improvement of the convergence due to explicit correlation in the CCSD(T) method has been tested on subsets of the S22[2] and the S66[69] test set. The Complete Basis Set limit (CBS) estimated from triple and quadruple basis sets has been used as the reference.

5.5.1 S22

Figure 9 demonstrates the effects of explicit correlation in CCSD(T) with the approximation CCSD(T*)-F12a for a subset of the S22 set. The deviation from CCSD(T)/CBS[3'4'] is plotted.

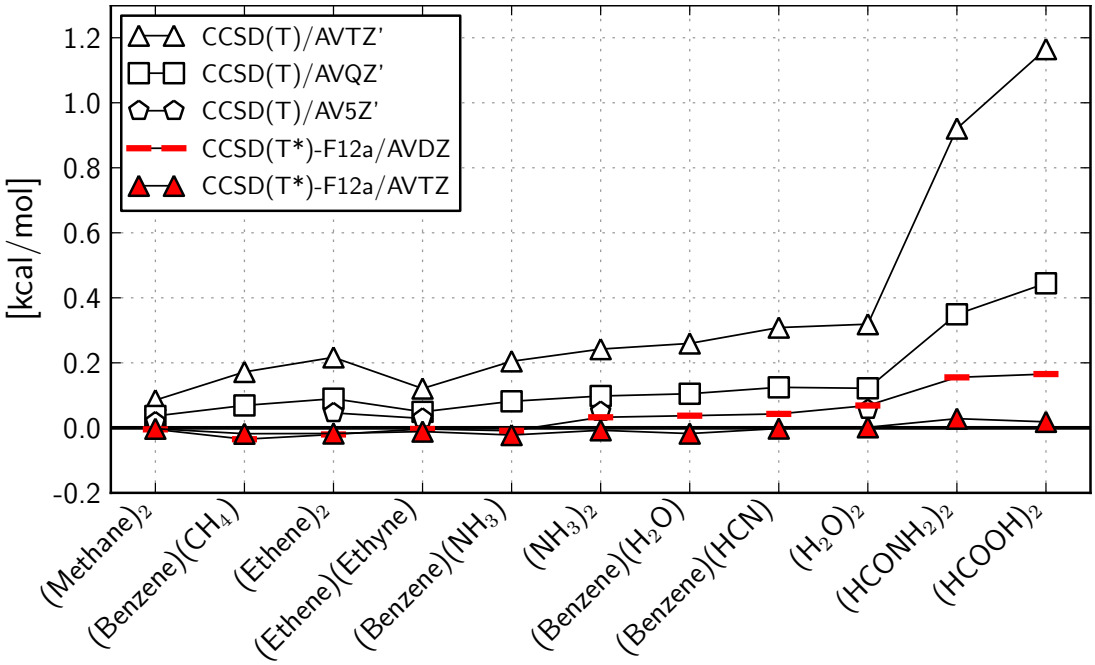


Figure 9: Basis set error for S22[2] subset, reference: CCSD(T)/CBS[3'4']

We see dramatic improvement of the interaction energies by the explicit correlation treatment.The **Figure** clearly demonstrates that the CCSD(T*)-F12a/AVDZ results are much more accurate than the conventional CCSD(T)/AVQZ' ones, and on the scale of the **Figure** the CCSD(T*)-F12a/AVTZ and CCSD(T)/CBS[3'4'] values are virtually identical.

Due to the size of the systems, CCSD(T)/CBS[34] estimates could only be obtained for 11 of the 22 dimers, and therefore the following discussion is restricted to this subset. As seen in **Table X**, the CCSD(T*)-F12a interaction energies are very close to the CCSD(T)/CBS[3'4'] reference values. Even for the small AVDZ basis, the mean absolute deviation (MAD) amounts to only 0.05 kcal/mol, and the maximum deviation (MAX) is 0.17 kcal/mol. With the AVTZ basis, the MAD (MAX) values are reduced to 0.014 (0.03) kcal/mol. According to previous experience,[58] it is not unlikely that the CCSD(T*)-F12a/AVTZ values are more accurate than the CCSD(T)/CBS[3'4'] ones.

The effects of explicit correlation in CCSD(T) for the two approximation CCSD(T*)-F12a CCSD(T*)-F12b in the S22[2] set are presented in **Table X**.

	CCSD(T)				CCSD(T*)-F12a	
Dimer	AVDZ'	AVTZ'	AVQZ'	CBS[3'4']	AVDZ	AVTZ
Hydrogen bonded complexes:						
$(NH_3)_2$	-2.420	-2.899	-3.044	-3.142	-3.109	-3.149
$(H_2O)_2$	-4.333	-4.673	-4.870	-4.992	-4.923	-4.990
Formic $acid_2$	-15.783	-17.629	-18.350	-18.794	-18.629	-18.776
$Formamide_2$	-13.811	-15.176	-15.747	-16.096	-15.941	-16.068
$Uracil_2$ (C_{2h})	—	—	—	(-20.65)	-20.630	-20.690
2-pyri. $\cdots$ 2-amino-py.	—	—	—	(-16.71)	-16.974	—
Adenine $\cdots$ Thymine	—	—	—	(-16.37)	-16.717	—
Complexes with predominant dispersion contribution:						
$(CH_4)_2$	-0.287	-0.443	-0.492	-0.528	-0.533	-0.531
$(C_2H_4)_2$	-0.824	-1.265	-1.392	-1.480	-1.502	-1.499
$C_6H_6 \cdots CH_4$	-0.955	-1.266	-1.369	-1.438	-1.473	-1.456
$(C_6H_6)_2$ $\parallel$	—	—	—	(-2.73)	-2.896	-2.703
$Pyrazine_2$	—	—	—	(-4.42)	-4.538	-4.299
Indole $\cdots$ C_6H_6 $\parallel$	—	—	—	(-5.22)	-4.914	—
$Uracil_2$ (C_2)	—	—	—	(-10.12)	-10.157	—
Adenine $\cdots$ Thymine $\parallel$	—	—	—	(-12.23)	-12.291	—
Mixed complexes:						
Ethene $\cdots$ Ethine	-1.154	-1.384	-1.455	-1.505	-1.508	-1.516
$C_6H_6 \cdots H_2O$	-2.624	-3.007	-3.162	-3.266	-3.229	-3.284
$C_6H_6 \cdots NH_3$	-1.771	-2.099	-2.222	-2.304	-2.313	-2.326
$C_6H_6 \cdots$ HCN	-3.650	-4.221	-4.405	-4.530	-4.487	-4.532
$(C_6H_6)_2$ $\perp$	—	—	—	(-2.74)	-2.783	-2.738
Indole $\cdots$ C_6H_6 $\perp$	—	—	—	(-5.73)	-5.721	—
$Phenol_2$	—	—	—	(-7.05)	-7.148	—

Table X: CCSD(T) and CCSD(T*')-F12 binding energies (in kcal/mol)for the S22 benchmark set from Ref. [2].

Values in parenthesis are estimated CBS limits from Ref. [2]. Those were obtained with smaller basis sets.

The results clearly demonstrate that the CCSD(T*)-F12a/AVDZ results are much more accurate than the conventional CCSD(T)/AVQZ' ones, CCSD(T*)-F12a/AVTZ values are virtually identical to the CCSD(T)/CBS[34] ones. As already mentioned, the CABS singles correction is very important to reach this excellent accuracy. Without it, the accurate correlation contributions would be spoiled by large errors of the HF contributions; the MAD (MAX) values for the CCSD(T*)-F12a/AVDZ values without the singles correction are 0.20 (0.854) kcal/mol. Furthermore, we note that the scaling correction of the triples energy reduces the MAD (MAX) values from 0.134 (0.380) to 0.05 (0.17) kcal/mol (basis AVDZ, including CABS singles correction).

5.5.2 S66

The S66 set is described by Pavel Hobza *et al.* in their paper[69] as following: "The (S66) data set consists of 66 complexes formed by combining 14 monomers in various configurations. The monomers were chosen so that they represent the motifs and functional groups most commonly found in bio-molecules. (...) Only complexes with interactions stronger than approximately 1.5 kcal/mol were included in the set"

The effects of explicit correlation in CCSD(T) with the approximation CCSD(T*)-F12a for a subset of 21 clusters of the S66[69] set is investigated. The subset consist of all systems that could be calculated with CCSD(T)/AVQZ' basis at that time and computational power available. The deviation from the CCSD(T)/CBS[3'4'] is shown.

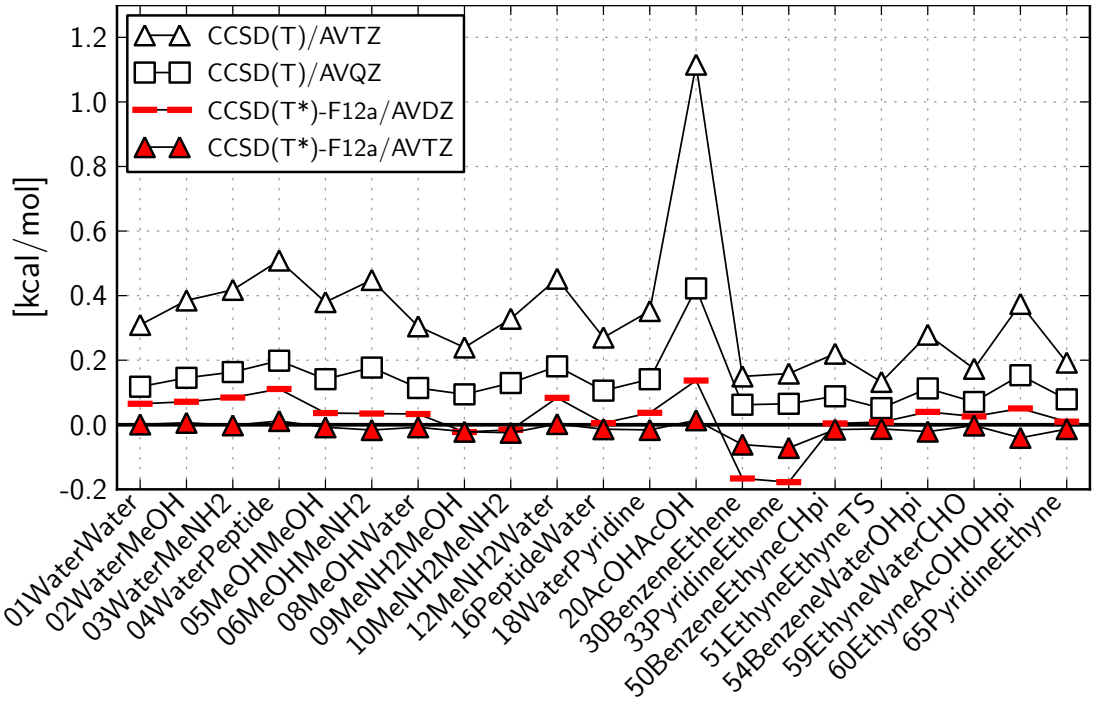


Figure 10: Basis set error for S66 subset, reference: CCSD(T)/CBS[3'4']

The statistical data for the S66 subset is presented in tabulated form:

RMS, MAD and Max [kcal/mol]

Method	Basis	RMS	MAD	Max
CCSD(T)	AVTZ	0.398	0.343	1.115
CCSD(T)	AVQZ	0.154	0.134	0.422
CCSD(T*)-F12a	AVDZ	0.077	0.058	-0.178
CCSD(T*)-F12a	AVTZ	0.026	0.018	-0.072
CCSD(T*)-F12b	AVDZ	0.140	0.118	0.344
CCSD(T*)-F12b	AVTZ	0.038	0.029	-0.024

Table XI: RMS MAD Max basis set error for the S66

For the S66 subset the CCSD(T*)-F12 improves over CCSD(T) in basis set convergence by more than two basis set cardinalities. For the triple zeta basis the maximal error drops from 1.115 kcal/mol -0.024 to kcal/mol (F12b).

All available CCSD(T), CCSD(T*)-F12a and CCSD(T*)-F12b for the S66[3] set are summarized in the **Tables XII, XIII and XIV**.

	CCSD(T)			CCSD(T*)-F12a		CCSD(T*)-F12b	
Dimer	AVTZ'	AVQZ'	CBS[3'4']	AVDZ	AVTZ	AVDZ	AVTZ
Hydrogen bonded complexes:							
$_{01}(H_2O)_2$	-4.680	-4.871	-4.989	-4.924	-4.988	-4.864	-4.958
$_{02}H_2O\cdots MeOH$	-5.298	-5.538	-5.683	-5.612	-5.677	-5.533	-5.639
$_{03}H_2O\cdots MeNH_2$	-6.596	-6.851	-7.014	-6.930	-7.016	-6.838	-6.970
$_{04}H_2O\cdots$ Peptide	-7.704	-8.014	-8.213	-8.102	-8.201	-7.986	-8.148
$_{05}MeOH_2$	-5.459	-5.698	-5.839	-5.803	-5.847	-5.719	-5.809
$_{06}MeOH\cdots MeNH_2$	-7.201	-7.473	-7.650	-7.615	-7.666	-7.504	-7.616
$_{07}MeOH\cdots$ Peptide	—	—	(-8.230)	-8.285	-8.349	-8.162	-8.296
$_{08}MeOH\cdots H_2O$	-4.768	-4.959	-5.073	-5.040	-5.081	-4.974	-5.052
$_{09}MeNH_2\cdots MeOH$	-2.851	-2.996	-3.091	-3.113	-3.112	-3.047	-3.085
$_{10}(MeNH_2)_2$	-3.874	-4.073	-4.202	-4.218	-4.227	-4.124	-4.187
$_{11}MeNH_2\cdots$ Peptide	—	—	(-5.419)	-5.509	-5.500	-5.378	-5.446
$_{12}MeNH_2\cdots H_2O$	-6.929	-7.200	-7.381	-7.298	-7.379	-7.195	-7.328
$_{13}$Peptide$\cdots$MeOH	—	—	(-6.187)	-6.264	-6.284	-6.171	-6.243
$_{14}$Peptide$\cdots MeNH_2$	—	—	(-7.454)	-7.559	-7.568	-7.452	-7.519
$_{15}Peptide_2$	—	—	(-8.630)	-8.736	-8.742	-8.609	-8.688
$_{16}$Peptide$\cdots H_2O$	-4.913	-5.078	-5.183	-5.178	-5.197	-5.115	-5.169
$_{17}Uracil_2BP$	—	—	(-17.182)	-17.451	—	-17.274	—
$_{18}H_2O\cdots$ Pyridine	-6.597	-6.808	-6.949	-6.912	-6.965	-6.806	-6.916
$_{19}MeOH\cdots$ Pyridine	—	—	(-7.410)	-7.505	-7.521	-7.386	-7.470
$_{20}(AcOH)_2$	-18.306	-18.999	-19.421	-19.284	-19.407	-19.077	-19.311
$_{21}(AcNH_2)_2$	—	—	(-16.265)	-16.399	-16.504	-16.250	-16.430
$_{22}$AcOH$\cdots$Uracil	—	—	(-19.491)	-19.697	-19.787	-19.506	-19.699
$_{23}$AcNH2$\cdots$Uracil	—	—	(-19.189)	-19.367	-19.457	-19.203	-19.378

Table XII: CCSD(T) and CCSD(T*)-F12 binding energies (in kcal/mol) for the S66[3] benchmark set of Ref. [3].

Values in parenthesis are estimated CBS limits from Ref. [3]. Those were obtained with smaller basis sets.

	CCSD(T)			CCSD(T*)-F12a		CCSD(T*)-F12b	
Dimer	AVTZ'	AVQZ'	CBS[3'4']	AVDZ	AVTZ	AVDZ	AVTZ
Complexes with predominant dispersion contribution:							
${}_{24}(C_6H_6)_2$ ‖	—	—	(-2.822)	-2.988	-2.803	-2.771	-2.728
${}_{25}(\text{Pyridine})_2$ ‖	—	—	(-3.895)	-4.077	-3.884	-3.842	-3.799
${}_{26}(\text{Uracil})_2$	—	—	(-9.829)	-10.192	—	-9.851	—
${}_{27}C_6H_6\cdots\text{Pyridine}$ ‖	—	—	(-3.439)	-3.614	-3.426	-3.386	-3.345
${}_{28}C_6H_6\cdots\text{Uracil}$ ‖	—	—	(-5.713)	-5.947	—	-5.637	—
${}_{29}\text{Pyridine}\cdots\text{Uracil}$ ‖	—	—	(-6.819)	-7.071	—	-6.762	—
${}_{30}C_6H_6\cdots\text{Ethene}$	-1.197	-1.285	-1.347	-1.513	-1.408	-1.362	-1.357
${}_{31}\text{Uracil}\cdots\text{Ethene}$	—	—	(-3.380)	-3.478	-3.388	-3.314	-3.327
${}_{32}\text{Uracil}\cdots\text{Ethyne}$	—	—	(-3.738)	-3.814	-3.746	-3.650	-3.685
${}_{33}\text{Pyridine}\cdots\text{Ethene}$	-1.628	-1.721	-1.786	-1.964	-1.858	-1.810	-1.804
${}_{34}(\text{Pent.})_2$	—	—	(-3.776)	-3.904	—	-3.748	—
${}_{35}\text{Neopent.}\cdots\text{Pent.}$	—	—	(-2.613)	-2.700	—	-2.592	—
${}_{36}\text{Neopent.}_2$	—	—	(-1.777)	-1.847	—	-1.771	—
${}_{37}\text{Cyclopent.}\cdots\text{Neopent.}$	—	—	(-2.404)	-2.497	—	-2.393	—
${}_{38}(\text{Cyclopent.})_2$	—	—	(-2.997)	-3.119	—	-2.995	—
${}_{39}C_6H_6\cdots\text{Cyclopent.}$	—	—	(-3.575)	-3.703	—	-3.529	—
${}_{40}C_6H_6\cdots\text{Neopent.}$	—	—	(-2.895)	-2.988	—	-2.850	—
${}_{41}\text{Uracil}\cdots\text{Pent.}$	—	—	(-4.848)	-5.053	—	-4.836	—
${}_{42}\text{Uracil}\cdots\text{Cyclopent.}$	—	—	(-4.138)	-4.325	—	-4.132	—
${}_{43}\text{Uracil}\cdots\text{Neopent.}$	—	—	(-3.712)	-3.863	—	-3.710	—
${}_{44}\text{Ethene}\cdots\text{Pent.}$	—	—	(-2.005)	-2.056	-2.029	-1.960	-1.994
${}_{45}\text{Ethyne}\cdots\text{Pent.}$	—	—	(-1.748)	-1.780	-1.756	-1.677	-1.719
${}_{46}\text{Peptide}\cdots\text{Pent.}$	—	—	(-4.264)	-4.406	—	-4.226	—

Table XIII: CCSD(T) and CCSD(T*)-F12 binding energies (in kcal/mol) for the S66[3] benchmark set of Ref. [3].

Values in parenthesis are estimated CBS limits from Ref. [3]. Those were obtained with smaller basis sets.

	CCSD(T)			CCSD(T*)-F12a		CCSD(T*)-F12b	
Dimer	AVTZ'	AVQZ'	CBS[3'4']	AVDZ	AVTZ	AVDZ	AVTZ
Mixed complexes:							
$_{47}(C_6H_6)_2$ $\perp$	—	—	(-2.876)	-2.934	-2.873	-2.809	-2.824
$_{48}$Pyridine$_2$ $\perp$	—	—	(-3.535)	-3.608	-3.544	-3.479	-3.492
$_{49}C_6H_6\cdots$Pyridine $\perp$	—	—	(-3.331)	-3.386	-3.332	-3.257	-3.281
$_{50}C_6H_6\cdots$Ethyne $^{(CHpi)}$	-2.631	-2.763	-2.851	-2.847	-2.866	-2.764	-2.828
$_{51}$Ethyne$_2$ $\perp$	-1.402	-1.484	-1.535	-1.527	-1.548	-1.486	-1.529
$_{52}C_6H_6\cdots$AcOH $^{(OHpi)}$	—	—	(-4.707)	-4.749	-4.757	-4.605	-4.696
$_{53}C_6H_6\cdots$AcNH$_2$ $^{(NHpi)}$	—	—	(-4.361)	-4.420	-4.431	-4.314	-4.384
$_{54}C_6H_6\cdots$H$_2$O $^{(OHpi)}$	-2.998	-3.164	-3.277	-3.237	-3.298	-3.135	-3.254
$_{55}C_6H_6\cdots$MeOH $^{(OHpi)}$	—	—	(-4.188)	-4.198	-4.204	-4.048	-4.144
$_{56}C_6H_6\cdots$MeNH$_2$ $^{(NHpi)}$	—	—	(-3.231)	-3.260	-3.234	-3.125	-3.182
$_{57}C_6H_6\cdots$Peptide $^{(NHpi)}$	—	—	(-5.282)	-5.368	-5.311	-5.195	-5.243
$_{58}$(Pyridine)$_2$ $^{(CHN)}$	—	—	(-4.146)	-4.272	-4.244	-4.186	-4.202
$_{59}$Ethyne$\cdots$H$_2$O $^{(CHO)}$	-2.740	-2.843	-2.914	-2.888	-2.916	-2.867	-2.902
$_{60}$Ethyne$\cdots$AcOH $^{(OHpi)}$	-4.568	-4.789	-4.943	-4.892	-4.983	-4.788	-4.934
$_{61}$Pentane$\cdots$AcOH	—	—	(-2.912)	-3.028	-2.964	-2.893	-2.914
$_{62}$Pentane$\cdots$AcNH$_2$	—	—	(-3.534)	-3.669	-3.593	-3.512	-3.535
$_{63}$Benzene$\cdots$AcOH	—	—	(-3.801)	-3.851	-3.800	-3.680	-3.733
$_{64}$Peptide$\cdots$Ethene	—	—	(-2.999)	-3.068	-3.032	-2.955	-2.987
$_{65}$Pyridine$\cdots$Ethyne	-3.884	-3.998	-4.077	-4.067	-4.090	-4.031	-4.067
$_{66}$MeNH$_2\cdots$Pyridine	—	—	(-3.968)	-4.052	-4.004	-3.911	-3.950

Table XIV: CCSD(T) and CCSD(T*)-F12 binding energies (in kcal/mol) for the S66[3] benchmark set of Ref. [3].

Values in parenthesis are estimated CBS limits from Ref. [3]. Those were obtained with smaller basis sets.

5.5.3 F12a vs. F12b

For CCSD(T)-F12 the two approximations F12a and F12b are plotted versus the reference CBS[3'4']. For both cases (AVDZ and AVTZ basis) F12a does give the better results but overshoots in some cases, while F12b does not show a strong the tendency to overshoot.

Please take into account the small scale for the plots:

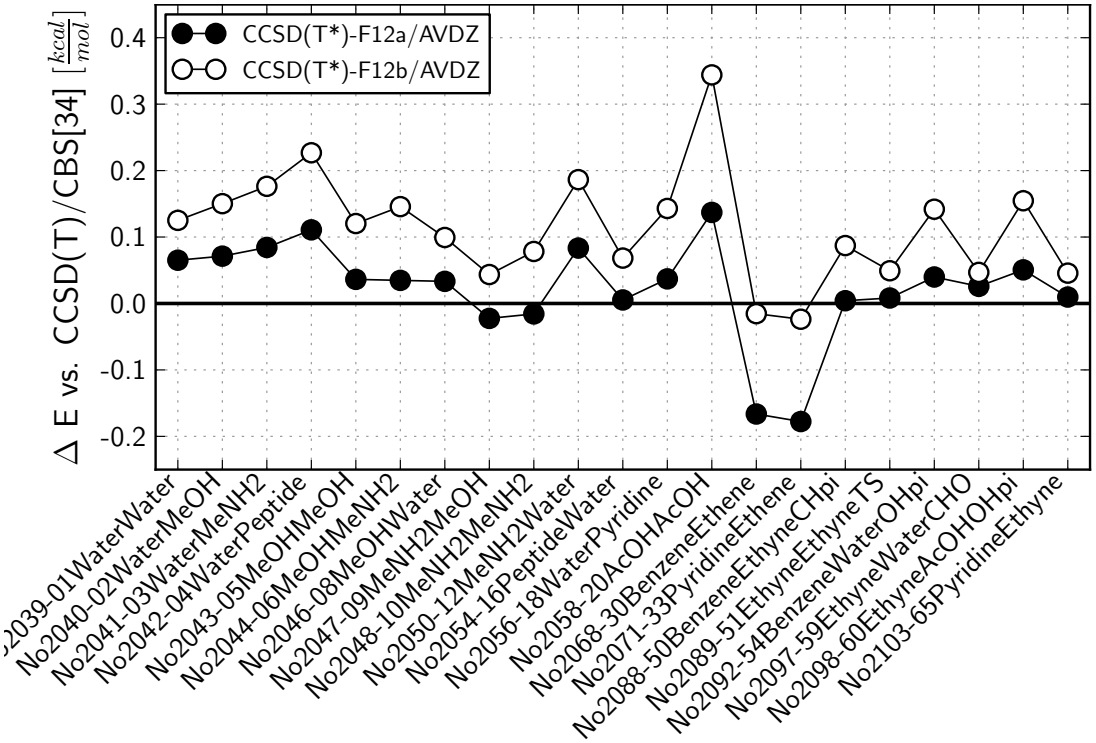


Figure 11: F12a vs. F12b with AVDZ basis, reference: CCSD(T)/CBS[3'4']

RMS, MAD and Max [kcal/mol]		
AVDZ	CCSD(T*)-F12a	CCSD(T*)-F12b
Max	0.178	0.344
MAD	0.058	0.118
RMS	0.077	0.140

Table XV: RMS MAD Max basis set error for the S66 subset

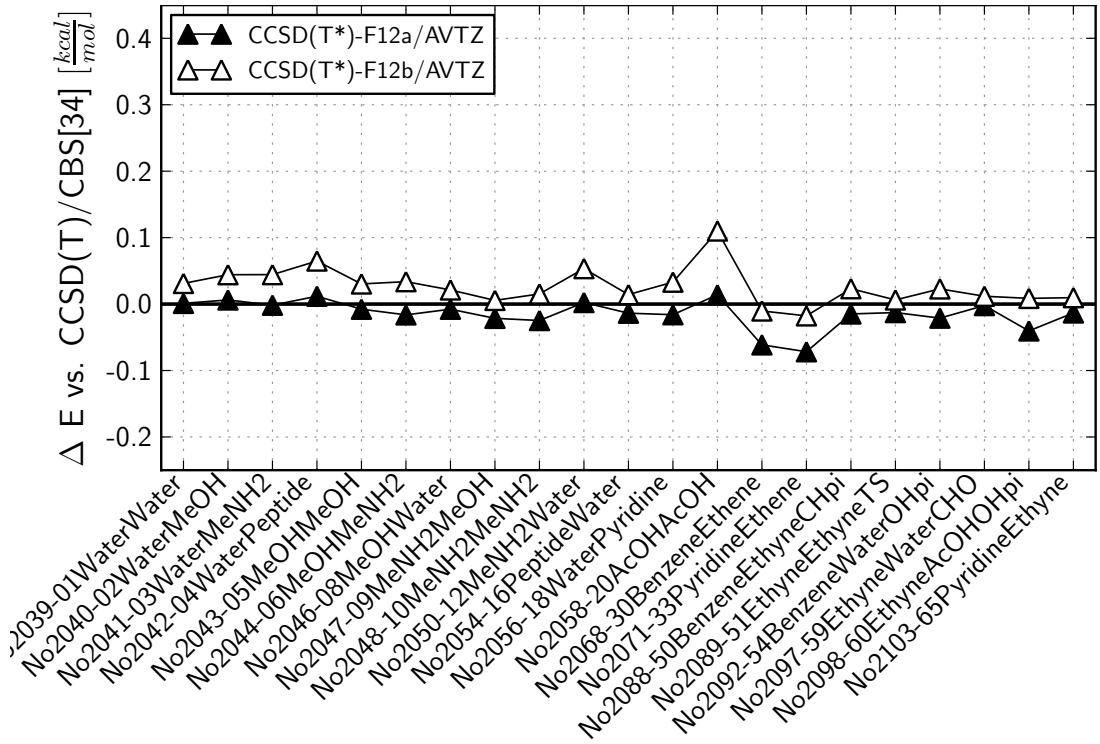


Figure 12: F12a vs. F12b with AVTZ basis, reference: CCSD(T)/CBS[3'4']

RMS, MAD and Max [kcal/mol]		
AVTZ	CCSD(T*)-F12a	CCSD(T*)-F12b
Max	0.072	0.110
MAD	0.018	0.029
RMS	0.026	0.038

Table XVI: RMS MAD Max basis set error for the S66 subset

6 Development of DW-MP2-F12

The CCSD(T)-F12 method is the best correlation method available, suitable for systems of a size of up to 35[b] first and second row atoms. It is known to cover most of the correlation energy and is therefore used as a reference. In combination with explicit correlation it is already close to the basis set limit with AVDZ basis (see **Figure 9** and **10**).

6.1 Deviation of MP2-F12 from CCSD(T*)-F12

Since the basis set used in all methods is the same and qualitatively different behavior for the convergence is not expected, **Figure 13** shows only the error of MP2-F12 relative to the CCSD(T*)-F12 reference. The usage of explicit correlation in both methods[56], [70] just improves the basis set convergence and has no strong influence on the deviation or should be at least not be noticeable on this scale.

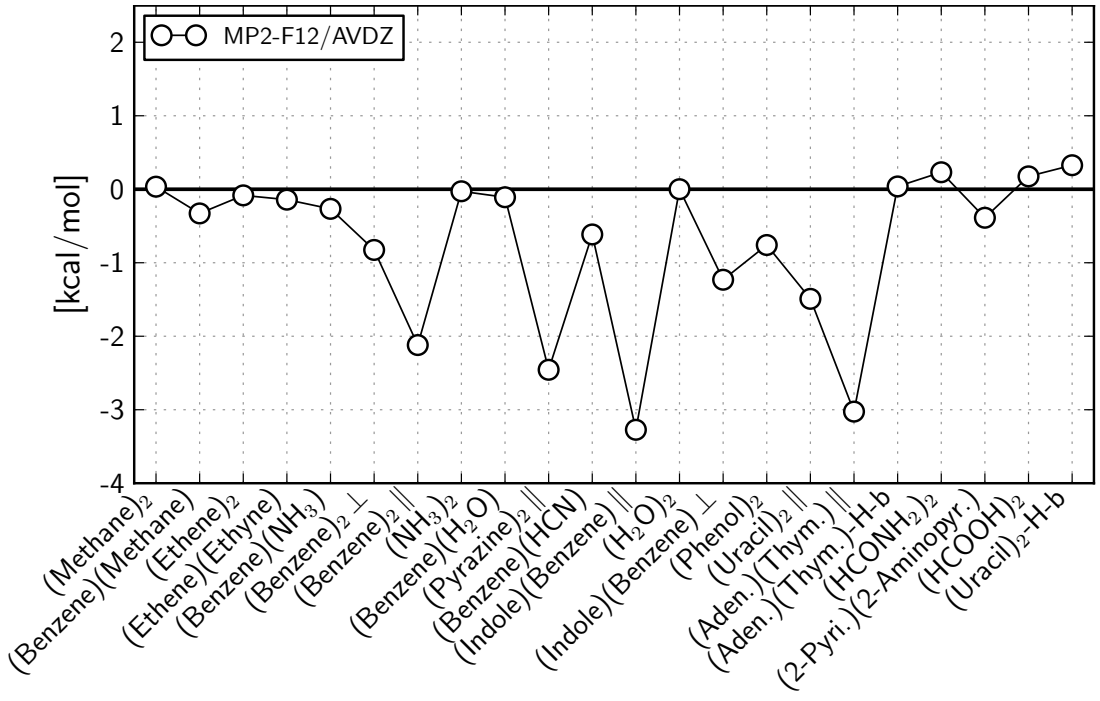


Figure 13: Deviation of MP2-F12 from CCSD(T*)-F12/same basis

Figure 13 confirms the well known fact that MP2 has a good error compensation for the description of clusters with hydrogen bonds, while the interaction energies of clusters in which the interaction of instantaneous multipoles is dominant are overestimated.

[b] depending on the number of hydrogen atoms and the computational power available

6.2 Improvement through spin component scaling

The usage of Spin Component Scaled MP2 (see **Section 3.9.2**) gives a good improvement for clusters in which the interaction of instantaneous multipoles is dominant but increases the error for clusters with hydrogen bonds.

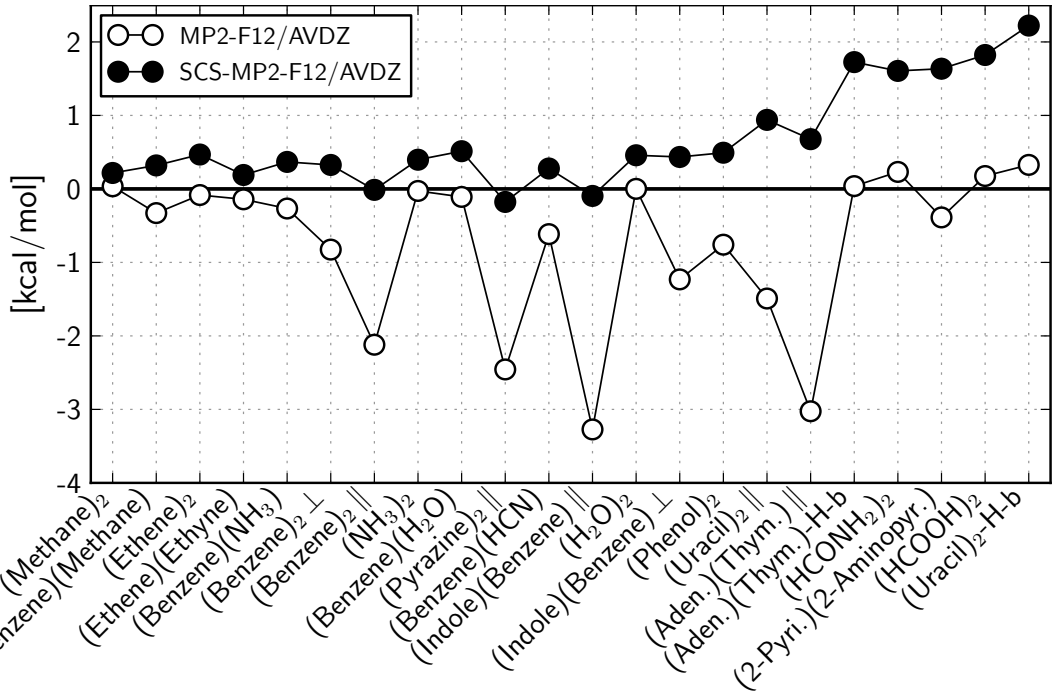


Figure 14: Deviation of MP2-F12 and SCS-MP2-F12 from CCSD(T*)-F12/same basis

This leads to the conclusion that one should use MP2 for the description of clusters with dominant hydrogen bonds and SCS-MP2 for clusters where London dispersion forces are dominating the interaction. This can be done by the user via selection of the method for each system. Such a manual selection goes along with the possible errors caused by the user due to misjudgement of the interaction type. A better solution for this problem is an automatic selection where MP2 and SCS-MP2 are smoothly combined.

6.3 Correlation between the HF/MP2 and the MP2/CCSD(T) behavior

As mentioned above it is well known that for London dispersion forces MP2 overshoots and SCS-MP2 gives good results while for hydrogen bonds SCS-MP2 undershoots and MP2 is very accurate. A very similar behavior is was found examining the relation between

MP2 and HF. The HF method describes interaction energies of hydrogen bonds at least qualitatively correct while for London dispersion force dominated interaction energies it gives a repulsive interaction energy.

This analogy in the behavior led the conclusion that the ratio of the HF energy contribution and the MP2 interaction energy could be used to develop a method that automatically switches between MP2 and SCS-MP2 for each cluster individually.

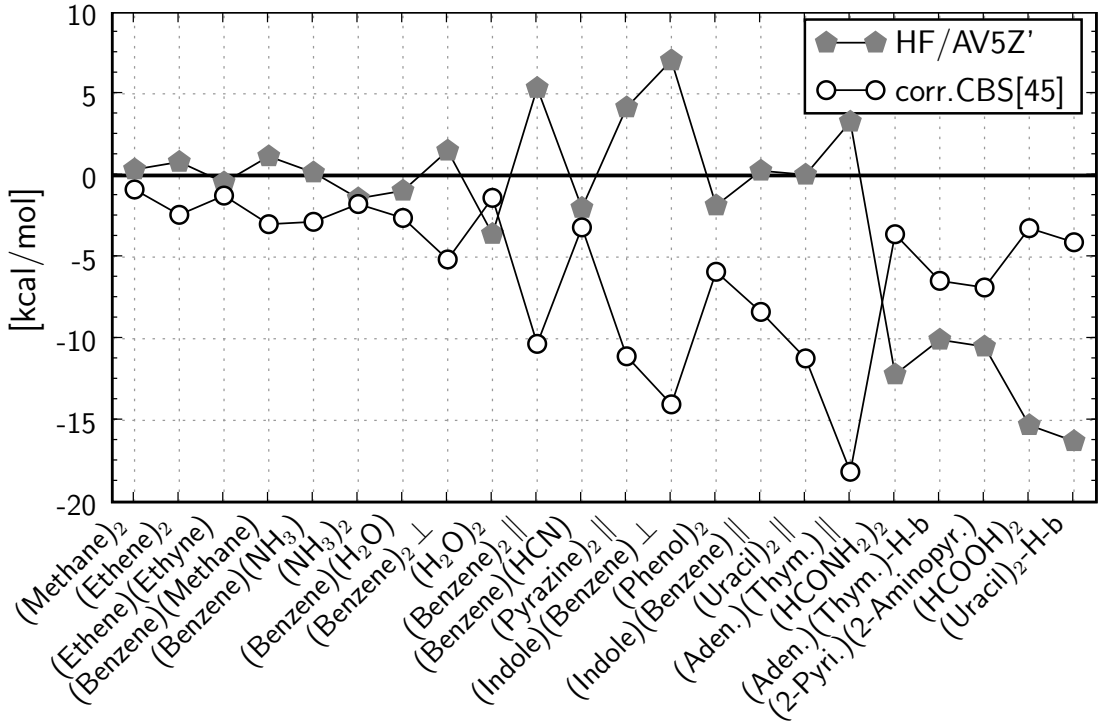


Figure 15: HF and MP2 correlation contribution on the interaction energy

The case of the Benzene dimer in the parallel orientation (||) is a good example for a London dispersion force dominated interaction of a molecular cluster and the interaction energy at the HF level is repulsive. The $(HCOOH)_2$ cluster is clearly dominated by hydrogen bonds and HF gives already the main contribution to the interaction energy. In the $(Uracil)_2$ in the parallel orientation (||) the interaction energy at HF level is ≈ 0 due to the canceling out of the different contributions.

6.4 Cutting function of DW-MP2-F12

From the analogy between the HF - MP2 relation and the MP2 - CCSD(T) relation a way of switching smoothly between MP2 and SCS-MP2 has been developed using a

tanh() type function. The energy on DW-MP2-F12 level for a system M ($\Delta E(M)$) is the HF contribution for the interaction energy $\Delta E_{HF_{+s}}(M)$ (including the MP2 singles) plus the correlation contribution. That correlation contribution is a mixture of MP2 and SCS-MP2:

$$\Delta E_{\text{DW-MP2}} = x \cdot \Delta E_{\text{MP2}} + (1 - x) \cdot \Delta E_{\text{SCS-MP2}} \tag{80}$$

That assures that the correlation contribution is always between MP2-F12 and SCS-MP2-F12 and not go beyond ($0 \leq x \leq 1$). This is ensured by:

$$x = \frac{1}{2} - \frac{1}{2} \tanh(g + v \cdot d) \tag{81}$$

The ratio of the interaction energy at HF and MP2 level is resembled by d:

$$d = \frac{\Delta E_{HF_{+s}}}{\Delta E_{\text{MP2-F12}}} \tag{82}$$

With this the DW-MP2-F12 cutting function takes the shape:

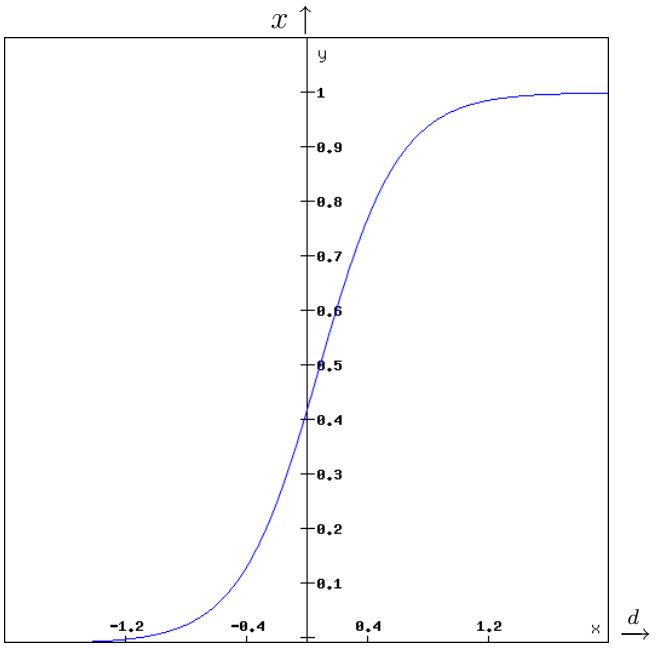


Figure 16: DW-MP2-F12 cutting function (x)

The DW-MP2-F12 energy for a system M split into singlet and triplet contributions:

$$\begin{aligned}\Delta E_{\text{DW-MP2}} &= \Delta E_{HF_{+s}} \\ &+ x \cdot \underbrace{[a \cdot \Delta E_{\text{MP2}_{\text{sing}}} + b \cdot \Delta E_{\text{MP2}_{\text{trip}}}]}_{\text{SCS}-\text{MP2}_{\text{ab}}^{\text{corr}}} \\ &+ (1-x) \cdot \underbrace{[\alpha \cdot \Delta E_{\text{MP2}_{\text{sing}}} + \beta \cdot \Delta E_{\text{MP2}_{\text{trip}}}]}_{\text{SCS}-\text{MP2}_{\alpha\beta}^{\text{corr}}} \end{aligned} \tag{83}$$

The SCS - factors a, b, α, β have predefined values and have not been optimized. For MP2 this is: $a = b = 1$ and for SCS-MP2 the values introduced by Stefan Grimme *et al.*[5] have been used: $\alpha = 1.2; \beta = 0.6222$

The parameters g and v (see **Equation (81)**) have been optimized on the S22 training set using an program written for this task in order to create DW-MP2-F12 $\equiv$ DW(2)-MP2-F12. For this they are $g = 0.15276$ and $v = -1.89952$. The number in parentheses indicates the number of parameters optimized. The values for g and v define the value for x where DW-MP2-F12 is a mixing of MP2 and SCS-MP2 half and half and the slope for the decrease of the mixing leading towards pure MP2 or SCS-MP2.

6.5 Effect of DW-MP2

The DW-MP2-F12 method can be applied to already finished MP2 calculations if the singlet and triplet contributions are stored separately. This procedure does not introduce any additional costs. The DW-MP2-F12 method gives values close to MP2 for systems that are already described qualitatively at the HF level and values close to SCS-MP2 for clusters that can not be characterized by HF at all. For systems that have a interaction energy close to 0 on HF level DW-MP2-F12 yields a mixture of MP2 and SCS-MP2.

7 Intrinsic error of DW-MP2-F12[1]

DW-MP2-F12 has been tested on S22,[2] the set it has been optimized on and 3 more (sub-)sets:

- S22
- T21 (own structures)[c]
- JSCH-2005[68] (subset)
- S66[3]

Due to the fact that the AVDZ basis gives already results close to the basis set limit and for larger basis sets some of the CCSD(T) calculations are not feasible, all calculations, CCSD(T*)-F12, MP2-F12, SCS-MP2-F12 and DW-MP2-F12 have been carried out using an AVDZ basis. The CCSD(T*)-F12 values for the same basis set are used as reference.

7.1 S22[2] training set

Since the DW-MP2-F12 has been optimized on the S22 training set, good performance on the very same set is expected. The number of parameters optimized in the DW-MP2-F12 is small with 2 for a set of 22 systems and nevertheless the improvement over MP2 and SCS-MP2 is significant.

[c] including geometries form Vogiatzis and Klopper, private communication

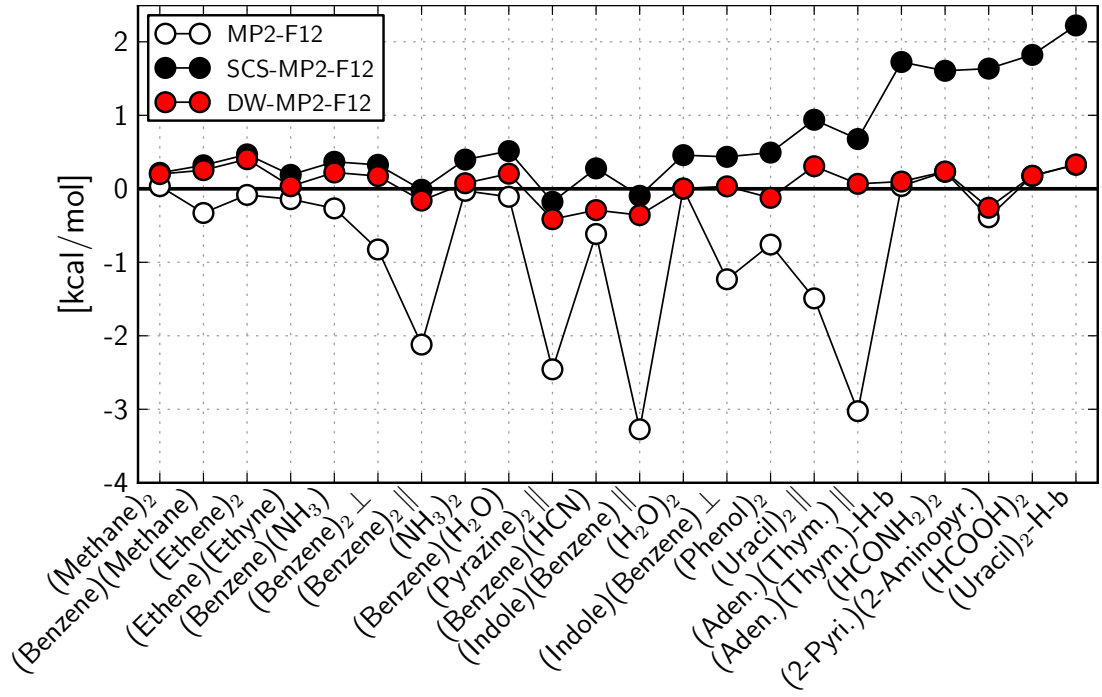


Figure 17: Intrinsic error of DW-MP2-F12 - S22 (training set), reference: CCSD(T*)-F12a/same basis (AVDZ)

By the usage of DW-MP2-F12 the RMS has decreased from 1.29 kcal/mol (MP2-F12) and 0.95 kcal/mol (SCS-MP2-F12) to 0.23 kcal/mol. The MAD went down from from 0.82 kcal/mol (MP2-F12) and 0.70 kcal/mol (SCS-MP2-F12) to 0.20 kcal/mol and for the largest error lowered from -3.27 kcal/mol (MP2-F12) and 2.22 kcal/mol (SCS-MP2-F12) to -0.41 kcal/mol.

RMS, MAD and Max [kcal/mol]			
Method	RMS	MAD	Max
MP2	1.29	0.82	-3.27
SCS-MP2	0.95	0.70	2.22
DW-MP2-F12	0.23	0.20	-0.41

Table XVII: RMS MAD Max intrinsic error (S22)

-F12/AVDZ	CCSD(T*)	HF	MP2	SCS-MP2	DW-MP2
Hydrogen bonded cluster:					
$(NH_3)_2$	-3.109	-1.397	-3.137	-2.712	-3.086
$(H_2O)_2$	-4.923	-3.585	-4.929	-4.472	-4.908
(Formic acid)$_2$	-18.629	-15.294	-18.510	-16.863	-18.459
(Formamide)$_2$	-15.941	-12.189	-15.745	-14.371	-15.694
Uracil$_2$ (C_{2h})	-20.630	-16.276	-20.341	-18.445	-20.276
2-pyri. $\cdots$ 2-amino-py.	-16.974	-10.503	-17.365	-15.341	-17.225
Adenine $\cdots$ Thymine	-16.717	-10.080	-16.585	-14.687	-16.456
cluster with predominant dispersion contribution:					
$(CH_4)_2$	-0.533	0.373	-0.495	-0.313	-0.326
$(C_2H_4)_2$	-1.502	0.835	-1.578	-1.027	-1.111
Benzene $\cdots$ CH_4	-1.473	1.171	-1.781	-1.132	-1.197
(Benzene)$_2$ $\parallel$	-2.896	5.375	-4.918	-2.811	-2.855
(Pyrazine)$_2$	-4.538	4.178	-6.894	-4.615	-4.887
Indole $\cdots$ Benzene $\parallel$	-4.914	7.066	-8.036	-4.859	-5.004
Uracil$_2$ (C_2)	-10.157	0.046	-11.119	-8.327	-9.924
Adenine $\cdots$ Thymine $\parallel$	-12.291	3.309	-14.830	-10.634	-12.177
Mixed cluster:					
Ethene $\cdots$ Ethine	-1.508	-0.434	-1.646	-1.314	-1.575
Benzene $\cdots$ H_2O	-3.229	-0.947	-3.453	-2.831	-3.324
Benzene $\cdots$ NH_3	-2.313	0.177	-2.600	-1.966	-2.290
Benzene $\cdots$ HCN	-4.487	-2.005	-5.099	-4.206	-4.972
(Benzene)$_2$ $\perp$	-2.783	1.515	-3.568	-2.416	-2.661
Indole $\cdots$ Benzene $\perp$	-5.721	0.285	-6.900	-5.234	-6.128
(Phenol)$_2$	-7.148	-1.845	-7.702	-6.272	-7.375

Table XVIII: CCSD(T*)-F12 and DW-MP2-F12 binding energies (in kcal/mol) for the S22 training set from Ref. [2].

7.2 Improvement through four additional degrees of freedom

In DW-MP2-F12 ≡ DW(2)-MP2 only two parameters have been optimized already giving a great improvement over MP2 and SCS-MP2. The idea of DW(6)-MP2 was to optimize the spin component factors (a, b, α and β) simultaneously to the parameters, g and v responsible for the shape of the cutting function. The improvement was rather small which indicates that the conventional MP2 and the pre optimized SCS-MP2 are already good for the description of the extreme cases (pure dispersion and dipole dominated structures).

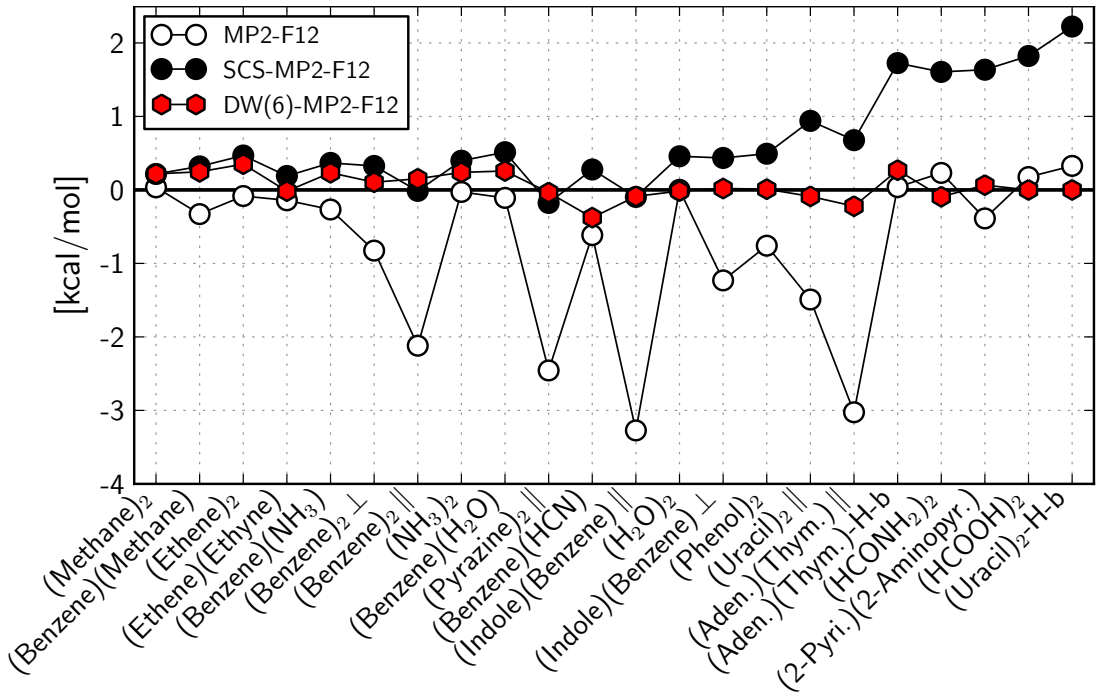


Figure 18: Intrinsic error of DW(6)-MP2 - S22 (training set), reference: CCSD(T*)-F12a/same basis (AVDZ)

The MAD,RMS and maximal deviations from the reference for MP2 SCS-MP2 and DW-MP2-F12 with 2 and 6 optimized parameters is shown in the following table:

RMS, MAD and Max [kcal/mol]			
Method	RMS	MAD	Max
MP2	1.29	0.82	-3.27
SCS-MP2	0.95	0.70	2.22
DW(6)-SCS-MP2	0.18	0.14	-0.38
DW(2)-SCS-MP2	0.23	0.20	-0.41

Table XIX: RMS MAD Max intrinsic error (S22)

In DW(6)-MP2-F12 the SCS-factors a, b, α, β along with the parameters g and v have been optimized simultaneously on the S22 training set using the previously mentioned program. For DW(6)-MP2-F12 they are $g = 0.16736$ and $v = -2.57676$, SCS-MP2$_{ab}$: $a = 0.2922; b = 1.0104$ and SCS-MP2$_{\alpha\beta}$: $\alpha = 0.6524; \beta = 1.0276$.
As in DW-MP2-F12 the values for g and v define the value for x where DW(6)-MP2-F12 is a mixing of SCS-MP2$_{ab}$ and SCS-MP2$_{\alpha\beta}$ half and half and the slope for the the mixing function.

-F12/AVDZ	CCSD(T*)	HF	MP2	SCS-MP2	DW(6)-MP2
Hydrogen bonded cluster:					
$(NH_3)_2$	-3.109	-1.397	-3.137	-2.712	-2.870
$(H_2O)_2$	-4.923	-3.585	-4.929	-4.472	-4.938
(Formic acid)$_2$	-18.629	-15.294	-18.510	-16.863	-18.629
(Formamide)$_2$	-15.941	-12.189	-15.745	-14.371	-16.034
Uracil$_2$ (C_{2h})	-20.630	-16.276	-20.341	-18.445	-20.630
2-pyri. $\cdots$ 2-amino-py.	-16.974	-10.503	-17.365	-15.341	-16.909
Adenine $\cdots$ Thymine	-16.717	-10.080	-16.585	-14.687	-16.448
cluster with predominant dispersion contribution:					
$(CH_4)_2$	-0.533	0.373	-0.495	-0.313	-0.314
$(C_2H_4)_2$	-1.502	0.835	-1.578	-1.027	-1.151
Benzene $\cdots$ CH_4	-1.473	1.171	-1.781	-1.132	-1.227
(Benzene)$_2$ $\parallel$	-2.896	5.375	-4.918	-2.811	-2.744
(Pyrazine)$_2$	-4.538	4.178	-6.894	-4.615	-4.574
Indole $\cdots$ Benzene $\parallel$	-4.914	7.066	-8.036	-4.859	-5.002
Uracil$_2$ (C_2)	-10.157	0.046	-11.119	-8.327	-10.245
Adenine $\cdots$ Thymine $\parallel$	-12.291	3.309	-14.830	-10.634	-12.513
Mixed cluster:					
Ethene $\cdots$ Ethine	-1.508	-0.434	-1.646	-1.314	-1.525
Benzene $\cdots$ H_2O	-3.229	-0.947	-3.453	-2.831	-2.971
Benzene $\cdots$ NH_3	-2.313	0.177	-2.600	-1.966	-2.080
Benzene $\cdots$ HCN	-4.487	-2.005	-5.099	-4.206	-4.864
(Benzene)$_2$ $\perp$	-2.783	1.515	-3.568	-2.416	-2.764
Indole $\cdots$ Benzene $\perp$	-5.721	0.285	-6.900	-5.234	-5.616
(Phenol)$_2$	-7.148	-1.845	-7.702	-6.272	-7.138

Table XX: CCSD(T*)-F12 and DW(6)-MP2-F12 binding energies (in kcal/mol) for the S22 training set from Ref.[2]

7.3 Test set with own structures

Besides the S22 test set most sets suitable the calculating weak interactions consist of rather big systems. A set of new small and medium size systems has been built to test DW-MP2-F12 on systems not included in the training set. The monomers have been optimized with MP2/AVTZ if not present in the S22 test set. The supersystems then have been optimized with MP2/AVDZ freezing the intramolecular degrees of freedom. All systems for which optimized structures could be created and did not happen to have multireference character have been used to test DW-MP2-F12. The set consists of 21 systems (including 2 geometries optimized by Vogiatzis[71]) and will be called T21 test set.

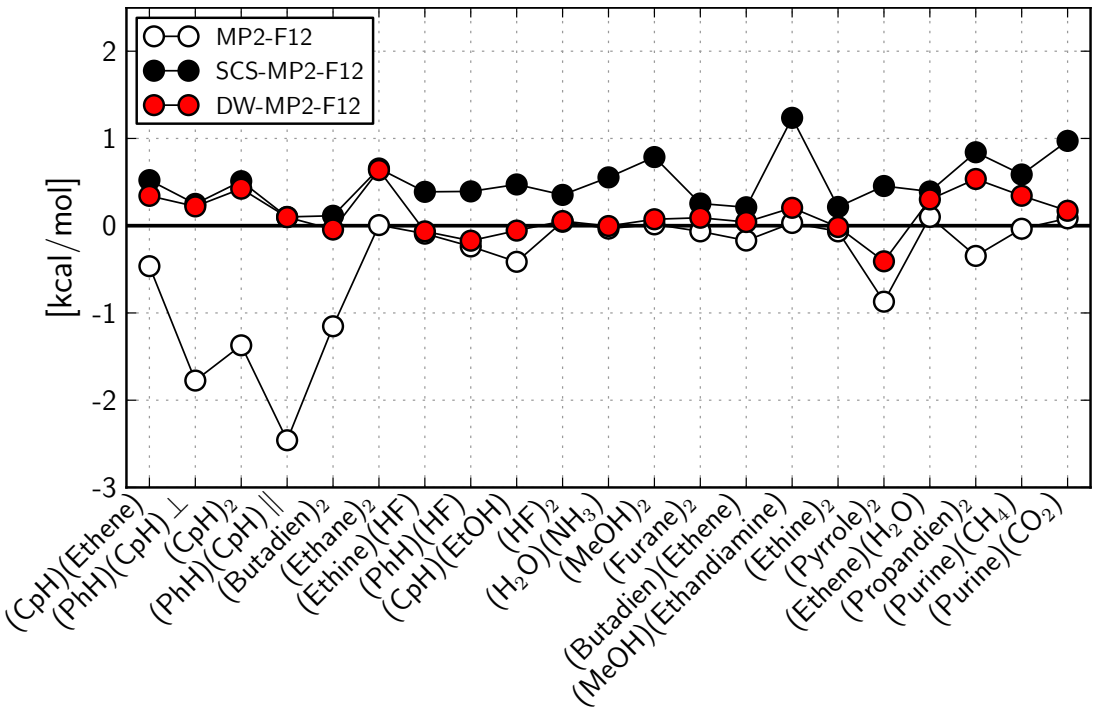


Figure 19: Deviation from CCSD(T*)-F12 of DW-MP2-F12 - T21 (test set), reference: CCSD(T*)-F12a/same basis (AVDZ)

The performance for the T21 set is comparable with the performance for the S22 training set. The following table summarizes the key statistics:

RMS, MAD and Max [kcal/mol]			
Method	RMS	MAD	Max
MP2-F12	0.81	0.47	-2.46
SCS-MP2-F12	0.56	0.49	1.24
DW-MP2-F12	0.27	0.20	0.64

Table XXI: RMS MAD Max of the deviation from CCSD(T*)-F12 (T21)

-F12/AVDZ	CCSD(T*)	HF	MP2	SCS-MP2	DW-MP2
Cluster					
(CpH)(Ethene)	-2.462	1.404	-2.969	-1.981	-2.162
(PhH)(CpH) $\perp$	-2.013	8.353	-4.473	-1.913	-1.915
$(CpH)_2$	-2.816	5.990	-4.608	-2.368	-2.390
(PhH)(CpH) $\parallel$	-2.192	4.308	-3.819	-2.086	-2.118
$(Butadien)_2$	-2.440	2.134	-3.593	-2.327	-2.484
$(Ethane)_2$	-1.457	1.480	-1.450	-0.801	-0.819
(Ethine)(HF)	-4.455	-2.860	-4.205	-3.717	-4.180
(PhH)(HF)	-4.583	-2.178	-4.487	-3.856	-4.421
(CpH)(EtOH)	-3.357	-0.085	-3.771	-2.884	-3.413
$(HF)_2$	-4.494	-3.856	-4.394	-4.086	-4.386
$(H_2O)(NH_3)$	-6.284	-4.387	-6.317	-5.731	-6.288
$(MeOH)_2$	-6.088	-3.506	-6.072	-5.301	-6.014
$(Furane)_2$	-1.208	0.088	-1.274	-0.954	-1.117
(Butadien)(Ethene)	-1.327	0.210	-1.497	-1.116	-1.285
(MeOH)(Ethanediamine)	-7.487	-2.896	-7.456	-6.251	-7.282
$(Ethine)_2$	-1.382	-0.464	-1.448	-1.168	-1.398
$(Pyrrole)_2$	-5.231	-0.511	-6.102	-4.776	-5.639
$(Ethene)(H_2O)$	-1.029	0.267	-0.928	-0.637	-0.728
$(Propadien)_2$	-3.532	1.391	-3.878	-2.690	-2.997
$(Purine)(CH_4)$	-2.180	2.144	-2.749	-1.682	-1.752
$(Purine)(CO_2)$	-5.956	-2.923	-5.813	-4.925	-5.726

Table XXII: CCSD(T*)-F12 and DW-MP2-F12 binding energies (in kcal/mol) for the T21 test set.

7.4 JSCH-2005 test sub set

The JSCH-2005[68] benchmark set contains over 100 complexes with structure of biomacromolecules like DNA, RNA and proteins. Those systems are determined by non-covalent interactions among the building blocks of DNA and RNA bases or amino acids. Most of them are to large for CCSD(T) calculations. Some small randomly selected uncharged non-repulsive complexes have been set up for calculations and the results are displayed in **Figure 20** and **Table XXIII**.

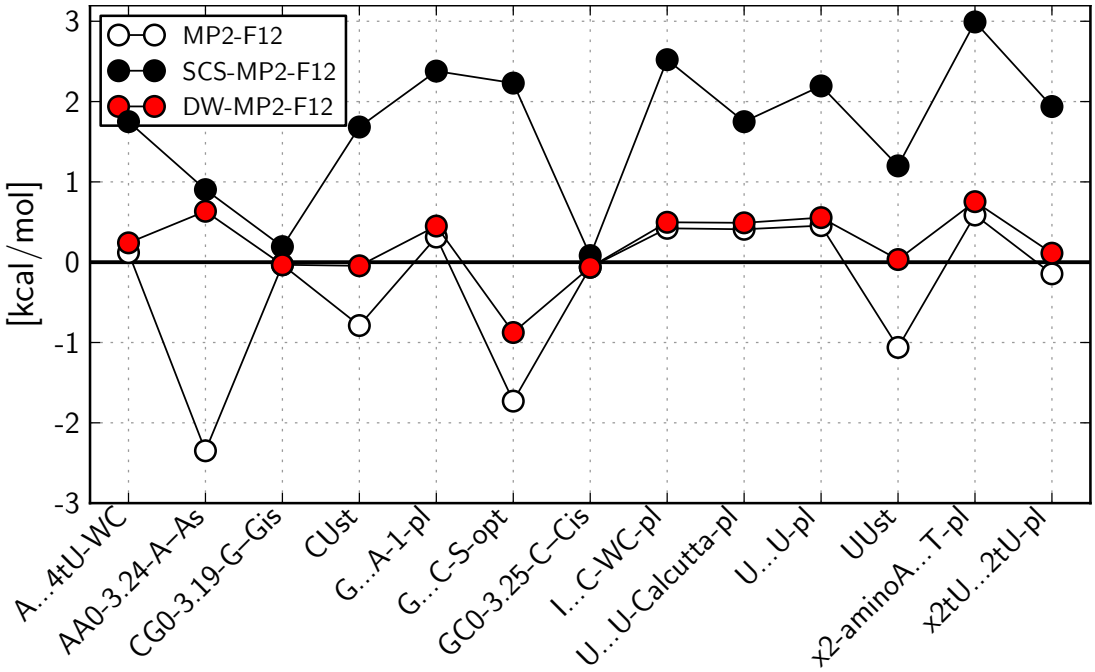


Figure 20: Deviation from CCSD(T*)-F12 of DW-MP2-F12 - JSCH-2005 (test sub set), reference: CCSD(T*)-F12a/same basis (AVDZ)

The performance for the JSCH-2005 test sub set is again comparable with the performance for the S22 training set. The following table summarizes the key statistics:

RMS, MAD and Max [kcal/mol]			
Method	RMS	MAD	Max
MP2-F12	0.97	0.69	2.35
SCS-MP2-F12	1.87	1.66	2.99
DW-MP2-F12	0.48	0.39	0.88

Table XXIII: RMS MAD Max of the deviation from CCSD(T*)-F12 (JSCH2005)

-F12/AVDZ Cluster	CCSD(T*)	HF	MP2	SCS-MP2	DW-MP2
A ··· 4tU-WC	-14.478	-8.356	-14.359	-12.728	-14.237
AA0-3.24-A–As	-6.289	6.161	-8.637	-5.385	-5.655
CG0-3.19-G–Gis	1.156	2.368	1.123	1.349	1.123
CUst	-10.378	-1.580	-11.167	-8.695	-10.423
G ··· A-1-pl	-19.249	-11.650	-18.936	-16.869	-18.799
G ··· C-S-opt	-19.252	-5.440	-20.982	-17.022	-20.127
GC0-3.25-C–Cis	2.988	4.142	2.925	3.070	2.925
I ··· C-WC-pl	-24.796	-19.159	-24.374	-22.275	-24.299
U ··· U-Calcutta	-10.411	-6.448	-10.001	-8.661	-9.921
U ··· U-pl	-14.051	-8.923	-13.591	-11.851	-13.491
UUst	-7.533	0.547	-8.594	-6.333	-7.500
x2-aminoA ··· T-pl	-20.403	-11.965	-19.815	-17.412	-19.649
x2tU ··· 2tU-pl	-12.421	-5.467	-12.566	-10.483	-12.309

Table XXIV: CCSD(T*)-F12 and DW-MP2-F12 binding energies (in kcal/mol) for the JSCH2005 test set.

7.5 S66[3]

The S66 set is a set of 66 clusters. It contains hydrogen bonded, dispersion dominated and mixed clusters like the S22 set but in contrast to the S22 set the S66 set has a number of alkane containing clusters which are of course also dispersion dominated but turned out to be difficult to be treated with DW-MP2-F12 and has already been pointed out by the authors of [3]. The alkane containing clusters are due to their pure dispersion dominated interaction treated with SCS factors close to pure SCS-MP2-F12 but turned out to be described close to CCSD(T*)-F12 level with MP2-F12.

The alkane containing clusters form a subset of 15 systems:

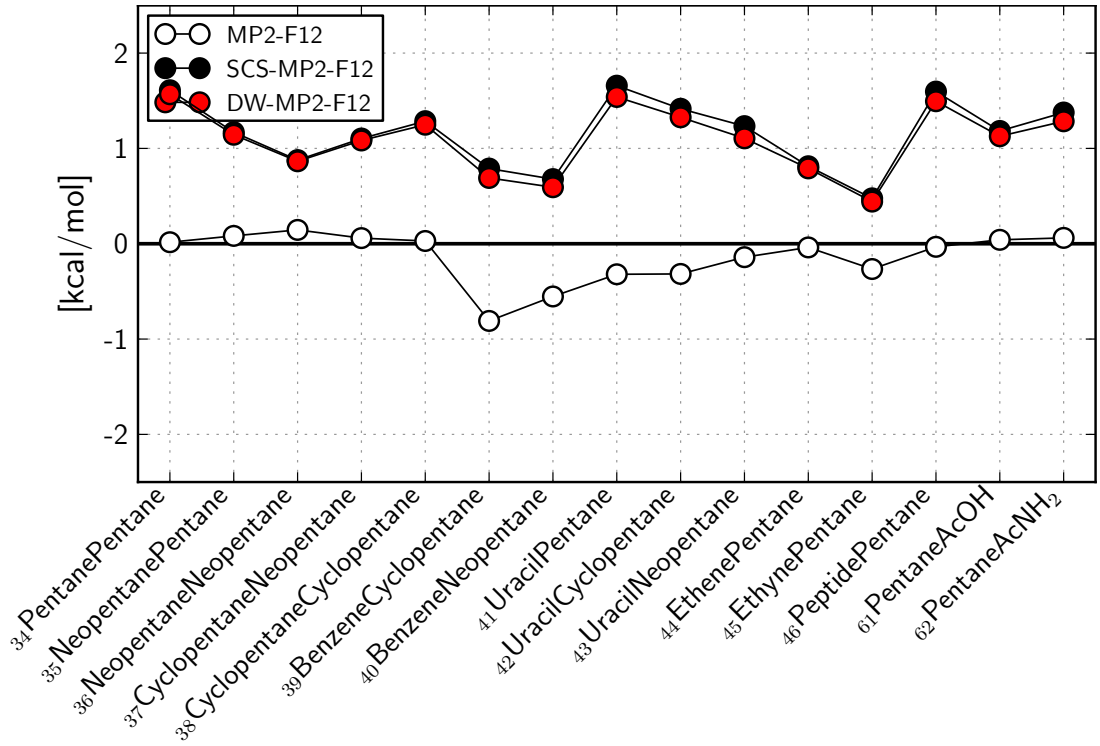


Figure 21: Deviation from CCSD(T*)-F12 of DW-MP2-F12 - S66 subset (alkane clusters), reference: CCSD(T*)-F12a/same basis (AVDZ)

One can see that the alkane containing clusters are not described well by DW-MP2-F12. They are better described by MP2-F12 and not as expected by SCS-MP2-F12. The following table summarizes the key statistics:

RMS, MAD and Max [kcal/mol]			
Method	RMS	MAD	Max
MP2-F12	0.09	0.44	-0.81
SCS-MP2-F12	1.45	1.07	1.66
DW-MP2-F12	1.19	1.04	1.57

Table XXV: RMS MAD Max of the deviation from CCSD(T*)-F12 (S66 alkane containing clusters)

The dispersion dominated clusters form a subset of 10 systems:

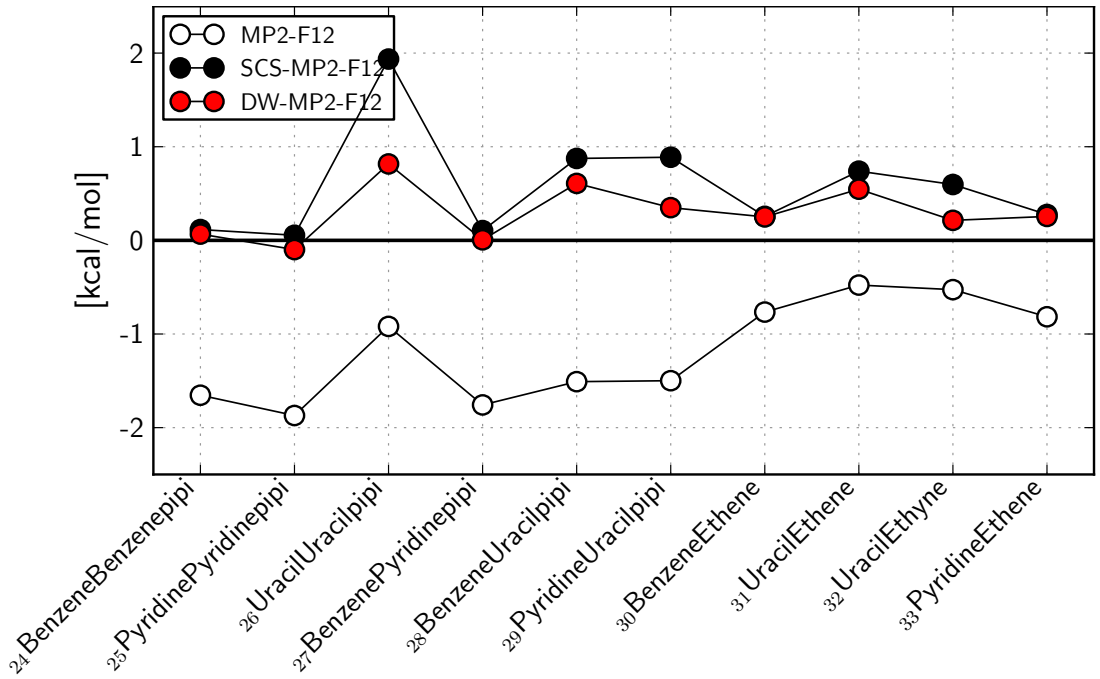


Figure 22: Deviation from CCSD(T*)-F12 of DW-MP2-F12 - S66 subset (dispersion dominated clusters), reference: CCSD(T*)-F12a/same basis (AVDZ)

The dispersion dominated clusters do well on DW-MP2-F12 level even if some systems are involved that are very different to the S22 set. The following table summarizes the key statistics:

RMS, MAD and Max [kcal/mol]			
Method	RMS	MAD	Max
MP2-F12	1.64	1.09	-1.87
SCS-MP2-F12	0.64	0.76	1.94
DW-MP2-F12	0.16	0.57	0.82

Table XXVI: RMS MAD Max of the deviation from CCSD(T*)-F12 (S66 dispersion dominated clusters)

The clusters with mixed interactions form a subset of 18 systems:

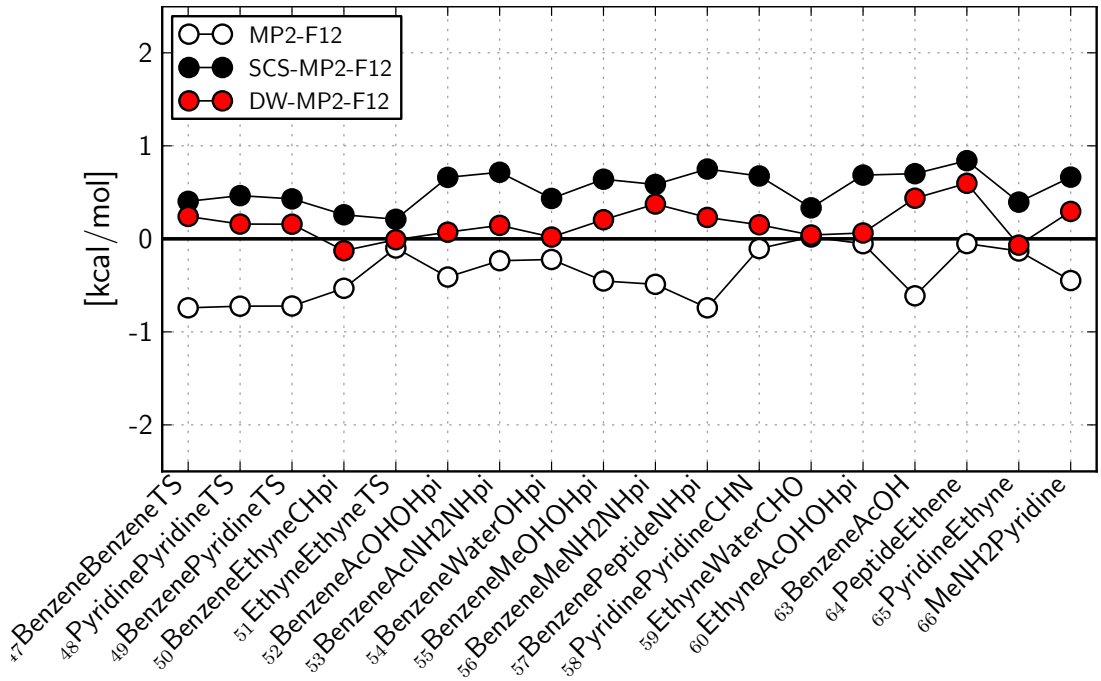


Figure 23: Deviation from CCSD(T*)-F12 of DW-MP2-F12 - S66 subset (clusters with mixed interactions), reference: CCSD(T*)-F12a/same basis (AVDZ)

The clusters with mixed interactions do well on DW-MP2-F12 level even if some systems are involved that are very different to the S22 set. The following table summarizes the key statistics:

RMS, MAD and Max [kcal/mol]			
Method	RMS	MAD	Max
MP2-F12	0.21	0.61	-0.74
SCS-MP2-F12	0.33	0.74	0.84
DW-MP2-F12	0.06	0.43	0.60

Table XXVII: RMS MAD Max of the deviation from CCSD(T*)-F12 (S66 mixed clusters)

The clusters dominated by hydrogen bond form a subset of 23 systems:

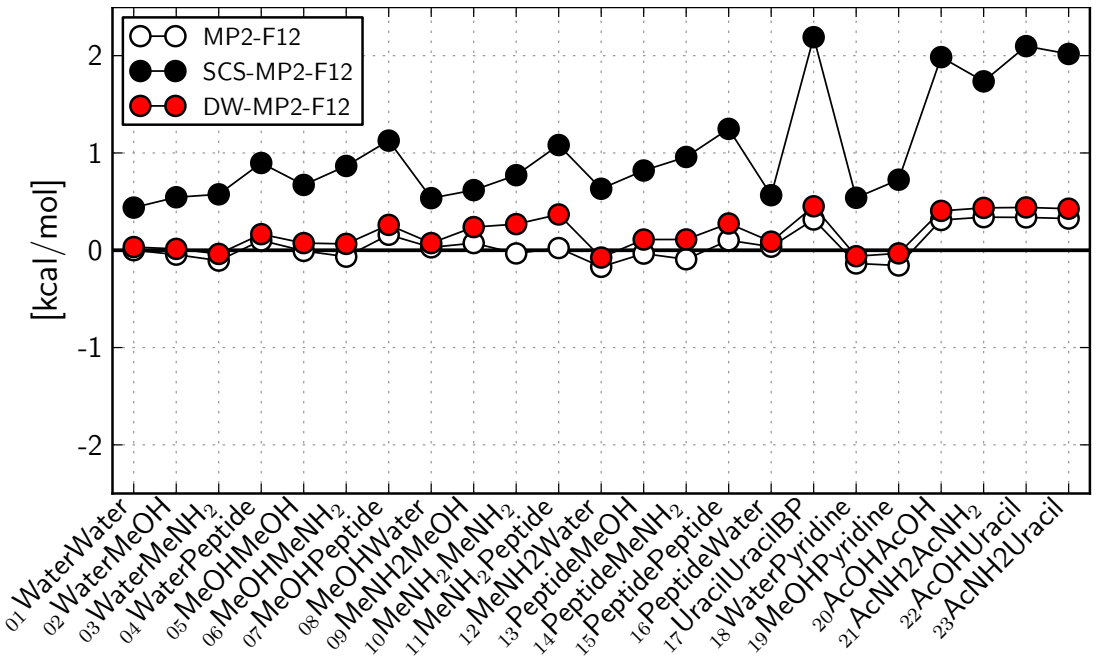


Figure 24: Deviation from CCSD(T*)-F12 of DW-MP2-F12 - S66 training set (hydrogen bond clusters), reference: CCSD(T*)-F12a/same basis (AVDZ)

The hydrogen bond dominated clusters also behave like expected and are described well by DW-MP2-F12. The following table summarizes the key statistics:

RMS, MAD and Max [kcal/mol]			
Method	RMS	MAD	Max
MP2-F12	0.03	0.36	0.34
SCS-MP2-F12	1.37	1.01	2.19
DW-MP2-F12	0.06	0.44	0.45

Table XXVIII: RMS MAD Max of the deviation from CCSD(T*)-F12 (S66 hydrogen bond)

For the whole S66 set the RMS MAD Max of the deviation from CCSD(T*)-F12 are summarized in the following table:

RMS, MAD and Max [kcal/mol]			
Method	RMS	MAD	Max
MP2-F12	0.58	0.37	-1.87
SCS-MP2-F12	1.00	0.86	2.19
DW-MP2-F12	0.58	0.42	1.57

Table XXIX: RMS MAD Max of the deviation from CCSD(T*)-F12 (S66 all)

With the exception of the methane dimer from the S22 set and the ethane dimer from the T21 set, alkane containing clusters have not been tested before the S66 set due to the fact that the do not play a role in biological systems and therefore didn't seem of interest. Looking back with the knowledge gained by using the S66 test set it is seen that the ethane dimer having the worst performance could have been an early indicator. The methane dimer with a interaction energy of approximately 0.5 kcal/mol did not give an indication.

7.6 Other properties

The good results[d] obtained for the intermolecular interaction energies where encouraging to use DW-MP2-F12 for a geometry optimization to obtain accurate geometries using a MP2 based method and compare them to CCSD(T*)-F12 optimized geometries.

7.6.1 Geometry optimization with DW-MP2-F12

Comparison of the geometries optimized with various MP2-F12 - based - methods to the geometries optimized using CCSD(T*)-F12a has been done using a nine step procedure. This nine step procedure is presented in a keyword scheme for reasons of understandability:

- Building of monomers (if not present in the S22 test set)
- Optimization of the all degrees of freedom in the monomer (intramolecular) using MP2/AVTZ analytical gradients (if not present in the S22 test set)
- Formation of (20) dimers in a random (non penetrating) structure
- Optimization of the all degrees of freedom in the dimer (intermolecular) with fixed monomers (freezing intramolecular degrees of freedom) using MP2/AVDZ analytical gradients with counterpoise correction[22] (pre-optimization) (see **Section 11.1** (Appendix))
- Optimization using numerical gradients of the intermolecular distance with counterpoise correction, freezing intramolecular and the other intermolecular degrees of freedom (see **Section 11.1** (Appendix)) using:
- MP2-F12/AVDZ
- SCS-MP2-F12/AVDZ
- DW-MP2-F12/AVDZ
- DW(6)-MP2-F12/AVDZ
- CCSD(T*)-F12a/AVDZ (reference)

The optimization of the clusters using different methods have be carried out on a single degree of freedom in order to able to compare the values with each other. For the optimization of all methods explicit correlation and the same basis set has been used to compare

[d] alkane containing clusters where not tested at that time

only of deviation from CCSD(T*)-F12. The distance[e] of the CCSD(T*)-F12a/AVDZ optimized structure has been used as a reference and compared with the same distance of the supersystem optimized using one of the other methods. The deviation of the distances from CCSD(T*)-F12a for each system is compared for all methods and is plot in the following diagram:

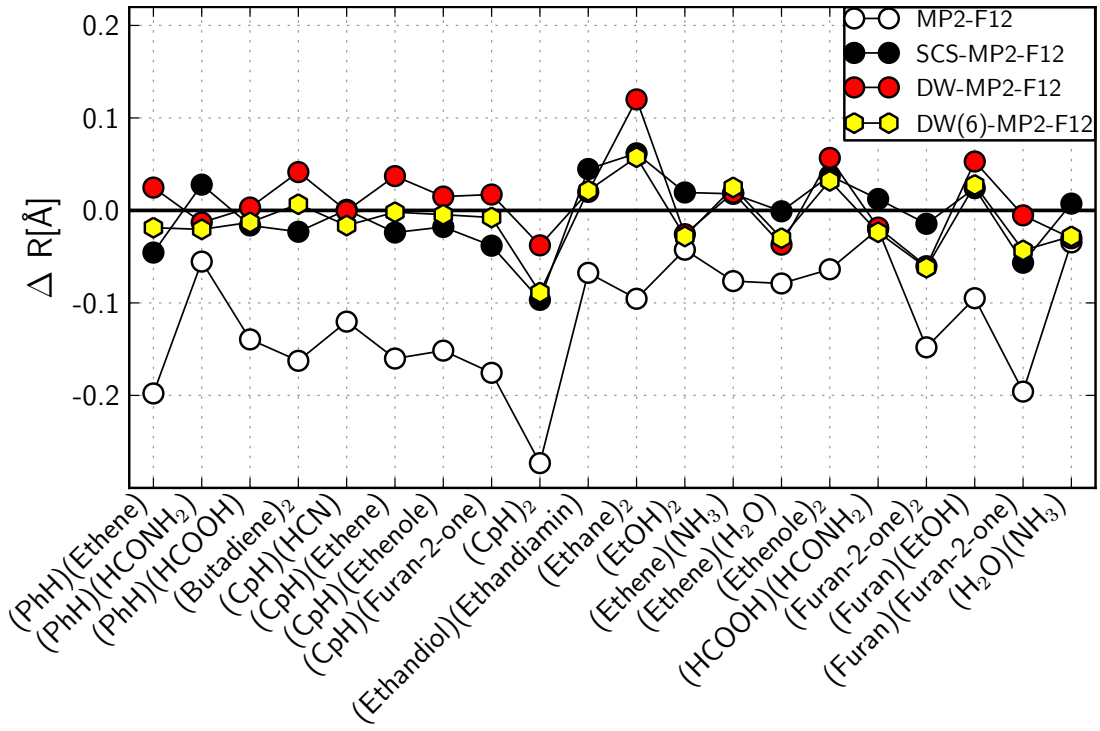


Figure 25: Numerical gradient geometry optimization with DW-MP2-F12 - reference: CCSD(T*)-F12a optimized structure

The statistics for the different methods are summarized in the following table:

RMS, MAD and Max [Å]			
Method	RMS	MAD	Max
MP2-F12	0.016	0.109	-0.27
SCS-MP2-F12	0.001	0.029	-0.10
DW-MP2-F12	0.002	0.030	0.12
DW(6)-MP2-F12	0.001	0.027	-0.09

Table XXX: Error of numerical gradient geometry optimization with MP2-F12, SCS-MP2-F12 and DW-MP2-F12

[e] distance of two random atoms each in a different monomer

The new developed DW-MP2-F12 method results in accurate geometries but the SCS-MP2-F12 method is giving similar results and the advantage of improving over all methods with the same costs is not seen for geometries. The DW-MP2-F12 method can be used for geometry optimization giving good results, but since SCS-MP2-F12 gives the same accuracy it is not advised to use DW-MP2-F12 for geometry optimizations.

-F12/AVDZ	CCSD(T*)	MP2	SCS-MP2	DW(2)-MP2	DW(6)-MP2
Cluster					
(PhH)(Ethene)	3.989	3.791	3.943	4.014	3.970
(PhH)($HCONH_2$)	6.980	6.925	7.008	6.967	6.960
(PhH)(HCOOH)	3.503	3.363	3.486	3.506	3.490
$(Butadiene)_2$	6.587	6.424	6.564	6.628	6.594
(CpH)(HCN)	6.217	6.097	6.218	6.217	6.201
(CpH)(Ethene)	3.761	3.601	3.737	3.798	3.759
(CpH)(Ethenole)	4.133	3.981	4.115	4.148	4.128
(CpH)(Furan-2-one)	5.098	4.922	5.059	5.115	5.090
$(CpH)_2$	4.043	3.770	3.947	4.006	3.955
(Ethandiol)(Ethandiamin)	4.333	4.266	4.378	4.353	4.355
$(Ethane)_2$	3.920	3.825	3.982	4.040	3.978
$(EtOH)_2$	5.257	5.215	5.277	5.232	5.229
(Ethene)(NH_3)	3.644	3.567	3.662	3.663	3.669
(Ethene)(H_2O)	3.477	3.398	3.476	3.440	3.447
$(Ethenole)_2$	4.095	4.031	4.132	4.151	4.127
(HCOOH)($HCONH_2$)	3.931	3.911	3.943	3.912	3.907
$(Furan\text{-}2\text{-}one)_2$	4.686	4.538	4.672	4.626	4.624
(Furan)(EtOH)	5.917	5.822	5.941	5.970	5.945
(Furan)(Furan-2-one)	3.409	3.213	3.352	3.403	3.366
(H_2O)(NH_3)	2.975	2.940	2.982	2.944	2.946

Table XXXI: CCSD(T*)-F12, MP2-F12, SCS-MP2-F12, DW(2)-MP2-F12 and DW(6)-MP2-F12 distances (in Å).

7.6.2 DW-CCSD(T**)-F12

Michael S. Marshall and C. David Sherrill have combined[72] the two CCSD(T*)-F12 approximations F12a and F12b to Dispersion-Weighted Explicitly Correlated Coupled-Cluster Theory (DW-CCSD(T**)-F12). They use the same *tanh*() type function as for DW-MP2-F12 (see **Section 6**) but here MP2-F12 and SCS-MP2-F12 are replaced by CCSD(T*)-F12a and CCSD(T*)-F12b:

$$\omega \cdot \text{CCSD(T*)-F12a} + (1 - \omega) \cdot \text{CCSD(T*)-F12b} \tag{84}$$

The notation (T**) used by Michael S. Marshall and C. David Sherrill denotes size consistent scaling and is identical to (T*) used in this work after **Section 5.4.1**.
The other equations are identical (see **Equations (81)** and **(82)**) just with interchanged variables:

$$\omega = \frac{1}{2} - \frac{1}{2}\tanh(\alpha + \beta \cdot d)$$

$$d = \frac{\Delta E_{HF_{+s}}(M)}{\Delta E_{MP2-F12}(M)}$$

They optimized the parameters on the S22 test set, which resulted in $\alpha = -1$ and $\beta = 4$. For the S22 set DW-CCSD(T**)-F12 improves over CCSD(T*)-F12a and CCSD(T*)-F12b:

RMS, MAD and Max [kcal/mol]			
	CCSD(T*)-F12a	CCSD(T*)-F12b	DW-CCSD(T**)-F12
Max	-0.57	0.33	0.13
MAD	0.12	0.10	0.05
RMS	0.19	0.13	0.07

Table XXXII: RMS MAD Max basis set error of DW-CCSD(T**)-F12 for the S22 training set from ref. [72].

These improvements in accuracy come free of computational cost. Notice that the set of systems (and the reference) which Marshall and Sherrill used to optimize Dispersion-Weighted Explicitly Correlated Coupled-Cluster Theory is different from the values used

here. The results therefore do not agree with the picture shown in **Section (5.5.3)** for the comparison of F12a and F12b.

7.6.3 DW-CCSD(T**)-F12 on S66

The DW-CCSD(T**)-F12 method has been tested on the same subset of the S66[3] test set as in **Figure 11** using the parameters α and β of Marshall and Sherrill [72]. CCSD(T)/CBS[3'4'] has been used as the reference and was available for only 21 of the 66 systems. The results confirm the improvement over F12a and F12b found by Marshall and Sherrill.

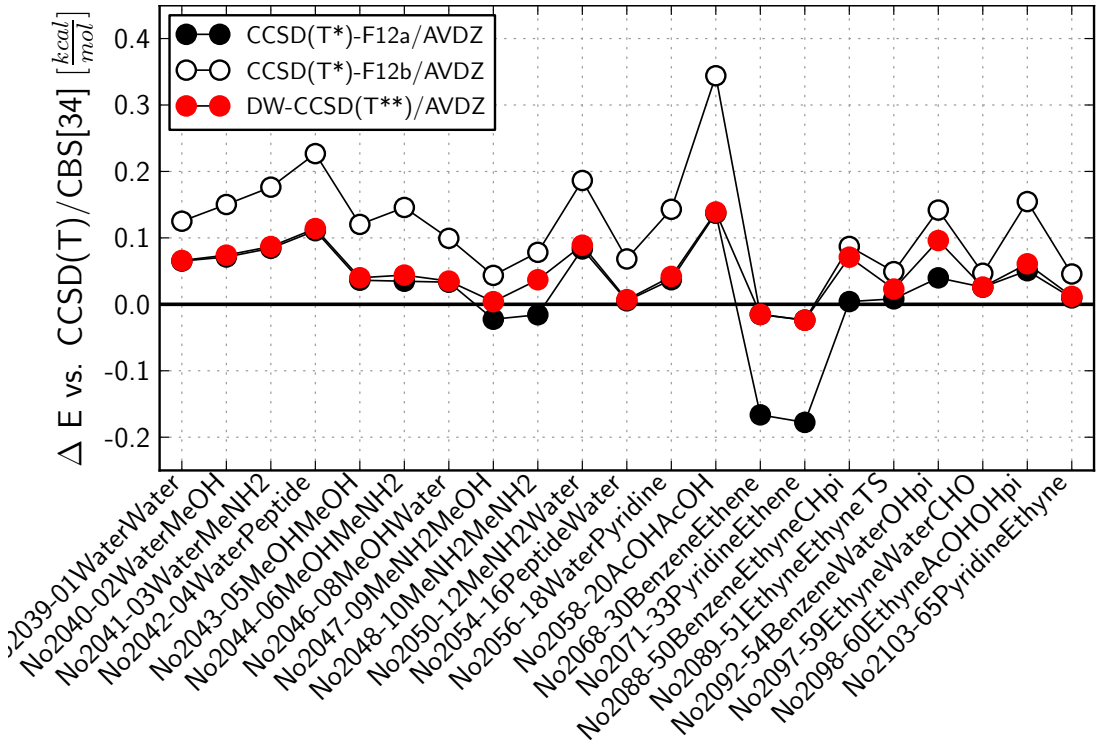


Figure 26: F12a vs. F12b vs DW-CCSD(T**)-F12 with AVDZ basis, reference: CCSD(T)/CBS[3'4']

The following table summarizes the key statistics:

RMS, MAD and Max [kcal/mol]			
	CCSD(T*)-F12a	CCSD(T*)-F12b	DW-CCSD(T**)-F12
Max	-0.178	0.344	0.139
MAD	0.058	0.118	0.053
RMS	0.077	0.140	0.064

Table XXXIII: RMS MAD Max basis set error for the S66 test set

Based on the data of the paper[72] of Marshall and Sherrill (see **Section 7.6.2**), and the results from the 21 systems of the S66 subset we calculated here, the following conclusions can be drawn: F12a works well for any weak interaction based on electrostatics, but F12b underestimates those interaction energies. On the other hand, for weak interactions caused by instantaneous multipoles (dispersion[7]) F12a overshoots and F12b gives interaction energies closer to the basis set limit.

	CCSD(T*)/AVDZ			CCSD(T)
Cluster	DW-F12	F12a	F12b	CBS[3'4']
No2039_01WaterWater	-4.923	-4.924	-4.864	-4.989
No2040_02WaterMeOH	-5.609	-5.612	-5.533	-5.683
No2041_03WaterMeNH2	-6.927	-6.930	-6.838	-7.014
No2042_04WaterPeptide	-8.099	-8.102	-7.986	-8.213
No2043_05MeOHMeOH	-5.799	-5.803	-5.719	-5.839
No2044_06MeOHMeNH2	-7.606	-7.615	-7.504	-7.650
No2046_08MeOHWater	-5.038	-5.040	-4.974	-5.073
No2047_09MeNH2MeOH	-3.087	-3.113	-3.047	-3.091
No2048_10MeNH2MeNH2	-4.165	-4.218	-4.124	-4.202
No2050_12MeNH2Water	-7.292	-7.298	-7.195	-7.381
No2054_16PeptideWater	-5.176	-5.178	-5.115	-5.183
No2056_18WaterPyridine	-6.907	-6.912	-6.806	-6.949
No2058_20AcOHAcOH	-19.282	-19.284	-19.077	-19.421
No2068_30BenzeneEthene	-1.362	-1.513	-1.362	-1.347
No2071_33PyridineEthene	-1.810	-1.964	-1.810	-1.786
No2088_50BenzeneEthyneCHpi	-2.780	-2.847	-2.764	-2.851
No2089_51EthyneEthyneTS	-1.512	-1.527	-1.486	-1.535
No2092_54BenzeneWaterOHpi	-3.181	-3.237	-3.135	-3.277
No2097_59EthyneWaterCHO	-2.888	-2.888	-2.867	-2.914
No2098_60EthyneAcOHOHpi	-4.882	-4.892	-4.788	-4.943
No2103_65PyridineEthyne	-4.065	-4.067	-4.031	-4.077

Table XXXIV: DW-CCSD(T**)-F12/AVDZ, CCSD(T*)-F12a/AVDZ, CCSD(T*)-F12b/AVDZ and CCSD(T)/CBS[3'4'] binding energies (in kcal/mol) for the S66 test (sub-)set

8 Other methods

A large number of methods have been developed for the description of weak interactions, and not all of them can be discussed in detail. Important DFT based methods are the empirical DFT-D3 correction[73] by Grimme and coworkers and the non-empirical dispersion functional[74] by Vydrov and Van Voorhis. Other approaches apply spin component scaling to CCSD, for example SCS-CCSD[75] and SCS-MI-CCSD[76] by Takatani, Hohenstein and Sherrill. We here focus on methods similar in spirit to DW-MP2-F12.

One example is the DF-SCSN-LMP2[77] by Hill and Platts, which is optimized for evaluating the interaction energy between nucleic acid base pairs. In this method the SCS-Factors are re-scaled to -0.53 for antiparallel and 2.28 for parallel spin. In their paper *Spin-Component Scaling Methods for Weak and Stacking Interactions* they compare DF-SCSN-LMP2 (SCSN) to two other Spin Component Scaled MP2 versions (SCS[5] and SOS[78]). The SCS results employed the default scaling factors of 6/5 for antiparallel spins and 1/3 for parallel spins, while for SOS data the parallel-spin scaling factor is set to zero and that of the antiparallel spins to 1.3. Here the results from Ref. [77], Table 2 are reproduced and compared with own DW-MP2-F12 values. It should be mentioned that the reference values (best[f]) they compared to, taken from paper *Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs*[68] are estimated CBS values. Since DW-MP2-F12 with AVDZ is already close to the basis set limit, it preforms well:

[f] best estimate

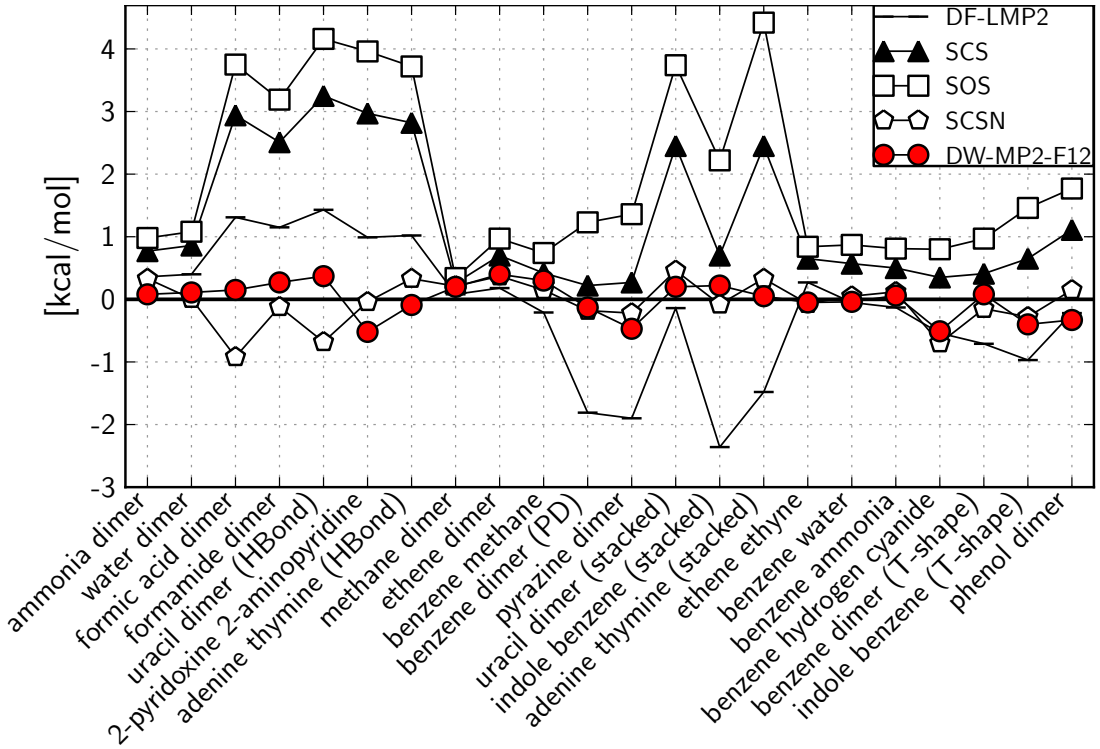


Figure 27: Intrinsic error DF-LMP2, SCS, SOS, SCSN and DW-MP2-F12 - S22 training set - reference "best estimate" from Ref. [68]

Notice, that the values for SCS and MP2 are different from those in **Table XVIII**.

complex	DF-LMP2	SCS	SOS	SCSN	best[68]	DW
ammonia dimer	-2.81	-2.40	-2.19	-2.84	-3.17	-3.09
water dimer	-4.62	-4.16	-3.94	-5.01	-5.02	-4.91
formic acid dimer	-17.30	-15.67	-14.86	-19.53	-18.61	-18.46
formamide dimer	-14.81	-13.45	-12.77	-16.08	-15.96	-15.69
uracil dimer (HBond)	-19.22	-17.40	-16.49	-21.33	-20.65	-20.28
2-pyridoxine 2-aminopyridine	-15.72	-13.74	-12.75	-16.75	-16.71	-17.23
adenine thymine (HBond)	-15.35	-13.55	-12.65	-16.04	-16.37	-16.46
methane dimer	-0.45	-0.27	-0.19	-0.32	-0.53	-0.33
ethene dimer	-1.33	-0.81	-0.54	-1.14	-1.51	-1.11
benzene methane	-1.71	-1.08	-0.76	-1.35	-1.50	-1.20
benzene dimer (PD)	-4.54	-2.51	-1.50	-2.91	-2.73	-2.86
pyrazine dimer	-6.32	-4.15	-3.06	-4.64	-4.42	-4.89
uracil dimer (stacked)	-10.26	-7.67	-6.38	-9.66	-10.12	-9.92
indole benzene (stacked)	-7.58	-4.52	-3.00	-5.30	-5.22	-5.00
adenine thymine (stacked)	-13.71	-9.78	-7.81	-11.90	-12.23	-12.18
ethene ethyne	-1.26	-0.88	-0.69	-1.60	-1.53	-1.58
benzene water	-3.33	-2.71	-2.41	-3.23	-3.28	-3.32
benzene ammonia	-2.48	-1.85	-1.54	-2.23	-2.35	-2.29
benzene hydrogen cyanide	-5.00	-4.11	-3.66	-5.16	-4.46	-4.97
benzene dimer (T-shape)	-3.45	-2.33	-1.77	-2.89	-2.74	-2.66
indole benzene (T-shape)	-6.70	-5.08	-4.27	-6.01	-5.73	-6.13
phenol dimer	-7.27	-5.94	-5.28	-6.90	-7.05	-7.38
RMSE	1.05	1.65	2.39	0.36	—	0.28
MD	0.15	-1.26	-1.97	0.04	—	0.00$^{\alpha}$
MAD	0.81	1.26	1.97	0.27	—	0.23
						α: -0.0007

Table XXXV: DF-LMP2, SCS, SOS, SCSN and DW-MP2-F12 binding energies (in kcal/mol) for the S22 training set from Ref. [77].

9 Summary

Several branches of science are interested in accurately describing van der Waals forces. These need to be taken into account when interactions of systems that are not connected by a covalent bond or otherwise are described. The calculation of weak interactions is a challenging task: (1) Due to the intrinsic accuracy of the methods used and (2) the applied basis set size. This has two main reasons.

First, some of the weak interactions, dispersion dominated in particular, are only covered by the correlation contribution and therefore Hartree-Fock (HF) is a very poor approximation. Actually, even a repulsive description of those systems is sometimes obtained. With HF being a poor approximation, the correlation contribution to the energy is rather large and a high-level correlation method is needed. Other complexes like hydrogen bound dominated ones are described qualitatively correct by HF and have a smaller correlation contribution. For those, lower level correlation methods are already sufficient.

Second, big basis sets are needed to gain accurate results, because with increasing importance of the correlation contribution, slow convergence towards the basis set limit is usually obtained.

These complications lead to an unbalanced description of the different types of weak interactions, making the comparison of their strength difficult. Investigating interesting weak interactions often means the consideration of large[g] systems, and the necessary high-level electron correlation methods become computationally challenging. The basis sets needed to gain accurate results then need to be enlarged increasing the overall costs.

The difficulty of the slow convergence of the energy with the basis set size is well known for a long time and also the idea of explicit correlation to overcome this problem originates back to the year 1929 when Hylleraas[43] used explicit electron-distance dependent functions to describe the Helium atom. This however is extremely costly and the whole idea of F12 seemed unpractical and did not find its way to large-scale applications. Only rather recent developments in explicit correlation (see **Section 4.3.1**) were able to make this idea feasible for practical applications in *ab initio* quantum mechanic calculations.

The introduction of explicit correlation takes care of the basis set quality problem by making an AVDZ basis sufficient to stay within a 0.1 kcal/mol error bar to the basis set limit, hence being around 10 times more accurate than without explicit correlation. For CCSD(T), the explicit correlation does not largely increase the computational cost. For

[g] larger than 30 to 40 first and second row atoms

an average case, it increases the cost by about 5%; for big systems even less. In the MP2 case the increase of the cost is higher, but still offset by the increase in accuracy.

To find a solution for the aforementioned problems, this thesis provides systematic investigations and suggests solutions for those problems. As a first step, the available and newly developed explicitly correlated Coupled Cluster methods (CCSD(T*)-F12)[1] where tested in order to evaluate their accuracy to describe intermolecular interactions.

As shown in this work in **Section 5**, various systems of different test sets where investigated. In the S22 test (sub-) set the maximum deviation (MAX) is 0.17 kcal/mol already with AVDZ basis while conventional CCSD(T) has a maximum deviation form the basis set limit of more than 1 kcal/mol. For the S66 the picture is virtually identical.

The explicitly correlated Coupled Cluster methods proved to be highly accurate and provide a good description of the investigated systems. For larger systems, however, the computational costs are too high for practical applications. In particular, the high increase of the computational cost by moving up the hierarchy of *ab initio* methods from MP2 to CCSD(T) illustrates this problem. While MP2 scales $\mathcal{O}(\mathcal{N}^5)$, CCSD(T) already scales with $\mathcal{O}(\mathcal{N}^7)$ making large systems inaccessible.

A method that combines the accuracy of CCSD(T) with the rather low computational demands of MP2 was therefore developed. In this method, MP2 and SCS-MP2 are combined to Dispersion-Weighted-MP2[h] (DW-MP2) as shown in **Section 6**. In the **Section 7** the performance of the new developed method is investigated. It is shown that the high intrinsic accuracy of CCSD(T) is reproduced within a deviation of around 0.5 kcal/mol using DW-MP2 for almost all systems, however, not for alkane clusters.

Furthermore the performance of DW-MP2-F12 is tested for numerical gradients in **Section 7.6.1**. The presented results are then compared to standard methods.

With this, the thesis shows that using a combination of both — DW-MP2 and explicit correlation (DW-MP2-F12) — the accurate description of weak interactions with very eligible computational effort is possible. The DW-MP2-F12 does not introduce any additional costs compared to MP2-F12 and therefore drastically reduces the computational demands compared to the CCSD(T*)-F12 method. The DW-MP2-F12 procedure can be used fully automatic (black-box), making accurate *ab initio* quantum mechanic calculations for weakly interacting systems with 100 atoms or more available.

The fundamental idea of DW-MP2-F12 is also appreciated in the quantum chemical com-

[h] DW-MP2 has only been used in combination with explicit correlation (DW-MP2-F12) in order to avoid basis set errors.

munity which can be seen by more than 20 articles citing the original publications resulting from this thesis. Furthermore, the general concept of DW-MP2-F12 was also applied in the development of Dispersion-Weighted explicitly correlated coupled-cluster theory (DW-CCSD(T**)-F12 by David Sherrill et al. (see **Section 7.6.2**), where the same optimisation principle is transferred in order to optimise CCSD(T*)-F12 results: MP2 and SCS-MP2 are herein replaced with CCSD(T*)-F12a and CCSD(T*)-F12b in order to obtain even further improved results.

10 Zusammenfassung

Verschiedene Bereiche der Wissenschaft sind an der genauen Beschreibung von Vander-Waals Kräften interessiert. Diese müssen in Betracht gezogen werden wenn die Wechselwirkung von zwei Systemen nicht durch eine chemische Bindung oder anderweitig beschrieben wird. Die Berechnung von schwachen Wechselwirkungen ist eine anspruchsvolle Aufgabe: (1) wegen der benötigten intrinsischen Genauigkeit der verwendeten Methoden und (2) und der verwendeten Basissatzgröße. Dafür gibt es zwei Gründe.

Erstens sind einige der schwachen Wechselwirkungen, im Besonderen dispersions dominierte ausschließlich durch den Korrelationsanteil beschrieben, weshalb Hartree-Fock (HF) eine sehr schlechte Nährung darstellt. Tatsächlich wird manchmal eine abstoßende Beschreibung dieser Systeme erhalten. Mit der schlechten Beschreibung durch HF wird der Korrelationsanteil zur Energie eher groß und Korrelationsmethoden auf hohem Niveau werden benötigt. Andere Komplexe wie durch Wasserstoffbücken dominierte werden durch HF qualitativ richtig beschrieben und haben einen kleineren Korrelationsbeitrag. Für diese sind Korrelationsmethoden auf niedrigem Niveau ausreichend.

Zweitens, es werden große Basissätze benötigt um genaue Ergebnisse zu erhalten da der Korrelationsbeitrag an Wichtigkeit gewinnt und die damit einhergehende langsame Konvergenz zum Basissatzlimit auftritt.

Diese Schwierigkeiten führen zu einer unausgewogenen Beschreibung der verschiedenen Typen von schwachen Wechselwirkungen und der Vergleich ihrer Stärke wir damit schwer. Die Untersuchung von interessanten schwachen Wechselwirkungen bedeutet häufig, daß große Systeme[i], für welche die Korrelationsmethoden auf hohen Niveau rechentechnisch zu einer Herausforderung werden, untersucht werden müssten. Die Basissätze, die benötigt werden um genaue Ergebnisse zu erhalten müssen dann vergrößert werden wodurch sich sich die Kosten weiter erhöhen.

Die langsame Konvergenz der Energie mit der Basissatzgröße ist sein längerer Zeit gut bekannt und auch die Idee der expliziten Korrelation zur Lösung dieses Problems geht auf das Jahr 1929 zurück, als Hylleraas[43] explizit auf Elektron-Elektron- Abstände beruhende Funktionen verwendete um das Heliumatom zu beschreiben. Dies ist äußerst kostspielig und die ganze Idee von F12 schien unpraktisch und fand keine breite Anwendung. Erst neuere Entwicklungen in der expliziten Korrelation (siehe **Section 4.3.1**) waren in der Lage diese Idee für praktische Anwendungen in der *ab initio* Quantenmechanik anwendbar

[i] mehr als 30 bis 40 Atome aus der Ersten und zweiten Periode

zu machen.

Die Einführung der expliziten Korrelation beseitigt das Basissatzproblem in dem Sie eine AVDZ Basis ausreichend macht um in der 0,1 kcal/mol Fehlergrenze zum Basissatzlimit zu bleiben und damit zehn mal genauer zu sein als ohne explizite Korrelation. Bei CCSD(T) vergrößert die Explizite Korrelation den Rechenaufwand nicht im größeren Umfang. Im durchschnittlichen Fall erhöht es die Kosten um 5% bei großen Systemen sogar weniger. Im Fall von MP2 ist die Erhöhung der Kosten größer wird aber durch die Erhöhung der Genauigkeit (mehr als) ausgeglichen.

Um eine Lösung für die vorher genannten Probleme zu finden untersucht die Arbeit diese und schlägt Lösungen vor. Als erster Schritt werden die aktuell entwickelten explizit korrelierten Coupled Cluster Methoden CCSD(T*)-F12[1] getestet um ihre Genauigkeit bei der Beschreibung von intermolekularen Wechselwirkungen zu ermitteln.

Wie in dieser Arbeit in **Section 5** zu sehen ist werden unterschiedliche Systeme verschieder Testsätze untersucht. Im S22 Testsatz ist die maximale Abweichung (MAX) schon bei einer AVDZ Basis (nur) 0,17 kcal/mol während bei konventionellem CCSD(T) die Abweichung vom Basissatz-Limit mehr als 1 kcal/mol beträgt. Für den S66 Satz stellt sich das praktisch identisch dar. Die explizit korrelierten Coupled Cluster Methoden zeigten sich als hochgenau und lieferten eine gute Beschreibung der untersuchten Systeme. Für große Systeme sind die Rechenkosten allerdings zu hoch um praktisch anwendbar zu sein. Im Speziellen zeigt sich diese große Problematik bei der Steigerung der Rechenkosten mit dem Aufstieg in der Hierarchie der *ab initio* Methoden von MP2 nach CCSD(T). Während MP2 mit $\mathcal{O}(\mathcal{N}^5)$ skaliert tut, das CCSD(T) schon mit $\mathcal{O}(\mathcal{N}^7)$ und ist damit für große Systeme unanwendbar.

Eine Methode, die die Genauigkeit von CCSD(T) mit dem eher niedrigen Rechenaufwand von MP2 verbindet wurde deshalb entwickelt. In dieser Methode werden MP2 und SCS-MP2 zu Dipersion-Weighted-MP2[j] (DW-MP2) kombiniert wie in **Section 6** beschrieben. In **Section 7** wird die Leistungsfähigkeit der neu entwickelten Methode gezeigt. Es wird gezeigt, daß die hohe intrinsische Genauigkeit von CCSD(T) mit der Verwendung von DW-MP2 bei einer Abweichung von ca. 0,5 kcal/mol für fast alle Systeme reproduziert werden kann, abgesehen von Alkan-Clustern.

Außerdem wird die Leistungsfähigkeit von DW-MP2-F12 für die Verwendung bei numerischen Gradienten in **Section 7.6.1** untersucht. Die (gezeigten) Resultate werden

[j] DW-MP2 wurde ausschließlich in Verbindung mit expliziter Korrelation verwendet (DW-MP2-F12) um Basissatz Fehler zu vermeiden

mit Standardmethoden verglichen.

Hiermit zeigt diese Arbeit, daß mit der Kombination von Beidem — DW-MP2 und expliziter Korrelation (DW-MP2-F12) — die genaue Beschreibung der schwachen Wechselwirkungen mit (sehr) annehmbarem Rechenaufwand möglich ist. Die DW-MP2-F12 Methode verursacht keinerlei zusätzliche Kosten im Vergleich zu MP2-F12 und reduziert damit die Rechenkosten drastisch verglichen mit der CCSD(T*)-F12 Methode. Die DW-MP2-F12 Methode kann voll automatisch verwendet werden (Black-Box) und macht damit genaue quantenmechanische *ab initio* Rechnungen für schwach wechselwirkende Systeme mit 100 Atomen oder mehr möglich.

Die grundsätzliche Idee von DW-MP2-F12 wurde in (der Gemeinde) der Quantenchemie anerkannt was an den mehr als 20 Zitaten zu der aus dieser Arbeit resultierenden Veröffentlichung zu sehen ist. Außerdem wurde das DW-MP2-F12 zugrundeliegende Konzept bei der Entwicklung von Dispersion-Weighted explicitly correlated Coupled Cluster (DW-CCSD(T**)-F12) von David Sherrill et al. (siehe **Section 7.6.2**) angewendet und das gleiche Optimierungsprinzip transferiert um CCSD(T*)-F12 zu optimieren: MP2 und SCS-MP2 werden hier durch CCSD(T*)-F12a und CCSD(T*)-F12b ersetzt um noch weiter optimierte Ergebnisse zu erhalten.

11 Appendix

11.1 Settings for the geometry optimization

11.1.1 Pre-optimization

Settings for the geometry optimization of the all degrees of freedom in the dimer (intermolecular) with fixed monomers (freezing intramolecular degrees of freedom) using MP2/AVDZ analytical gradients with counterpoise correction:

```
{optg,MAXIT=300,gradient=1.d-5,startcmd=label:  !find next energy
hessian,numerical=4;
active,R1,AN1,AN2,DI1,DI2,DI3}
```

For cases where linear monomers are involved, the number of degrees of freedom for the geometry optimization decreased. The accuracy for the geometry optimization is thrgrad=1.d-7

11.1.2 Intermolecular distance geometry optimization

Optimization of the intermolecular distance with counterpoise correction using numerical gradients, freezing intramolecular and the other intermolecular degrees of freedom:

```
{optg,MAXIT=300,proc=energies_method,variable=Opt_var,astep=2,rstep=0.02
numdia=1,gradient=thrgrad;hessian,model=vdw;method,diis
active,R1}
```

`Opt_var` is the counterpoise corrected interaction energy of the method used in the procedure `energies_method` in the geometry optimization. The accuracy for the geometry optimization is thrgrad=1.d-7 .

.

11.2 Geometries (T21 test set)

(CpH)(Ethene)

```
1 C1 -1.405928113 -1.619908630 -2.589198612
2 C2 -2.211227149 0.872481099 -1.520931069
3 C3 0.208882283 2.325071815 -1.335816488
4 C4 2.127911086 0.861289411 -2.182795371
5 C5 1.124556638 -1.590256556 -2.961032183
6 H1 -2.675287116 -3.167452383 -2.984121596
7 H2 2.242342436 -3.124045017 -3.713253863
8 H3 4.094458469 1.401343173 -2.276520152
9 H4 0.354433039 4.235404817 -0.634739302
10 H5 -3.118377820 0.656691099 0.326280630
11 H6 -3.583030364 1.822789209 -2.744380033
12 C6 0.394385429 -1.538907345 4.413666753
13 C7 -0.067944806 0.681230272 5.518122109
14 H7 0.887527843 -1.666254686 2.431653824
15 H8 0.032846069 2.437978639 4.473393238
16 H9 0.292249700 -3.292126953 5.462659321
17 H10 -0.560518809 0.802918760 7.499828160
```

$(CpH)_2$

```
1 C1 0.000000000 0.000000000 0.000000000
2 C2 0.000000000 0.000000000 2.828729156
3 C3 0.000000000 2.750371327 3.489542611
4 C4 -0.000590787 4.126859829 1.333624535
5 C5 0.000038265 2.417806070 -0.834689364
6 H1 0.000440929 -1.684222918 -1.151312762
7 H2 -0.000530330 3.010085948 -2.788402192
8 H3 -0.001543542 6.164871885 1.214222009
9 H4 -0.002057553 3.476494446 5.396059699
10 H5 1.658580179 -0.971836015 3.594447953
11 H6 -1.658744584 -0.971520931 3.594542721
12 C6 6.442243058 -2.074460280 4.416278670
13 C7 6.318273282 -2.336729932 1.602463686
14 C8 6.122988474 0.336158701 0.697617256
15 C9 6.134820239 1.904454528 2.718263664
16 C10 6.332572258 0.406479210 5.028980955
17 H7 6.594110735 -3.642033277 5.713112976
18 H8 6.382944011 1.176440612 6.919063418
19 H9 6.017757525 3.941575895 2.653580852
20 H10 5.997575501 0.881351271 -1.264295859
21 H11 4.689537568 -3.467061142 1.009800160
22 H12 7.997574678 -3.280287837 0.846552162
```

(PhH)(CpH) $\perp$

```
1 C1 0.000000000 0.000000000 0.000000000
2 C2 0.000000000 0.000000000 2.635485762
3 C3 0.000000000 2.282064968 3.953302129
4 C4 -0.000119490 4.564451869 2.636043418
5 C5 -0.000119507 4.565095843 0.000557734
6 C6 -0.000039842 2.282731981 -1.317749968
7 H1 0.004539674 -1.769575379 3.656858982
8 H2 0.005311836 2.281819337 5.996444426
9 H3 0.004327541 6.333777860 3.657848999
10 H4 0.003339262 6.334744228 -1.020821760
11 H5 0.002616089 2.282980714 -3.361031614
12 H6 0.003551415 -1.769398577 -1.021811873
13 C7 -6.044190096 1.047604208 0.543523639
14 C8 -6.214250003 -1.661416956 1.339760162
15 C9 -6.198503431 -1.522046451 4.164922948
16 C10 -6.041718510 0.929148934 4.878900098
17 C11 -5.945205156 2.525813409 2.628611663
18 H7 -6.008504585 1.677289312 -1.396669991
19 H8 -5.816371611 4.563177658 2.647008020
20 H9 -5.994385104 1.615769648 6.800894519
21 H10 -6.300528814 -3.143879017 5.398339494
22 H11 -4.624643505 -2.772689367 0.618716797
23 H12 -7.935291266 -2.562488653 0.627128630
```

(PhH)(CpH) $\parallel$

```
1 C1 0.000000000 0.000000000 0.000000000
2 C2 0.000000000 0.000000000 2.635485762
3 C3 0.000000000 2.282064968 3.953302129
4 C4 -0.000119490 4.564451869 2.636043418
5 C5 -0.000119507 4.565095843 0.000557734
6 C6 -0.000039842 2.282731981 -1.317749968
7 H1 0.004539674 -1.769575379 3.656858982
8 H2 0.005311836 2.281819337 5.996444426
9 H3 0.004327541 6.333777860 3.657848999
10 H4 0.003339262 6.334744228 -1.020821760
11 H5 0.002616089 2.282980714 -3.361031614
12 H6 0.003551415 -1.769398577 -1.021811873
13 C7 -6.044190096 1.047604208 0.543523639
14 C8 -6.214250003 -1.661416956 1.339760162
15 C9 -6.198503431 -1.522046451 4.164922948
16 C10 -6.041718510 0.929148934 4.878900098
17 C11 -5.945205156 2.525813409 2.628611663
18 H7 -6.008504585 1.677289312 -1.396669991
19 H8 -5.816371611 4.563177658 2.647008020
20 H9 -5.994385104 1.615769648 6.800894519
21 H10 -6.300528814 -3.143879017 5.398339494
22 H11 -4.624643505 -2.772689367 0.618716797
23 H12 -7.935291266 -2.562488653 0.627128630
```

(Butadien)$_2$

```
1 C1 1.978153737 -3.132783793 3.475786516
2 C2 -0.114934166 -1.431844184 2.992101365
3 C3 -0.082380589 1.107727633 4.020536634
4 C4 -2.175468492 2.808667242 3.536851483
5 H1 -1.711364216 -2.061685890 1.856446787
6 H2 1.514049462 1.737569339 5.156191212
7 H3 1.619845834 -4.939650757 2.558263752
8 H4 3.712760916 -2.315879381 2.728253295
9 H5 -3.581752441 1.853433869 2.377155205
10 H6 -3.027779744 3.372812998 5.322994225
11 C5 -1.733122677 -2.908442720 -3.715454679
12 C6 -0.949166577 -0.315017238 -4.124992172
13 C7 1.233940279 0.681453891 -2.802398662
14 C8 2.017896379 3.274879373 -3.211936154
15 H7 -1.999977042 0.884350547 -5.425882283
16 H8 2.284750744 -0.517913894 -1.501508551
17 H9 -3.393849607 -3.312201587 -4.861738082
18 H10 -0.205312963 -4.179406747 -4.249728690
19 H11 0.749683882 4.169607672 -4.563268510
20 H12 1.954556378 4.300358684 -1.428854265
```

(Ethane)$_2$

```
1 C1 0.000000000 0.000000000 0.000000000
2 C2 0.000000000 0.000000000 2.880422615
3 H1 0.000000000 1.919003708 3.624716567
4 H2 -1.661909137 -0.959486773 3.624728917
5 H3 1.661891312 -0.959517647 3.624728917
6 H4 0.000000000 -1.919003708 -0.744293953
7 H5 -1.661905961 0.959501854 -0.744293953
8 H6 1.661905961 0.959501854 -0.744293953
9 C3 6.825658167 -1.534933412 1.340126375
10 C4 6.856114834 1.345286437 1.355632823
11 H7 5.233937954 2.101313128 2.372251991
12 H8 8.555828494 2.066775740 2.265014690
13 H9 6.802198102 2.100519632 -0.558332832
14 H10 8.447835047 -2.290960102 0.323507206
15 H11 6.879596174 -2.290154504 3.254096206
16 H12 5.125933503 -2.256410305 0.430755231
```

(Ethine)(HF)

```
1 C1 0.000000000 0.000000000 0.000000000
2 C2 0.000000000 0.000000000 2.281396439
3 H1 0.000000000 0.000000000 4.295632845
4 H2 -0.000000000 0.000000000 -2.009583901
5 F1 -1.121892080 5.715960147 1.140473741
6 H3 -0.788121535 4.015530471 1.138173354
```

(PhH)(HF)

```
1 C1 -2.558938367 -0.780569803 0.203315440
2 C2 -2.023780047 1.789661609 0.434186106
3 C3 0.287177415 2.570173340 1.431551407
4 C4 2.063395558 0.780782974 2.198096724
5 C5 1.528813875 -1.789590775 1.967474918
6 C6 -0.782542555 -2.570515572 0.970024209
7 H1 -3.402659890 3.177085147 -0.156040282
8 H2 0.699730419 4.562782160 1.615314379
9 H3 3.853376240 1.386004819 2.975461863
10 H4 2.904588329 -3.177036290 2.565072637
11 H5 -1.198278406 -4.563260691 0.793565731
12 H6 -4.352154807 -1.385781444 -0.566734631
13 F1 1.005546933 0.000145445 -4.767993936
14 H7 0.248347911 -0.001841309 -3.209303830
```

(CpH)(EtOH)

```
1 C1 0.000000000 0.000000000 0.000000000
2 C2 0.000000000 0.000000000 2.828729156
3 C3 0.000000000 2.750371327 3.489542611
4 C4 -0.000590787 4.126859829 1.333624535
5 C5 0.000038265 2.417806070 -0.834689364
6 H1 0.000440929 -1.684222918 -1.151312762
7 H2 -0.000530330 3.010085948 -2.788402192
8 H3 -0.001543542 6.164871885 1.214222009
9 H4 -0.002057553 3.476494446 5.396059699
10 H5 1.658580179 -0.971836015 3.594447953
11 H6 -1.658744584 -0.971520931 3.594542721
12 C6 6.687065216 -3.828988544 0.953850066
13 C7 6.490067173 -1.373721713 0.440440506
14 O1 6.373215384 0.406292383 2.312569949
15 H7 6.236393698 2.052304106 1.557485076
16 H8 6.412976605 -0.668591260 -1.480055452
17 H9 6.772727499 -5.158735165 -0.585676257
18 H10 6.761579956 -4.503282131 2.877596744
```

(HF)$_2$

```
1 F1 0.000000000 0.018352652 -2.680645054
2 H1 0.000000000 -0.221873963 -0.955266955
3 F2 0.000000000 -0.087027014 2.559401066
4 H2 0.000000000 1.516299427 3.240563847
```

$(H_2O)(NH_3)$

```
1 O1 0.000000000 0.000000000 0.000000000
2 H1 0.000000000 0.000000000 1.812258698
3 H2 0.000000000 1.772560336 -0.403691961
4 N1 -0.012399513 5.254082260 -1.825554960
5 H3 -0.718436366 6.720670987 -0.819774664
6 H4 1.745872205 5.777714616 -2.368927993
7 H5 -1.064825212 5.126726997 -3.422175021
```

$(MeOH)_2$

```
1 C1 0.000000000 0.000000000 0.000000000
2 O1 0.000000000 0.000000000 2.697081385
3 H1 0.000000000 1.718593052 3.271681018
4 H2 1.674168288 -0.983014413 -0.682162918
5 H3 -1.674168288 -0.983014413 -0.682162918
6 H4 0.000000000 1.907254000 -0.773951492
7 C2 -3.381032492 -5.433481243 4.545652154
8 O2 -4.256583653 -2.941391497 4.000540363
9 H5 -2.817824225 -1.871633819 3.737259834
10 H6 -4.062869578 -6.005468297 6.401031370
11 H7 -4.115589275 -6.738533477 3.134352032
12 H8 -1.326075490 -5.550627209 4.538776339
```

$(Furane)_2$

```
1 C1 0.000000000 0.000000000 0.000000000
2 O1 0.000000000 0.000000000 2.570105018
3 C2 0.000000000 2.459862953 3.314983747
4 C3 0.000000000 4.022456390 1.262746592
5 C4 0.000000000 2.416978244 -0.900845512
6 H1 0.000000000 -1.828864756 -0.884422821
7 H2 -0.000000000 2.964671139 -2.859161664
8 H3 -0.000000000 6.055455640 1.306148318
9 H4 -0.000000000 2.776220289 5.321678041
10 C5 -8.036242363 -6.581486206 4.245155070
11 O2 -8.818289390 -5.407466680 6.393531295
12 C6 -7.058127079 -3.618486015 6.947739638
13 C7 -5.171561423 -3.635664480 5.188776728
14 C8 -5.809946929 -5.569526216 3.424879453
15 H5 -9.244289400 -8.062587814 3.556743563
16 H6 -4.771690003 -6.141521242 1.772664605
17 H7 -3.542725056 -2.418516082 5.168490546
18 H8 -7.413215926 -2.515514800 8.616352944
```

(Butadien)(Ethene)

```
1 C1 0.000000000 0.000000000 0.000000000
2 C2 0.000000000 0.000000000 2.740102891
3 C3 0.000000000 2.372998712 4.110154336
4 C4 -0.000000000 2.372998712 6.850257226
5 H1 -0.000000000 -1.782203860 3.769058769
6 H2 0.000000000 4.155202573 3.081198457
7 H3 0.000000000 -1.940220453 -0.685963114
8 H4 -1.680280201 0.970110226 -0.685963114
9 H5 -0.000000000 0.432778259 7.536220341
10 H6 1.680280201 3.343108939 7.536220341
11 C5 7.071713625 -2.574491436 5.001077952
12 C6 7.677608235 -1.377533560 2.865014872
13 H7 5.185083761 -2.472171528 5.787169822
14 H8 6.305164260 -0.259428334 1.838359977
15 H9 8.447541563 -3.690917695 6.023470247
16 H10 9.565114684 -1.483127881 2.083500087
```

(MeOH)(Ethanediamine)

```
1 C1 0.000000000 0.000000000 0.000000000
2 C2 0.000000000 0.000000000 2.866491554
3 N1 0.000000000 2.607269069 3.767352367
4 N2 -2.345608795 1.161106426 -0.914500568
5 H1 1.702422733 -0.945106015 3.544148533
6 H2 -2.227519414 1.554266074 -2.782602790
7 H3 0.290899777 -1.929634813 -0.689546193
8 H4 -1.627710988 -1.111723435 3.523079742
9 H5 -0.267693996 2.679440923 5.659721014
10 H6 1.574231664 1.158631612 -0.650185573
11 H7 -3.808442148 -0.057234662 -0.703141697
12 H8 -1.468579470 3.523429489 2.943880734
13 C3 -7.738031373 -3.062322259 -0.004131938
14 O1 -6.356177511 -0.783403778 0.409701917
15 H9 -5.544131740 -0.889309985 2.026208136
16 H10 -9.730028754 -2.615003805 -0.261727886
17 H11 -7.037226099 -3.985869475 -1.704260035
18 H12 -7.560092266 -4.372620293 1.573231517
```

$(Ethine)_2$

```
1 C1 0.000000000 0.000000000 0.000000000
2 C2 0.000000000 0.000000000 2.281396439
3 H1 0.000000000 0.000000000 4.295632845
4 H2 -0.000000000 0.000000000 -2.009583901
5 C3 -1.166808685 9.397594363 1.126598393
6 C4 -0.888985230 7.160805287 1.479243354
7 H3 -0.643695937 5.185952555 1.790592274
8 H4 -1.411531405 11.367885559 0.815968630
```

$(Pyrrole)_2$

```
1 C1 0.000000000 0.000000000 0.000000000
2 C2 0.000000000 0.000000000 2.613362738
3 C3 0.000000000 2.553945160 3.416805698
4 C4 0.000000000 4.050210637 1.274191603
5 N1 0.000000000 2.469096154 -0.774042345
6 H1 -0.000000000 6.070911937 1.046533655
7 H2 -0.000000000 3.236408605 5.333809203
8 H3 0.000000000 -1.657126318 3.794245243
9 H4 0.000000000 -1.526345602 -1.343714717
10 H5 -0.000000000 3.038931872 -2.585850094
11 C5 6.635048012 -0.953802847 2.323230171
12 C6 6.331241569 1.457739108 1.363109424
13 C7 8.765479639 2.553461895 1.158089762
14 C8 10.495414785 0.783901829 1.998082085
15 N2 9.168695523 -1.325513443 2.694766030
16 H6 12.521768139 0.854171299 2.153053599
17 H7 9.218059759 4.417104969 0.477895288
18 H8 4.553904876 2.317521357 0.870768192
19 H9 5.280632329 -2.405507948 2.763014882
20 H10 9.943286618 -2.918344328 3.380519973
```

$(Ethene)(H_2O)$

```
1 C1 0.000000000 0.000000000 0.000000000
2 C2 0.000000000 0.000000000 2.522415889
3 H1 0.000000000 1.744195400 -1.070310443
4 H2 0.000000000 1.744195400 3.592726332
5 H3 0.000000000 -1.746475565 -1.065091177
6 H4 -0.000000000 -1.746475565 3.587507066
7 O1 -1.796451628 6.177494491 1.240446145
8 H5 -2.934654004 5.577944074 2.516892870
9 H6 -2.907290720 7.014919006 0.070106862
```

$(Propadien)_2$

```
1 C1 0.000000000 0.000000000 0.000000000
2 C2 0.000000000 0.000000000 2.490604693
3 C3 0.000000000 0.000000000 -2.490604693
4 H1 0.000000000 -1.774358936 -3.504932681
5 H2 0.000000000 1.774358981 -3.504932603
6 H3 0.000000000 -1.774358995 3.504932541
7 H4 0.000000000 1.774358930 3.504932693
8 C4 -8.574011146 0.003224721 -0.003505839
9 C5 -6.083406720 0.002287995 -0.004176869
10 C6 -11.064615572 0.004161447 -0.002834808
11 H5 -12.079455780 -0.088319645 -1.774488698
12 H6 -12.078431047 0.097405527 1.769365697
13 H7 -5.069591308 -0.090956086 -1.776377388
14 H8 -5.068566502 0.094769086 1.767477014
```

$(Purine)(CH_4)$

```
1 C1 0.023846762 0.017107286 -4.187516353
2 C2 0.008695967 1.384836918 -1.936499611
3 C3 -0.017501716 -0.062824246 0.295867521
4 N1 -0.028592519 -2.570791112 0.440801911
5 C4 -0.011874043 -3.648367625 -1.844162179
6 N2 0.013354225 -2.501910724 -4.119157655
7 N3 -0.028114231 1.675344439 2.216884491
8 C5 -0.008588626 4.004769281 1.091929250
9 N4 0.013835646 3.932442012 -1.399006268
10 H1 0.044119921 0.927186239 -6.020286731
11 H2 -0.019529105 -5.691474567 -1.875037016
12 H3 -0.011870755 5.715066058 2.197533431
13 H4 -0.045882200 1.284718703 4.080069007
14 C6 0.031936365 -1.875839217 7.522395718
15 H5 0.050095282 -3.217846098 9.073045630
16 H6 -1.652588537 -0.709610489 7.684956585
17 H7 1.715991296 -0.704717856 7.647601208
18 H8 0.013881895 -2.890318202 5.736946934
```

$(Purine)(CO_2)$

```
1 N1 -0.004930384 -2.443333373 -4.793866404
2 C1 -0.002108370 -0.137232609 -3.596043571
3 C2 0.003472613 -0.437952189 -0.956340079
4 N2 0.004026447 -2.993829001 -0.572676823
5 C3 -0.001108055 -4.077904356 -2.916843053
6 N3 0.007019200 1.410001509 0.740658427
7 C4 0.004846004 3.691978179 -0.334307515
8 N4 -0.000252929 4.245523475 -2.815507115
9 C5 -0.003814295 2.345250931 -4.464919326
10 H1 -0.008051211 2.804313362 -6.454605751
11 H2 0.007410073 5.284196963 0.939816566
12 H3 -0.001913233 -6.096876740 -3.147414873
13 H4 0.007096763 -3.855184568 1.124278111
14 C6 -0.002237503 -0.385493107 5.857612115
15 O1 0.002440768 -2.450055781 5.063518252
16 O2 -0.007146238 1.626591195 6.738657931
```

11.3 Geometries (geometry optimization)

(PhH)(Ethene)

```
1 c1
2 c2 1 cc2
3 c3 2 cc3 1 ccc3
4 c4 3 cc4 2 ccc4 1 dih4
5 c5 4 cc5 3 ccc5 2 dih5
6 c6 5 cc6 4 ccc6 3 dih6
7 h1 2 hc7 3 hcc7 4 dih7
8 h2 3 hc8 2 hcc8 1 dih8
9 h3 4 hc9 3 hcc9 2 dih9
10 h4 5 hc10 4 hcc10 3 dih10
11 h5 6 hc11 5 hcc11 4 dih11
12 h6 1 hc12 2 hcc12 3 dih12
13 c7 2 r1 1 an1 6 di1
14 c8 13 bcc2 2 an2 1 di2
15 h7 13 bhc3 14 bhcc3 2 di3
16 h8 14 bhc4 13 bhcc4 15 bdih4
17 h9 13 bhc5 14 bhcc5 16 bdih5
18 h10 14 bhc6 13 bhcc6 15 bdih6

cc2 = 1.394639; cc3 = 1.394506; ccc3 = 120.005;
cc4 = 1.394506; ccc4 = 120.004; dih4 = 0.003;
cc5 = 1.394639; ccc5 = 120.005; dih5 = -0.003;
cc6 = 1.394773; ccc6 = 119.997; dih6 = 0.001;
hc7 = 1.081209; hcc7 = 120.002; dih7 = 179.856;
hc8 = 1.081188; hcc8 = 119.998; dih8 = -179.828;
hc9 = 1.081209; hcc9 = 120.002; dih9 = -179.856;
hc10 = 1.081243; hcc10 = 120.006; dih10 = -179.885;
hc11 = 1.081259; hcc11 = 120.004; dih11 = -179.914;
hc12 = 1.081243; hcc12 = 120.006:  dih12 = 179.885;
bcc2 = 1.334805; bhc3 = 1.082912; bhcc3 = 121.535;
bhc4 = 1.082912; bhcc4 = 121.535; bdih4 = 0.000;
bhc5 = 1.082500; bhcc5 = 121.377; bdih5 = 180.000;
bhc6 = 1.082500; bhcc6 = 121.377; bdih6 = 180.000;
R1 = 3.98883587; AN1 = 74.37001214;
AN2 = 101.57671544; DI1 = 74.46311281;
DI2 = 31.18661278; DI3 = 21.44947797;
```

(PhH)($HCONH_2$)

```
1 c1
2 c2 1 cc2
3 c3 2 cc3 1 ccc3
4 c4 3 cc4 2 ccc4 1 dih4
5 c5 4 cc5 3 ccc5 2 dih5
6 c6 5 cc6 4 ccc6 3 dih6
7 h1 2 hc7 3 hcc7 4 dih7
8 h2 3 hc8 2 hcc8 1 dih8
9 h3 4 hc9 3 hcc9 2 dih9
10 h4 5 hc10 4 hcc10 3 dih10
11 h5 6 hc11 5 hcc11 4 dih11
12 h6 1 hc12 2 hcc12 3 dih12
13 c7 1 r1 2 an1 3 di1
14 n1 13 Bnc2 1 an2 2 di2
15 o1 13 Boc3 14 Bocn3 3 di3
16 h7 14 Bhn4 13 Bhnc4 15 Bdih4
17 h8 14 Bhn5 13 Bhnc5 15 Bdih5
18 h9 13 Bhc6 14 Bhcn6 16 Bdih6

Bnc2 = 1.342414; Boc3 = 1.229066; Bocn3 = 125.487;
Bhn4 = 1.003261; Bhnc4 = 119.219; Bdih4 = 180.000;
Bhn5 = 1.022613; Bhnc5 = 120.730; Bdih5 = 0.000;
Bhc6 = 1.100428; Bhcn6 = 113.600; Bdih6 = 0.000;
cc2 = 1.394639; cc3 = 1.394506; ccc3 = 120.005;
cc4 = 1.394506; ccc4 = 120.004; dih4 = 0.003;
cc5 = 1.394639; ccc5 = 120.005; dih5 = -0.003;
cc6 = 1.394773; ccc6 = 119.997; dih6 = 0.001;
hc7 = 1.081209; hcc7 = 120.002; dih7 = 179.856;
hc8 = 1.081188; hcc8 = 119.998; dih8 = -179.828;
hc9 = 1.081209; hcc9 = 120.002; dih9 = -179.856;
hc10 = 1.081243; hcc10 = 120.006; dih10 = -179.885;
hc11 = 1.081259; hcc11 = 120.004; dih11 = -179.914;
hc12 = 1.081243; hcc12 = 120.006; dih12 = 179.885;
R1 = 6.98030680; AN1 = 36.77772466;
AN2 = 125.61482765; DI1 = 0.04913351;
DI2 = -0.23116205; DI3 = -0.21826302;
```

(PhH)(HCOOH)

```
1 c1
2 c2 1 cc2
3 c3 2 cc3 1 ccc3
4 c4 3 cc4 2 ccc4 1 dih4
5 c5 4 cc5 3 ccc5 2 dih5
6 c6 5 cc6 4 ccc6 3 dih6
7 h1 2 hc7 3 hcc7 4 dih7
8 h2 3 hc8 2 hcc8 1 dih8
9 h3 4 hc9 3 hcc9 2 dih9
10 h4 5 hc10 4 hcc10 3 dih10
11 h5 6 hc11 5 hcc11 4 dih11
12 h6 1 hc12 2 hcc12 3 dih12
13 c7 1 r1 2 an1 3 di1
14 o1 13 Aoc2 1 an2 2 di2
15 o2 13 Aoc3 14 Aoco3 1 di3
16 h8 13 Ahc4 14 Ahco4 15 Adih4
17 h9 15 Aho5 13 Ahoc5 14 Adih5

Aoc2 = 1.220719; Aoc3 = 1.314335; Aoco3 = 126.428;
Ahc4 = 1.093459; Ahco4 = 121.904; Adih4 = 180.000;
Aho5 = 0.995011; Ahoc5 = 109.447; Adih5 = 0.000;
cc2 = 1.394639; cc3 = 1.394506; ccc3 = 120.005;
cc4 = 1.394506; ccc4 = 120.004; dih4 = 0.003;
cc5 = 1.394639; ccc5 = 120.005; dih5 = -0.003;
cc6 = 1.394773; ccc6 = 119.997; dih6 = 0.001;
hc7 = 1.081209; hcc7 = 120.002; dih7 = 179.856;
hc8 = 1.081188; hcc8 = 119.998; dih8 = -179.828;
hc9 = 1.081209; hcc9 = 120.002; dih9 = -179.856;
hc10 = 1.081243; hcc10 = 120.006; dih10 = -179.885;
hc11 = 1.081259; hcc11 = 120.004; dih11 = -179.914;
hc12 = 1.081243; hcc12 = 120.006; dih12 = 179.885;
R1 = 3.50260767; AN1 = 84.77557603;
AN2 = 102.61503038; DI1 = 75.73374811;
DI2 = 180.00000000; DI3 = 121.57079296;
```

$(Butadiene)_2$

```
1 c1
2 c2 1 cc2
3 c3 2 cc3 1 ccc3
4 c4 3 cc4 2 ccc4 1 dih4
5 h1 2 hc5 3 hcc5 4 dih5
6 h2 3 hc6 2 hcc6 1 dih6
7 h3 1 hc7 2 hcc7 3 dih7
8 h4 1 hc8 2 hcc8 3 dih8
9 h5 4 hc9 3 hcc9 2 dih9
10 h6 4 hc10 3 hcc10 2 dih10
11 c5 7 r1 1 an1 6 di1
12 c6 11 cc2 7 an2 1 di2
13 c7 12 cc3 11 ccc3 1 di3
14 c8 13 cc4 12 ccc4 11 dih4
15 h7 12 hc5 13 hcc5 14 dih5
16 h8 13 hc6 12 hcc6 11 dih6
17 h9 11 hc7 12 hcc7 13 dih7
18 h10 11 hc8 12 hcc8 13 dih8
19 h11 14 hc9 13 hcc9 12 dih9
20 h12 14 hc10 13 hcc10 12 dih10

cc2 = 1.341151; cc3 = 1.452897; ccc3 = 123.567;
cc4 = 1.341165; ccc4 = 123.604; dih4 = 179.845;
hc5 = 1.085421; hcc5 = 116.889; dih5 = -0.172;
hc6 = 1.085497; hcc6 = 116.872; dih6 = -0.241;
hc7 = 1.080357; hcc7 = 121.430; dih7 = -179.997;
hc8 = 1.082687; hcc8 = 120.846; dih8 = -0.011;
hc9 = 1.082539; hcc9 = 120.880; dih9 = -0.097;
hc10 = 1.080420; hcc10 = 121.390; dih10 = 179.959;
R1 = 6.58704641; AN1 = 351.43124596;
AN2 = 968.58017781; DI1 = 25.74587861;
DI2 = -898.76335069; DI3 = 93.19887069;
```

(CpH)(HCN)

```
1 c1
2 c2 1 cc2
3 c3 2 cc3 1 ccc3
4 c4 3 cc4 2 ccc4 1 dih4
5 c5 4 cc5 3 ccc5 2 dih5
6 h1 1 hc6 2 hcc6 3 dih6
7 h2 5 hc7 4 hcc7 3 dih7
8 h3 4 hc8 3 hcc8 2 dih8
9 h4 3 hc9 2 hcc9 1 dih9
10 h5 2 hc10 3 hcc10 4 dih10
11 h6 2 hc11 3 hcc11 4 dih11
12 c6 1 r1 2 an1 3 di1
13 n1 12 Anc2 1 an2 2 di2
14 h7 12 Ahc3 2 Ahcn3 1 di3

Anc2 = 1.167338; Ahc3 = 1.067879; Ahcn3 = 179.972;
cc2 = 1.496899; cc3 = 1.496853; ccc3 = 103.510;
cc4 = 1.353567; ccc4 = 109.047; dih4 = 0.014;
cc5 = 1.460994; ccc5 = 109.198; dih5 = -0.023;
hc6 = 1.079590; hcc6 = 124.356; dih6 = 179.985;
hc7 = 1.080324; hcc7 = 124.890; dih7 = 179.990;
hc8 = 1.080319; hcc8 = 125.910; dih8 = 179.972;
hc9 = 1.079582; hcc9 = 124.360; dih9 = 179.930;
hc10 = 1.094985; hcc10 = 111.729; dih10 = 120.375;
hc11 = 1.094995; hcc11 = 111.719; dih11 = -120.345;
R1 = 6.21741443; AN1 = -2.10232237;
AN2 = -0.01132568; DI1 = 0.00000000;
DI2 = 180.00000000; DI3 = 41.36382808;
```

(CpH)(Ethene)

```
1 c1
2 c2 1 cc2
3 c3 2 cc3 1 ccc3
4 c4 3 cc4 2 ccc4 1 dih4
5 c5 4 cc5 3 ccc5 2 dih5
6 h1 1 hc6 2 hcc6 3 dih6
7 h2 5 hc7 4 hcc7 3 dih7
8 h3 4 hc8 3 hcc8 2 dih8
9 h4 3 hc9 2 hcc9 1 dih9
10 h5 2 hc10 3 hcc10 4 dih10
11 h6 2 hc11 3 hcc11 4 dih11
12 c6 2 r1 1 an1 6 di1
13 c7 12 bcc2 2 an2 1 di2
14 h7 12 bhc3 13 bhcc3 2 di3
15 h8 13 bhc4 12 bhcc4 14 bdih4
16 h9 12 bhc5 13 bhcc5 15 bdih5
17 h10 13 bhc6 12 bhcc6 14 bdih6

cc2 = 1.496899; cc3 = 1.496853; ccc3 = 103.510;
cc4 = 1.353567; ccc4 = 109.047; dih4 = 0.014;
cc5 = 1.460994; ccc5 = 109.198; dih5 = -0.023;
hc6 = 1.079590; hcc6 = 124.356; dih6 = 179.985;
hc7 = 1.080324; hcc7 = 124.890; dih7 = 179.990;
hc8 = 1.080319; hcc8 = 125.910; dih8 = 179.972;
hc9 = 1.079582; hcc9 = 124.360; dih9 = 179.930;
hc10 = 1.094985; hcc10 = 111.729; dih10 = 120.375;
hc11 = 1.094995; hcc11 = 111.719; dih11 = -120.345;
bcc2 = 1.334805; bhc3 = 1.082912; bhcc3 = 121.535;
bhc4 = 1.082912; bhcc4 = 121.535; bdih4 = 0.000;
bhc5 = 1.082500; bhcc5 = 121.377; bdih5 = 180.000;
bhc6 = 1.082500; bhcc6 = 121.377; bdih6 = 180.000;
R1 = 3.76078024; AN1 = 84.66568445;
AN2 = 90.59215555; DI1 = 102.40772782;
DI2 = 170.96825966; DI3 = 36.83454052;
```

(CpH)(Ethenole)

```
1 c1
2 c2 1 cc2
3 c3 2 cc3 1 ccc3
4 c4 3 cc4 2 ccc4 1 dih4
5 c5 4 cc5 3 ccc5 2 dih5
6 h1 1 hc6 2 hcc6 3 dih6
7 h2 5 hc7 4 hcc7 3 dih7
8 h3 4 hc8 3 hcc8 2 dih8
9 h4 3 hc9 2 hcc9 1 dih9
10 h5 2 hc10 3 hcc10 4 dih10
11 h6 2 hc11 3 hcc11 4 dih11
12 c6 2 r1 1 an1 6 di1
13 c7 12 acc2 2 an2 1 di2
14 o1 13 aoc3 12 aocc3 2 di3
15 h7 14 aho4 13 ahoc4 12 adih4
16 h8 13 ahc5 14 ahco5 15 adih5
17 h9 12 ahc6 13 ahcc6 14 adih6
18 h10 12 ahc7 13 ahcc7 14 adih7

acc2 = 1.331460; aoc3 = 1.368409; aocc3 = 121.838;
aho4 = 0.961040; ahoc4 = 109.043; adih4 = 180.000;
ahc5 = 1.083387; ahco5 = 116.126; adih5 = 0.000;
ahc6 = 1.077458; ahcc6 = 119.098; adih6 = 180.000;
ahc7 = 1.079447; ahcc7 = 121.189; adih7 = 0.007;
cc2 = 1.496899; cc3 = 1.496853; ccc3 = 103.510;
cc4 = 1.353567; ccc4 = 109.047; dih4 = 0.014;
cc5 = 1.460994; ccc5 = 109.198; dih5 = -0.023;
hc6 = 1.079590; hcc6 = 124.356; dih6 = 179.985;
hc7 = 1.080324; hcc7 = 124.890; dih7 = 179.990;
hc8 = 1.080319; hcc8 = 125.910; dih8 = 179.972;
hc9 = 1.079582; hcc9 = 124.360; dih9 = 179.930;
hc10 = 1.094985; hcc10 = 111.729; dih10 = 120.375;
hc11 = 1.094995; hcc11 = 111.719; dih11 = -120.345;
R1 = 4.13280950; AN1 = 75.92334912;
AN2 = 61.76195852; DI1 = 55.88134304;
DI2 = 72.16703709; DI3 = 65.51815595;
```

(CpH)(Furan-2-one)

```
1 c1
2 c2 1 cc2
3 c3 2 cc3 1 ccc3
4 c4 3 cc4 2 ccc4 1 dih4
5 c5 4 cc5 3 ccc5 2 dih5
6 h1 1 hc6 2 hcc6 3 dih6
7 h2 5 hc7 4 hcc7 3 dih7
8 h3 4 hc8 3 hcc8 2 dih8
9 h4 3 hc9 2 hcc9 1 dih9
10 h5 2 hc10 3 hcc10 4 dih10
11 h6 2 hc11 3 hcc11 4 dih11
12 c6 1 r1 2 an1 3 di1
13 c7 12 Acc2 1 an2 2 di2
14 o1 13 Aoc3 12 Aocc3 1 di3
15 c8 14 Aco4 13 Acoc4 12 Adih4
16 c9 15 Acc5 14 Acco5 13 Adih5
17 o2 13 Aoc6 14 Aoco6 15 Adih6
18 h7 12 Ahc7 13 Ahcc7 14 Adih7
19 h8 15 Ahc8 14 Ahco8 13 Adih8
20 h9 16 Ahc9 15 Ahcc9 14 Adih9
21 h10 15 Ahc10 14 Ahco10 13 Adih10

cc2 = 1.496899; cc3 = 1.496853; ccc3 = 103.510;
cc4 = 1.353567; ccc4 = 109.047; dih4 = 0.014;
cc5 = 1.460994; ccc5 = 109.198; dih5 = -0.023;
hc6 = 1.079590; hcc6 = 124.356; dih6 = 179.985;
hc7 = 1.080324; hcc7 = 124.890; dih7 = 179.990;
hc8 = 1.080319; hcc8 = 125.910; dih8 = 179.972;
hc9 = 1.079582; hcc9 = 124.360; dih9 = 179.930;
hc10 = 1.094985; hcc10 = 111.729; dih10 = 120.375;
hc11 = 1.094995; hcc11 = 111.719; dih11 = -120.345;
Acc2 = 1.488564; Aoc3 = 1.388872; Aocc3 = 107.691;
Aco4 = 1.441607; Acoc4 = 109.519; Adih4 = 0.001;
Acc5 = 1.504806; Acco5 = 105.327; Adih5 = -0.001;
Aoc6 = 1.214222; Aoco6 = 122.465; Adih6 = 179.999;
Ahc7 = 1.089116; Ahcc7 = 122.500; Adih7 = -180.000;
Ahc8 = 1.102176; Ahco8 = 108.486; Adih8 = -120.806;
Ahc9 = 1.091599; Ahcc9 = 123.591; Adih9 = 180.000;
Ahc10 = 1.102176; Ahco10 = 108.486;
R1 = 5.09764225; AN1 = 60.17391281;
AN2 = 50.85411454; DI1 = 53.46540122;
DI2 = 57.59393939; DI3 = 53.43671277;
```

$(CpH)_2$

```
1 c1
2 c2 1 cc2
3 c3 2 cc3 1 ccc3
4 c4 3 cc4 2 ccc4 1 dih4
5 c5 4 cc5 3 ccc5 2 dih5
6 h1 1 hc6 2 hcc6 3 dih6
7 h2 5 hc7 4 hcc7 3 dih7
8 h3 4 hc8 3 hcc8 2 dih8
9 h4 3 hc9 2 hcc9 1 dih9
10 h5 2 hc10 3 hcc10 4 dih10
11 h6 2 hc11 3 hcc11 4 dih11
12 c6 2 r1 1 an1 6 di1
13 c7 12 cc2 2 an2 1 di2
14 c8 13 cc3 12 ccc3 2 di3
15 c9 14 cc4 13 ccc4 12 dih4
16 c10 15 cc5 14 ccc5 13 dih5
17 h7 12 hc6 13 hcc6 14 dih6
18 h8 16 hc7 15 hcc7 14 dih7
19 h9 15 hc8 14 hcc8 13 dih8
20 h10 14 hc9 13 hcc9 12 dih9
21 h11 13 hc10 14 hcc10 15 dih10
22 h12 13 hc11 14 hcc11 15 dih11

cc2 = 1.496899; cc3 = 1.496853; ccc3 = 103.510;
cc4 = 1.353567; ccc4 = 109.047; dih4 = 0.014;
cc5 = 1.460994; ccc5 = 109.198; dih5 = -0.023;
hc6 = 1.079590; hcc6 = 124.356; dih6 = 179.985;
hc7 = 1.080324; hcc7 = 124.890; dih7 = 179.990;
hc8 = 1.080319; hcc8 = 125.910; dih8 = 179.972;
hc9 = 1.079582; hcc9 = 124.360; dih9 = 179.930;
hc10 = 1.094985; hcc10 = 111.729; dih10 = 120.375;
hc11 = 1.094995; hcc11 = 111.719; dih11 = -120.345;
R1 = 4.04336716; AN1 = 103.34079631;
AN2 = 74.71485122; DI1 = 69.68602207;
DI2 = -2.12394629; DI3 = 66.58645080;
```

(Ethandiol)(Ethandiamin)

```
1 c1
2 c2 1 acc2
3 n1 2 anc3 1 ancc3
4 n2 1 anc4 2 ancc4 3 adih4
5 h1 2 ahc5 1 ahcc5 4 adih5
6 h2 4 ahn6 1 ahnc6 2 adih6
7 h3 1 ahc7 2 ahcc7 3 adih7
8 h4 2 ahc8 1 ahcc8 4 adih8
9 h5 3 ahn9 2 ahnc9 1 adih9
10 h6 1 ahc10 2 ahcc10 3 adih10
11 h7 4 ahn11 1 ahnc11 2 adih11
12 h8 3 ahn12 2 ahnc12 1 adih12
13 c3 2 r1 1 an1 6 di1
14 c4 13 cc2 2 an2 1 di2
15 o1 14 oc3 13 occ3 2 di3
16 o2 13 oc4 14 occ4 15 dih4
17 h9 14 hc5 13 hcc5 16 dih5
18 h10 16 ho6 13 hoc6 14 dih6
19 h11 13 hc7 14 hcc7 15 dih7
20 h12 14 hc8 13 hcc8 16 dih8
21 h13 15 ho9 14 hoc9 13 dih9
22 h14 13 hc10 14 hcc10 15 dih10

acc2 = 1.516882; anc3 = 1.459743; ancc3 = 109.061;
anc4 = 1.467106; ancc4 = 109.260; adih4 = -63.664;
ahc5 = 1.091015; ahcc5 = 109.189; adih5 = 177.299;
ahn6 = 1.012144; ahnc6 = 110.856; adih6 = 166.912;
ahc7 = 1.095229; ahcc7 = 109.461; adih7 = 171.427;
ahc8 = 1.099426; ahcc8 = 108.423; adih8 = 60.669;
ahn9 = 1.012089; ahnc9 = 111.026; adih9 = 171.376;
ahc10 = 1.090076; ahcc10 = 108.399; adih10 = 53.647;
ahn11 = 1.013607; ahnc11 = 110.081;
adih11 = -75.393; ahn12 = 1.014335;
ahnc12 = 108.147; adih12 = 53.732;
cc2 = 1.505229; oc3 = 1.423528; occ3 = 108.311;
oc4 = 1.423459; occ4 = 108.318; dih4 = -74.878;
hc5 = 1.094290; hcc5 = 108.930; dih5 = 164.622;
ho6 = 0.961385; hoc6 = 108.453; dih6 = 162.629;
hc7 = 1.094307; hcc7 = 108.930; dih7 = 164.619;
hc8 = 1.092731; hcc8 = 108.636; dih8 = 46.164;
ho9 = 0.961496; hoc9 = 108.445; dih9 = 162.807;
hc10 = 1.092731; hcc10 = 108.631; dih10 = 46.170;
R1 = 4.33344685; AN1 = 65.71620314;
AN2 = 94.49496231; DI1 = 89.71854674;
DI2 = -117.33214155; DI3 = 18.09159044;
```

$(Ethane)_2$

```
1 c1
2 c2 1 cc2
3 h1 2 hc3 1 hcc3
4 h2 2 hc4 3 hch4 1 dih4
5 h3 2 hc5 4 hch5 3 dih5
6 h4 1 hc6 2 hcc6 3 dih6
7 h5 1 hc7 2 hcc7 3 dih7
8 h6 1 hc8 2 hcc8 3 dih8
9 c3 2 r1 1 an1 6 di1
10 c4 9 cc2 2 an2 1 di2
11 h7 10 hc3 9 hcc3 2 di3
12 h8 10 hc4 11 hch4 9 dih4
13 h9 10 hc5 12 hch5 11 dih5
14 h10 9 hc6 10 hcc6 11 dih6
15 h11 9 hc7 10 hcc7 11 dih7
16 h12 9 hc8 10 hcc8 11 dih8

cc2 = 1.524254; hc3 = 1.089199; hcc3 = 111.199;
hc4 = 1.089199; hch4 = 107.689; dih4 = -122.060;
hc5 = 1.089199; hch5 = 107.689; dih5 = 115.880;
hc6 = 1.089199; hcc6 = 111.199; dih6 = 180.000;
hc7 = 1.089199; hcc7 = 111.199; dih7 = -60.000;
hc8 = 1.089199; hcc8 = 111.199; dih8 = 60.000;
R1 = 3.92038786; AN1 = 78.24110349;
AN2 = 78.54837816; DI1 = 77.95498500;
DI2 = 87.93195487; DI3 = 20.27335624;
```

$(EtOH)_2$

```
1 c1
2 c2 1 cc2
3 o1 2 oc3 1 occ3
4 h1 3 ho4 2 hoc4 1 dih4
5 h2 2 hc5 3 hco5 4 dih5
6 h3 1 hc6 2 hcc6 3 dih6
7 h4 1 hc7 2 hcc7 3 dih7
8 h5 1 hc8 2 hcc8 3 dih8
9 h6 2 hc9 1 hcc9 6 dih9
10 c3 1 r1 2 an1 3 di1
11 c4 10 cc2 1 an2 2 di2
12 o2 11 oc3 10 occ3 1 di3
13 h7 12 ho4 11 hoc4 10 dih4
14 h8 11 hc5 12 hco5 13 dih5
15 h9 10 hc6 11 hcc6 12 dih6
16 h10 10 hc7 11 hcc7 12 dih7
17 h11 10 hc8 11 hcc8 12 dih8
18 h12 11 hc9 10 hcc9 15 dih9

cc2 = 1.510238; oc3 = 1.428745; occ3 = 107.364;
ho4 = 0.962098; hoc4 = 108.307; dih4 = -179.900;
hc5 = 1.093518; hco5 = 110.396; dih5 = -59.598;
hc6 = 1.089376; hcc6 = 110.393; dih6 = -179.574;
hc7 = 1.088457; hcc7 = 110.168; dih7 = -59.455;
hc8 = 1.088301; hcc8 = 110.216; dih8 = 60.248;
hc9 = 1.093753; hcc9 = 110.280; dih9 = -59.277;
R1 = 5.25737037; AN1 = 101.74146406;
AN2 = 31.57543541; DI1 = -30.69657920;
DI2 = 50.40960730; DI3 = -67.32738795;
```

$(Ethene)(NH_3)$

```
1 c1
2 c2 1 cc2
3 h1 1 hc3 2 hcc3
4 h2 2 hc4 1 hcc4 3 dih4
5 h3 1 hc5 2 hcc5 4 dih5
6 h4 2 hc6 1 hcc6 3 dih6
7 n1 1 r1 2 an1 3 di1
8 h5 7 hn2 1 an2 2 di2
9 h6 7 hn3 8 hnh3 1 di3
10 h7 7 hn4 9 hnh4 8 adih4

hn2 = 1.012509; hn3 = 1.012509; hnh3 = 106.177;
hn4 = 1.014174; hnh4 = 106.632; adih4 = 113.434;
cc2 = 1.334805; hc3 = 1.082912; hcc3 = 121.535;
hc4 = 1.082912; hcc4 = 121.535; dih4 = 0.000;
hc5 = 1.082500; hcc5 = 121.377; dih5 = 180.000;
hc6 = 1.082500; hcc6 = 121.377; dih6 = 180.000;
R1 = 3.64373839; AN1 = 79.85551764;
AN2 = -35.50501742; DI1 = 40.04069187;
DI2 = -106.48454186; DI3 = 256.24423510;
```

(Ethene)(H_2O)

```
1 c1
2 c2 1 cc2
3 h1 1 hc3 2 hcc3
4 h2 2 hc4 1 hcc4 3 dih4
5 h3 1 hc5 2 hcc5 4 dih5
6 h4 2 hc6 1 hcc6 3 dih6
7 o1 2 r1 1 an1 3 di1
8 h5 7 ho2 2 an2 1 di2
9 h6 7 ho3 8 hoh3 2 di3

cc2 = 1.334805; hc3 = 1.082912; hcc3 = 121.535;
hc4 = 1.082912; hcc4 = 121.535; dih4 = 0.000;
hc5 = 1.082500; hcc5 = 121.377; dih5 = 180.000;
hc6 = 1.082500; hcc6 = 121.377; dih6 = 180.000;
ho2 = 0.959006; ho3 = 0.962017; hoh3 = 102.830;
R1 = 3.47668315; AN1 = 79.01039190;
AN2 = -15.88741263; DI1 = 71.95504409;
DI2 = -133.01842508; DI3 = 224.14256976;
```

(Ethenole)$_2$

```
1 c1
2 c2 1 cc2
3 o1 2 oc3 1 occ3
4 h1 3 ho4 2 hoc4 1 dih4
5 h2 2 hc5 3 hco5 4 dih5
6 h3 1 hc6 2 hcc6 3 dih6
7 h4 1 hc7 2 hcc7 3 dih7
8 c3 1 r1 2 an1 3 di1
9 c4 8 cc2 1 an2 2 di2
10 o2 9 oc3 8 occ3 1 di3
11 h5 10 ho4 9 hoc4 8 dih4
12 h6 9 hc5 10 hco5 11 dih5
13 h7 8 hc6 9 hcc6 10 dih6
14 h8 8 hc7 9 hcc7 10 dih7

cc2 = 1.331460; oc3 = 1.368409; occ3 = 121.838;
ho4 = 0.961040; hoc4 = 109.043; dih4 = 180.000;
hc5 = 1.083387; hco5 = 116.126; dih5 = 0.000;
hc6 = 1.077458; hcc6 = 119.098; dih6 = 180.000;
hc7 = 1.079447; hcc7 = 121.189; dih7 = 0.007;
R1 = 4.09456533; AN1 = 448.67536258;
AN2 = 448.61021849; DI1 = 23.92025043;
DI2 = -165.25362657; DI3 = 24.03823995;
```

(HCOOH)($HCONH_2$)

```
1 c1
2 o1 1 Aoc2
3 o2 1 Aoc3 2 Aoco3
4 h1 1 Ahc4 2 Ahco4 3 Adih4
5 h2 3 Aho5 1 Ahoc5 2 Adih5
6 c2 1 r1 2 an1 3 di1
7 n1 6 nc2 1 an2 2 di2
8 o3 6 oc3 7 ocn3 3 di3
9 h3 7 hn4 6 hnc4 8 dih4
10 h4 7 hn5 6 hnc5 8 dih5
11 h5 6 hc6 7 hcn6 9 dih6

Aoc2 = 1.220719; Aoc3 = 1.314335; Aoco3 = 126.428;
Ahc4 = 1.093459; Ahco4 = 121.904; Adih4 = 180.000;
Aho5 = 0.995011; Ahoc5 = 109.447; Adih5 = 0.000;
nc2 = 1.342414; oc3 = 1.229066; ocn3 = 125.487;
hn4 = 1.003261; hnc4 = 119.219; dih4 = 180.000;
hn5 = 1.022613; hnc5 = 120.730; dih5 = 0.000;
hc6 = 1.100428; hcn6 = 113.600; dih6 = 0.000;
R1 = 3.93084818; AN1 = 62.87315613;
AN2 = 68.81651771; DI1 = 0.00000000;
DI2 = 0.00000000; DI3 = 0.00000000;
```

(Furan-2-one)$_2$

```
1 c1
2 c2 1 cc2
3 o1 2 oc3 1 occ3
4 c3 3 co4 2 coc4 1 dih4
5 c4 4 cc5 3 cco5 2 dih5
6 o2 2 oc6 3 oco6 4 dih6
7 h1 1 hc7 2 hcc7 3 dih7
8 h2 4 hc8 3 hco8 2 dih8
9 h3 5 hc9 4 hcc9 3 dih9
10 h4 4 hc10 3 hco10 2 dih10
11 c5 1 r1 2 an1 3 di1
12 c6 11 cc2 1 an2 2 di2
13 o3 12 oc3 11 occ3 1 di3
14 c7 13 co4 12 coc4 11 dih4
15 c8 14 cc5 13 cco5 12 dih5
16 o4 12 oc6 13 oco6 14 dih6
17 h5 11 hc7 12 hcc7 13 dih7
18 h6 14 hc8 13 hco8 12 dih8
19 h7 15 hc9 14 hcc9 13 dih9
20 h8 14 hc10 13 hco10 12 dih10

cc2 = 1.488564; oc3 = 1.388872; occ3 = 107.691;
co4 = 1.441607; coc4 = 109.519; dih4 = 0.001;
cc5 = 1.504806; cco5 = 105.327; dih5 = -0.001;
oc6 = 1.214222; oco6 = 122.465; dih6 = 179.999;
hc7 = 1.089116; hcc7 = 122.500; dih7 = -180.000;
hc8 = 1.102176; hco8 = 108.486; dih8 = -120.806;
hc9 = 1.091599; hcc9 = 123.591; dih9 = 180.000;
hc10 = 1.102176; hco10 = 108.486; dih10 = 120.804;
R1 = 4.68602580; AN1 = -48.06056601;
AN2 = 41.60667441; DI1 = 103.68270546;
DI2 = 176.92467655; DI3 = 70.96577326;
```

(Furan)(EtOH)

```
1 c1
2 o1 1 oc2
3 c2 2 co3 1 coc3
4 c3 3 cc4 2 cco4 1 dih4
5 c4 4 cc5 3 ccc5 2 dih5
6 h1 1 hc6 2 hco6 3 dih6
7 h2 5 hc7 4 hcc7 3 dih7
8 h3 4 hc8 3 hcc8 2 dih8
9 h4 3 hc9 2 hco9 1 dih9
10 c5 1 r1 2 an1 3 di1
11 c6 10 acc2 1 an2 2 di2
12 o2 11 aoc3 10 aocc3 1 di3
13 h5 12 aho4 11 ahoc4 10 adih4
14 h6 11 ahc5 12 ahco5 13 adih5
15 h7 10 ahc6 11 ahcc6 12 adih6
16 h8 10 ahc7 11 ahcc7 12 adih7
17 h9 10 ahc8 11 ahcc8 12 adih8
18 h10 11 ahc9 10 ahcc9 15 adih9

oc2 = 1.360041; co3 = 1.360075; coc3 = 106.847;
cc4 = 1.364967; cco4 = 110.439; dih4 = 0.000;
cc5 = 1.425707; ccc5 = 106.137; dih5 = 0.000;
hc6 = 1.075018; hco6 = 115.808; dih6 = 180.000;
hc7 = 1.076062; hcc7 = 127.798; dih7 = 180.000;
hc8 = 1.076062; hcc8 = 126.063; dih8 = 180.000;
hc9 = 1.075012; hco9 = 115.806; dih9 = 180.000;
acc2 = 1.510238; aoc3 = 1.428745; aocc3 = 107.364;
aho4 = 0.962098; ahoc4 = 108.307; adih4 = -179.900;
ahc5 = 1.093518; ahco5 = 110.396; adih5 = -59.598;
ahc6 = 1.089376; ahcc6 = 110.393; adih6 = -179.574;
ahc7 = 1.088457; ahcc7 = 110.168; adih7 = -59.455;
ahc8 = 1.088301; ahcc8 = 110.216; adih8 = 60.248;
ahc9 = 1.093753; ahcc9 = 110.280; adih9 = -59.277;
R1 = 5.91716494; AN1 = 52.18408514;
AN2 = 73.44769996; DI1 = 3.36095344;
DI2 = 123.08139248; DI3 = 82.97227812;
```

(Furan)(Furan-2-one)

```
1 c1
2 o1 1 oc2
3 c2 2 co3 1 coc3
4 c3 3 cc4 2 cco4 1 dih4
5 c4 4 cc5 3 ccc5 2 dih5
6 h1 1 hc6 2 hco6 3 dih6
7 h2 5 hc7 4 hcc7 3 dih7
8 h3 4 hc8 3 hcc8 2 dih8
9 h4 3 hc9 2 hco9 1 dih9
10 c5 1 r1 2 an1 3 di1
11 c6 10 Acc2 1 an2 2 di2
12 o2 11 Aoc3 10 Aocc3 1 di3
13 c7 12 Aco4 11 Acoc4 10 Adih4
14 c8 13 Acc5 12 Acco5 11 Adih5
15 o3 11 Aoc6 12 Aoco6 13 Adih6
16 h5 10 Ahc7 11 Ahcc7 12 Adih7
17 h6 13 Ahc8 12 Ahco8 11 Adih8
18 h7 14 Ahc9 13 Ahcc9 12 Adih9
19 h8 13 Ahc10 12 Ahco10 11 Adih10

oc2 = 1.360041; co3 = 1.360075; coc3 = 106.847;
cc4 = 1.364967; cco4 = 110.439; dih4 = 0.000;
cc5 = 1.425707; ccc5 = 106.137; dih5 = 0.000;
hc6 = 1.075018; hco6 = 115.808; dih6 = 180.000;
hc7 = 1.076062; hcc7 = 127.798; dih7 = 180.000;
hc8 = 1.076062; hcc8 = 126.063; dih8 = 180.000;
hc9 = 1.075012; hco9 = 115.806; dih9 = 180.000;
Acc2 = 1.488564; Aoc3 = 1.388872; Aocc3 = 107.691;
Aco4 = 1.441607; Acoc4 = 109.519; Adih4 = 0.001;
Acc5 = 1.504806; Acco5 = 105.327; Adih5 = -0.001;
Aoc6 = 1.214222; Aoco6 = 122.465; Adih6 = 179.999;
Ahc7 = 1.089116; Ahcc7 = 122.500; Adih7 = -180.000;
Ahc8 = 1.102176; Ahco8 = 108.486; Adih8 = -120.806;
Ahc9 = 1.091599; Ahcc9 = 123.591; Adih9 = 180.000;
Ahc10 = 1.102176; Ahco10 = 108.486;
Adih10 = 120.804;
R1 = 3.40877785; AN1 = -86.45490818;
AN2 = -71.03036899; DI1 = -262.41528016;
DI2 = -300.02614924; DI3 = 84.92176352;
```

(H_2O)(NH_3)

```
1 o1
2 h1 1 ho2
3 h2 1 ho3 2 hoh3
4 n1 1 r1 2 an1 3 di1
5 h3 4 hn2 1 an2 2 di2
6 h4 4 hn3 5 hnh3 1 di3
7 h5 4 hn4 6 hnh4 5 dih4

hn2 = 1.012509; hn3 = 1.012509; hnh3 = 106.177;
hn4 = 1.014174; hnh4 = 106.632; dih4 = 113.434;
ho2 = 0.959006; ho3 = 0.962017; hoh3 = 102.830;
R1 = 2.97455824; AN1 = -250.76546720;
AN2 = 124.84974359; DI1 = -0.01208484;
DI2 = 13.44272003; DI3 = 240.02643311;
```

References

[1] Marchetti, O.; Werner, H. *J. Phys. Chem. A* **2009**, *113*, 11580–11585.

[2] Jureč, P.; Černy, J.; Šponer, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, 1985–1993.

[3] Reza c, J.; Riley, K. E.; Hobza, P. *J. Chem. Theory Comput.* **2011**, *7*, 2427–2438.

[4] Allen, L. C.; Karo, A. M. *Rev. Mod. Phys.* **1960**, *32*, 275–285.

[5] Grimme, S. *J. Chem. Phys.* **2003**, *118*, 9095–9102.

[6] Pitoňák, M.; Neogrády, P.; Černỳ, J.; Grimme, S.; Hobza, P. *ChemPhysChem* **2009**, *10*, 282–289.

[7] London, F. *Zeitschrift für Physik* **1930**, *63*, 245–279.

[8] McNaught, A. D.; Wilkinson, A. *Blackwell Scientific Publications* **Last update: 2012-08-19; version: 2.3.2.**, 1007, . Compiled by . Blackwell Scientific Publications, Oxford (1997). XML on-line corrected version: (2006-) created by M. Nic, J. Jirat, B. Kosata; updates compiled by A. Jenkins. ISBN 0-9678550-9-8. doi:10.1351/goldbook.

[9] Samadikhah, K.; Larsson, R.; Bazooyar, F.; Bolton, K. *Comp. Mater. Sci.* **2012**, *53*, 37–43.

[10] Puthoff, J.; Holbrook, M.; Wilkinson, M.; Autumn, K. *MRS Online Proceedings Library* **2012**, *1465*, year.

[11] Schrödinger, E. *Naturwissenschaften* **1935**, *23*, 152.

[12] Stone, A. *The theory of intermolecular forces*; Clarendon Press Oxford, 1996; Vol. 32.

[13] Power, E. *Intermolecular Forces*; Advanced Chemical Physics, 1967; Vol. 12.

[14] Schrödinger, E. *Annalen der Physik* **1926**, *79*, 361, 489, 734.

[15] Helgaker, T.; Jørgensen, P.; Olsen, J. *Molecular electronic-structure theory*, 2000.

[16] Dunning, T. *J. Chem. Phys.* **1989**, *90*, 1007.

[17] Patkowski, K. *J. Chem. Phys.* **2013**, *138*, 154101.

[18] Patkowski, K. *J. Chem. Phys.* **2012**, *137*, 034103.

[19] Yousaf, K.; Peterson, K. *J. Chem. Phys.* **2008**, *129*, 184108.

[20] Lowdin, P. *Adv. Chem. Phys* **1959**, *2*, 207.

[21] Davidson, E.; Chakravorty, S. *Chem. Phys. Lett.* **1994**, *217*, 48–54.

[22] Bernardi, F.; Boys, S. F. *Mol. Phys.* **1970**, *19*, 553.

[23] Schütz, M.; Werner, H.-J.; Lindh, R.; Manby, F. R. *J. Chem. Phys.* **2004**, *121*, 737.

[24] Sherrill, C. D. *School of Chemistry and Biochemistry* **1996**.

[25] Bak, K.; Jørgensen, P.; Olsen, J.; Helgaker, T.; Klopper, W. *J. Chem. Phys.* **2000**, *112*, 9229–9242.

[26] Pople, J.; Head-Gordon, M.; Raghavachari, K. *J. Chem. Phys.* **1987**, *87*, 5968–5975.

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

[28] Boughton, J. W.; Pulay, P. *J. Comput. Chem.* **1993**, *14*, 736–740.

[29] Boys, S. F. Localized Orbitals and Localized Adjustment Functions. In *Quantum Theory of Atoms, Molecules, and the Solid State*; Löwdin, P. O., Ed.; Academic Press: New York, 1966; pp 253–262.

[30] Pipek, J.; Mezey, P. G. *J. Chem. Phys.* **1989**, *90*, 4916–4926.

[31] Watts, J. D.; Gauss, J.; Bartlett, R. J. *J. Chem. Phys.* **1993**, *98*, 8718.

[32] Deegan, M. J.; Knowles, P. J. *Chem. Phys. Lett.* **1994**, *227*, 321–326.

[33] Karton, A.; Rabinovich, E.; Martin, J. M.; Ruscic, B. *J. Chem. Phys.* **2006**, *125*, 144108–144108.

[34] Stanton, J. F. *Chem. Phys. Lett.* **1997**, *281*, 130–134.

[35] Urban, M.; Noga, J.; Cole, S. J.; Bartlett, R. J. *J. Chem. Phys.* **1985**, *83*, 4041.

[36] Pople, J. A.; Head-Gordon, M.; Raghavachari, K. *J. Chem. Phys.* **1987**, *87*, 5968.

[37] Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**, *157*, 479.

[38] Jensen, F. *Introduction to computational chemistry*; Wiley, 2007.

[39] Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. *Chem. Phys. Lett.* **1998**, *286*, 243–252.

[40] Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Olsen, J. *Chem. Phys. Lett.* **1999**, *302*, 437–446.

[41] Karton, A.; Martin, J. *Theor. Chem. Acc.* **2006**, *115*, 330–333.

[42] Dunning, Jr., T. H. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

[43] Hylleraas, E. A. *Zeitschrift für Physik* **1929**, *54*, 347–366.

[44] Kutzelnigg, W.; Klopper, W. *J. Chem. Phys.* **1991**, *94*, 1985–2001.

[45] Klopper, W.; Samson, C. C. M. *J. Chem. Phys.* **2002**, *116*, 6397–6410.

[46] Valeev, E. F. *Chem. Phys. Lett.* **2004**, *395*, 190–195.

[47] Ten-no, S. *Chem. Phys. Lett.* **2004**, *398*, 56–61.

[48] Samson, C. C. M.; Klopper, W.; Helgaker, T. *Comp. Phys. Comm.* **2002**, *149*, 1–10.

[49] May, A. J.; Manby, F. R. *J. Chem. Phys.* **2004**, *121*, 4479–4485.

[50] May, A. J.; Valeev, E.; Polly, R.; Manby, F. R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 2710.

[51] Tew, D.; Klopper, W. *J. Chem. Phys.* **2005**, *123*, 074101.

[52] Manby, F. R.; Werner, H.-J.; Adler, T. B.; May, A. J. *J. Chem. Phys.* **2006**, *124*, 094103.

[53] Valeev, E. F. *J. Chem. Phys.* **2006**, *125*, 244106.

[54] Werner, H.-J.; Adler, T. B.; Manby, F. R. *J. Chem. Phys.* **2007**, *126*, 164102.

[55] Adler, T. B.; Knizia, G.; Werner, H.-J. *J. Chem. Phys.* **2007**, *127*, 221106.

[56] Marchetti, O.; Werner, H. *Phys. Chem. Chem. Phys.* **2008**, *10*, 3400–3409.

[57] Werner, H.; Adler, T.; Manby, F. *J. Chem. Phys.* **2007**, *126*, 164102.

[58] Knizia, G.; Adler, T.; Werner, H. *J. Chem. Phys.* **2009**, *130*, 054104.

[59] Adler, T.; Knizia, G.; Werner, H. *J. Chem. Phys.* **2007**, *127*, 221106.

[60] Rauhut, G.; Knizia, G.; Werner, H. *J. Chem. Phys.* **2009**, *130*, 054105.

[61] Prof. Dr. H.-J. Werner, *private communication*, 2007.

[62] Tew, D. P.; Klopper, W. *J. Chem. Phys.* **2006**, *125*, 094302.

[63] Klopper, W.; Kutzelnigg, W. *Chem. Phys. Lett.* **1987**, *134*, 17–22.

[64] Klopper, W.; Kutzelnigg, W. *J. Phys. Chem.* **1990**, *94*, 5625.

[65] Wind, P.; Klopper, W.; Helgaker, T. *Theor. Chem. Acc.* **2002**, *107*, 173–179.

[66] Kato, T. *Communications on Pure and Applied Mathematics* **1957**, *10*, 151–177.

[67] Werner, H.-J. et al. *MOLPRO, version 2012.1, a package of ab initio programs*, 2012, see.

[68] Jurečka, P.; Šponer, J.; Černỳ, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.

[69] Reza c, J.; Riley, K. E.; Hobza, P. *J. Chem. Theory Comput.* **2011**, *7*, 2427–2438.

[70] Werner, H.-J.; Adler, T. B.; Manby, F. R. *J. Chem. Phys.* **2007**, *126*, year.

[71] Vogiatzis,; Klopper, P. D. *private communication*, 2009.

[72] Marshall, M. S.; Sherrill, C. D. *J. Chem. Theory Comput.* **2011**, *7*, 3978–3982.

[73] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.

[74] Vydrov, O. A.; Van Voorhis, T. *Phys. Rev. Lett.* **2009**, *103*, 063004.

[75] Takatani, T.; Hohenstein, E. G.; Sherrill, C. D. *J. Chem. Phys.* **2008**, *128*, 124111.

[76] Pitoňák, M.; Řezáč, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2010**, *12*, 9611–9614.

[77] Hill, J. G.; Platts, J. A. *J. Chem. Theory Comput.* **2007**, *3*, 80–85.

[78] Jung, Y.; Lochan, R. C.; Dutoi, A. D.; Head-Gordon, M. *J. Chem. Phys.* **2004**, *121*, 9793.

List of Figures

List of Tables